



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIA E INGENIERÍA DE LA COMPUTACIÓN

CONTRIBUCIÓN DE LA INFORMACIÓN ENTONATIVA AL
RECONOCIMIENTO DE ACTOS DE DIÁLOGO EN UN CORPUS DE
DIÁLOGOS PRÁCTICOS

T E S I S

QUE PARA OBTENER EL GRADO DE:

**DOCTOR EN CIENCIAS
(COMPUTACIÓN)**

P R E S E N T A:

SERGIO RAFAEL CORIA OLGUÍN

DIRECTOR DE TESIS: DR. LUIS ALBERTO PINEDA CORTÉS

México, D.F.

2008.



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

Agradezco las enseñanzas y orientaciones del Dr. Luis Alberto Pineda Cortés, director de esta tesis; así como las de los integrantes de mi Comité Doctoral: Dr. James Allen, de la Universidad de Rochester, Dra. Elizabeth Shriberg, de la Universidad de California en Berkeley, Dr. Joaquim Llisterri, de la Universidad Autónoma de Barcelona (UAB), Dra. Katya Rodríguez, de la UNAM y Dr. Christian Lemaitre, de la Universidad Autónoma Metropolitana (UAM). También a la Dra. Montserrat Riera, de la UAB y al Mtro. Javier Cuétara, de la Fac. de Filosofía.

Gracias también a mis compañeros estudiantes del grupo DIME en el IIMAS-UNAM; en particular, a Varinia Estrada, por su participación y valiosos comentarios en el proceso de etiquetación de actos de diálogo del corpus DIME.

Agradezco al Posgrado en Ciencia e Ingeniería de la Computación.

El trabajo aquí reportado se ha desarrollado dentro del contexto del proyecto DIME, con apoyo parcial del Consejo Nacional de Ciencia y Tecnología, CONACyT. También agradezco el apoyo de la Dirección General de Estudios de Posgrado, DGEP, de la UNAM.

Ciudad Universitaria, D.F., octubre de 2007.

Sergio Coria

Contribución de la Información Entonativa al Reconocimiento de Actos de Diálogo en un Corpus de Diálogos Prácticos: Tesis Doctoral

Índice

Agradecimientos

Introducción

1. El acto de diálogo y la entonación

El acto de habla

El acto de diálogo

El grounding o aterrizamiento

Esquemas de etiquetaje de actos de diálogo

El esquema DAMSL

El esquema DIME-DAMSL

Concepto de acto de habla en el presente trabajo

La entonación

Relación entre el acto de diálogo y la entonación

Niveles de representación lingüística para el reconocimiento del acto de diálogo

Niveles de representación útiles para la modelación computacional del acto de diálogo

Necesidad del parseo para la interpretación de los actos de diálogo

2. Estado del arte

Algunos actos de diálogo cuya detección podría beneficiarse de la información prosódica

El trabajo de Shriberg et al.

El proyecto VERBMOBIL

Trabajo de Fernández y Picard

Comparación entre Shriberg et al. y VERBMOBIL

Comparación entre los enfoques previos y el del presente trabajo

Importancia de la relación entre contorno entonativo y la modalidad del enunciado

El tonema en español: noción y tipología

Principales estudios sobre la relación entre el tonema y la modalidad del enunciado

3. Teoría de DIME-DAMSL

Motivación y objetivos

Esquema DAMSL

Esquema de VERBMOBIL

Esquema de Levin et al.

DIME-DASML: teoría

Transacción

Los planos de comunicación de DIME-DAMSL

Plano de las obligaciones

Plano del common ground

Subplano del acuerdo

Subplano del entendimiento

Cargos y abonos
Consistencia de los etiquetajes

4. Representación de la entonación
Requerimientos de la representación
Enfoque del procesamiento de señales
Enfoque de la Fonética: INTSINT
Enfoque de la Fonología: ToBI
Adaptación de INTSINT al presente trabajo
Ventajas de la representación

5. Experimentos y resultados 5.1 Atributos del acto de diálogo
Atributos de la etiquetación de entonación
Atributos de la duración del enunciado
Atributos del rol del hablante
Atributos de la modalidad del enunciado
Análisis estadístico
Configuración de los experimentos
Resultados de aprendizaje automático
Configuración 1: Atributos entonativos como predictores
Configuración 2: Evaluando atributos no entonativos
Configuración 3: Evaluando el acto de diálogo previo como predictor
Discusión sobre las reglas

6. Conclusiones
Postulación final
Acto de habla
Niveles de representación lingüística
Comparación con el estado del arte
Esquemas para etiquetaje de actos de diálogo
Esquema DIME-DAMSL
DIME-DAMSL: Aplicación
Uso del estadístico Kappa
Representación de la entonación
Experimentos y resultados
Análisis estadístico
Configuración de los experimentos
Configuración 1: Sólo atributos entonativos
Configuración 2: Atributos entonativos con otros no entonativos
Configuración 3: Utilización del acto de diálogo previo
Resultados de los experimentos
Discusión de las reglas
Aplicaciones potenciales
Etiquetación automática de la modalidad del enunciado
Soporte a la etiquetación automática de actos de diálogo
Trabajo futuro

Referencias

Tablas

- 2.1 Cuadro comparativo de los enfoques de trabajos previos y del presente
- 2.2 Patrones entonativos de diversas modalidades del enunciado, resumido por Garrido (1996)
- 2.3 Modalidades de enunciado tipificadas por Navarro Tomás (1948)
- 2.4 Modalidades de enunciado presentadas en (Quilis, 1981)
- 2.5 Porcentajes de contornos entonativos finales por tipo de enunciado (Garrido, 1996, p. 147)
- 3.1 Actos de diálogo de la función hacia delante en DAMSL (Allen y Core, 1997)
- 3.2 Actos de diálogo de la función hacia atrás en DAMSL (Allen y Core, 1997)
- 3.3 Los 18 actos de diálogo de mayor nivel de VERBMOBIL-1 (Jekat et al., 1995), jerarquizados por Jurafsky (2006)
- 3.4 Categorías de actos de diálogo en el esquema de Levin et al. (1998)
- 3.5 Etiquetas DIME-DAMSL del plano de obligaciones, con ejemplos del corpus DIME
- 3.6 Etiquetas DIME-DAMSL del plano de common ground, subplano del acuerdo, con ejemplos del corpus DIME
- 3.7 Etiquetas DIME-DAMSL del plano de common ground, subplano del entendimiento, con ejemplos del corpus DIME
- 3.8 Pares equilibrados en el plano de las obligaciones (Pineda et al., 2006a)
- 3.9 Pares equilibrados en el plano del common ground (Pineda et al., 2006a)
- 3.10 Etiquetaje DIME-DAMSL Preliminar de una transacción del corpus DIME
- 3.11 Cantidades de transacciones identificadas por 2 anotadores en 12 diálogos del corpus DIME
- 3.12 Valores Kappa de etiquetajes de inicio/fin de transacción, ocurrencia de segmento, y cargos/abonos (2 anotadores, fecha de cómputo: octubre de 2006)
- 3.13 Porcentajes básicos de acuerdo y valores Kappa en etiquetajes de acto de diálogo con DIME-DAMSL Preliminar (2 anotadores, fecha de cómputo: octubre de 2006)
- 4.1 Algunos parámetros usados por Shriberg et al. (1998)
- 4.2 Algunos parámetros usados por Garrido (1991)
- 4.3 Algunos parámetros usados por Fernández y Picard (2002)
- 4.4 Algunos parámetros usados por Kompe et al. (1995)
- 5.1 Duraciones, número de turnos y de enunciados de los 12 diálogos seleccionados
- 5.2 Análisis de Pareto de las etiquetaciones de actos de diálogo
- 5.3 Atributos implementados en el dataset usando datos del corpus
- 5.4 Resultados de los experimentos
- 5.5 Las 5 reglas con mayor soporte para modalidad del enunciado (modelo 1.3)
- 5.6 Medidas F, recalls y precisions de modalidad predicha optimizada (modelo 1.3)
- 5.7 Medidas F, recalls y precisions de modalidad predicha optimizada (modelo 2.1)

- 5.8 Las 5 reglas con mayor soporte para modalidad predicha optimizada (modelo 2.1)
- 5.9 Medidas F, recalls y precisions del modelo para actos de diálogo de obligaciones (modelo 2.2.A)
- 5.10 Medidas F, recalls y precisions del modelo para actos de diálogo de common ground (modelo 2.2.B)
- 5.11 Las 5 reglas con mayor soporte para actos de diálogo de obligaciones (modelo 2.2.A)
- 5.12 Las 5 reglas con mayor soporte para actos de diálogo de common ground (modelo 2.2.B)
- 5.13 Las cinco reglas con mayor soporte para actos de diálogo de obligaciones (modelo 3.1.A)
- 5.14 Las 5 reglas con mayor soporte para actos de diálogo de common ground (modelo 3.1.B)
- 5.15 Porcentajes de presencia de los atributos en las reglas de los árboles del modelo 3.1:
 - A) Obligaciones y
 - B) Common ground
- 5.16 Medidas F, recalls y precisions de actos de diálogo de obligaciones (modelo 3.1.A)
- 5.17 Medidas F, recalls y precisions de actos de diálogo de common ground (modelo 3.1.B)

Figuras

- 2.1 Representación gráfica de la tipología de tonemas del español de Navarro Tomás según la interpretación de García-Lecumberri
- 4.1 Frecuencia fundamental original
- 4.2 F0 estilizada (contorno grueso) con puntos de inflexión (círculos)
- 4.3 Regiones del contorno estilizado que requieren modificación
- 4.4 F0 estilizada modificada en la verificación perceptiva
- 4.5 Etiquetaje INTSINT de los puntos de inflexión
- 4.6 Contenido del archivo de etiquetaje INTSINT en formato eti
- 4.7 Etiquetaje INTSINT transformado en secuencia alfabética de etiquetas sin información de tiempo
- 4.8 Etiquetaje INTSINT sin etiquetas S múltiples
- 4.9 Muestra de secuencias INTSINT de algunos enunciados del corpus DIME

Apéndices

- Apéndice I. El Corpus DIME
 - I.1 Motivación y objetivos
 - I.2 Condiciones de la grabación
 - I.3 Determinación de los niveles de etiquetación
 - I.3.1 Alfabeto MexBet y herramienta Transcribemex
 - I.3.2 Alófonos
 - I.3.3 Sílabas fonéticas
 - I.3.4 Palabras
 - I.3.5 Índices de disyunción (break indices)
 - I.3.6 Entonación
 - I.3.7 Partes del habla
 - I.3.8 Marcadores del discurso

I.3.9 Reparaciones del habla (speech repairs)
I.3.10 Modalidad del enunciado
I.3.11 Actos de diálogo: DIME-DAMSL
I.4 Control del proceso y de la calidad de la etiquetación

Apéndice II. DIME-DAMSL detallado y uso del estadístico Kappa II.1 DIME-DAMSL Detallado
II.2 DIME-DAMSL: Aplicación
II.3 Uso del estadístico Kappa

Apéndice III. Análisis estadístico de la modalidad del enunciado

Apéndice IV. Análisis estadístico del rol del hablante

Apéndice V. Análisis estadístico de etiquetaciones INTSINT

Apéndice VI. Análisis estadístico de las etiquetaciones de DIME-DAMSL Preliminar

Apéndice VII. Análisis de correlación entre actos de diálogo y modalidades de enunciado

Apéndice VIII. Análisis de correlación entre rol del hablante y acto de diálogo

Apéndice IX. Análisis de correlaciones entre actos de diálogo de dos planos

Apéndice X. Reglas del modelo 2.1 para modalidad predicha optimizada (ordenadas por soporte)

Apéndice XI. Reglas del modelo 3.1.A para actos de diálogo de obligaciones (ordenadas por soporte)

Apéndice XII. Reglas del modelo 3.1.B para actos de diálogo de common ground

Apéndice XIII. Árboles de decisión generados con WEKA para la configuración 1

Apéndice XIV. Árboles de decisión generados con WEKA para la configuración 2

Apéndice XV. Árboles de decisión generados con WEKA para la configuración 3

Introducción

En el campo de la Inteligencia Artificial (IA), el procesamiento del lenguaje hablado es, en la primera década del siglo XXI, un área que presenta diversos retos a la investigación. Algunos de los principales son el mejoramiento de la robustez de los sistemas de reconocimiento automático del habla y el aprovechamiento de las diversas fuentes de información que los seres humanos usan al participar en el diálogo. Aunque el contenido léxico del habla es la principal fuente informativa para los participantes de un diálogo, otras fuentes, tales como la entonación, el rol que el hablante desempeña en el diálogo y los actos de diálogo que preceden a un enunciado, son valiosas para una adecuada comprensión de la intención.

Dado que la IA desea lograr un eficiente procesamiento del lenguaje hablado, toma en cuenta las aportaciones de las diversas áreas del conocimiento que estudian los fenómenos lingüísticos, entre ellas la Lingüística, la Filosofía del Lenguaje, la Psicolingüística y la Fonética.

Una de las aportaciones más reconocidas en la Filosofía del Lenguaje proviene de Searle, quien afirma que el acto de habla es la unidad mínima de comunicación lingüística; esta idea introduce un nuevo enfoque en el estudio lingüístico porque coloca en segundo plano a otros elementos del lenguaje, tales como el símbolo, la palabra o el fonema. La Lingüística, por su parte, ha sugerido que uno de los medios de expresión del acto de habla es la entonación del enunciado.

El estudio empírico de la relación entre el acto de habla y la entonación puede ofrecer un mejor entendimiento del fenómeno y además proporcionar medios para mejorar las implementaciones de sistemas computacionales para reconocimiento automático del habla y de administración del diálogo.

El análisis y la modelación computacional del diálogo como fenómeno lingüístico presentan un alto grado de complejidad; por ello, la Lingüística Computacional lo estudia bajo condiciones controladas y dentro de ámbitos delimitados. De este modo, Allen identifica los *diálogos prácticos*, aquellos cuyos interlocutores han predefinido (implícita o explícitamente) una meta que debe ser alcanzada en forma colaborativa, estableciendo estructuras comunicativas más simples que en los diálogos informales de índole genérica. Además de su valor teórico, el estudio de los diálogos prácticos también tiene valor ingenieril, puesto que a esta clase pertenecen la mayoría de los diálogos susceptibles de establecerse entre un humano y una computadora mediante una interfaz de lenguaje hablado.

En Lingüística Computacional, el acto de habla se ha estudiado frecuentemente bajo la forma de *acto de diálogo*; es decir, concentrándose en el aspecto de su función comunicativa dentro del contexto del diálogo, más que considerándolo en forma aislada.

El enfoque tradicionalmente abordado para reconocer el acto de diálogo ha sido mediante el reconocimiento automático del habla, aplicando parseo, modelos de lenguaje y en algunos casos, planeación. En la década de 1990, se ha abordado un enfoque más amplio y se ha evaluado la contribución de otras fuentes de información, tales como la entonación del habla y la estructura inherente al diálogo.

Tomando en cuenta lo anterior, este trabajo aborda el análisis y la modelación de la relación entre el acto de diálogo y la entonación; en forma complementaria, también se analiza y modela la contribución de otras fuentes de información que contribuyen al reconocimiento del acto de diálogo. El objetivo es al mismo tiempo teórico y práctico, ya que se intenta describir y comprender el fenómeno, así como crear modelos que conduzcan a la implementación de sistemas eficientes para la administración del diálogo hablado. El problema a resolver se plantea como la evaluación empírica de la contribución de la información entonativa del habla al reconocimiento de actos de diálogo usando algoritmos de Aprendizaje Automático (*Machine Learning*) para crear árboles de clasificación con datos de un corpus de diálogos prácticos. Además de las contribuciones teóricas, este trabajo podría ofrecer modelos susceptibles de implementación en sistemas de administración de diálogo. El trabajo se ubica dentro del contexto del Proyecto denominado Diálogos Inteligentes Multimodales en Español (DIME, Pineda *et al.*, 2002). El análisis de fuentes de información léxica del habla no es parte del alcance de la presente investigación.

Se requiere un enfoque empírico porque el fenómeno necesita analizarse en forma cuantitativa a partir de datos reales del habla. Por ello, en el presente trabajo se realizan análisis estadísticos y se crean modelos basados en Aprendizaje Automático; el primero permite describir el fenómeno y el segundo explicar cómo sus variables están relacionadas entre sí. Adicionalmente, los modelos de Aprendizaje Automático podrían facilitar la creación o robustecimiento de sistemas de reconocimiento automático del habla o de administración de diálogo. La visión empírica también ha sido tomada por algunos trabajos previos en el área, como el de Shriberg *et al.* (1998), el proyecto VERBMOBIL (Wahlster, 1993) y el de Fernández y Picard (2002).

El presente trabajo se apoya en las teorías sobre el acto de habla y el acto de diálogo y en las teorías sobre la entonación del español. Asimismo, se usan esquemas de etiquetaje en forma intensiva, tanto para actos de diálogo como para entonación. Se toman en cuenta los resultados empíricos de trabajos previos en el área y se presentan los resultados propios. A continuación se muestra un panorama del contenido temático.

En el **capítulo 1** se presentan las bases teóricas que guían el análisis de la relación entre el acto de habla y la entonación; se revisan las principales nociones relacionadas con el acto de habla desde los puntos de vista de Grice (1957 y 1975) y de Searle (1969). Se repasa también el concepto de *grounding* de Clark y Schaefer (1989), con el propósito de hacerlo distinguible del mismo término bajo el concepto introducido posteriormente en el esquema DIME-DAMSL.

El acto de habla ha sido frecuentemente estudiado en la Lingüística Computacional con el término *acto de diálogo*, entendiéndolo como una de las unidades componentes del diálogo y se le estudia bajo la perspectiva de la función que desempeña dentro del contexto particular de la conversación. También se han desarrollado esquemas de etiquetaje para su análisis y modelación. Uno de los esquemas más conocidos es DAMSL (*Dialogue Act Markup in Several Layers*), *Marcaje de Actos de Diálogo en Varias Capas*, de Allen y Core (1997).

Con base en DAMSL y en una variante del concepto de *grounding* de Clark y Schaefer, se desarrolla el esquema DIME-DAMSL (Pineda *et al.*, 2006, 2006a), descrito en el capítulo 3. En este esquema el acto de diálogo se da a lo largo de planos de comunicación, dentro de una estructura de mayor alcance que el enunciado, la *transacción*, en la cual se observa un estado de equilibrio entre actos de diálogo. El *grounding* en DIME-DAMSL consiste en un conjunto de actos que construyen o refuerzan los conocimientos compartidos, las creencias y las presuposiciones de los participantes del diálogo; incluye también actos cuya función es administrar el flujo comunicativo. Esta acepción difiere de la correspondiente a Clark y Schaefer, quienes refieren así al proceso mismo de incrementar los conocimientos compartidos, creencias y presuposiciones.

A partir de las teorías sobre el acto de habla y de los esquemas de etiquetaje de acto de diálogo, en el presente trabajo se sintetiza un concepto operacional del acto de diálogo, útil para los propósitos empíricos propios.

Trabajos teóricos y empíricos previos han establecido que la detección e interpretación de los actos de diálogo depende de diversas fuentes de información involucradas en el habla, representadas en diversos niveles. Entre algunos de éstos existe un considerable grado de redundancia, por lo cual la evaluación empírica de sus correspondientes grados de contribución es de gran utilidad. El nivel de información más estudiado ha sido el del contenido léxico del enunciado y en segundo término otros niveles, tales como la entonación, la modalidad del enunciado, el contexto que le antecede, etc. Dado que el contenido léxico es considerado la principal fuente, se han desarrollado múltiples esfuerzos para mejorar las técnicas de parseo; sin embargo, al considerar otros niveles se podría mejorar la precisión y robustez de los sistemas de reconocimiento automático del habla y de administración de diálogo. Un ejemplo de los beneficios que podrían obtenerse es el de los enunciados que, teniendo un mismo contenido léxico, pueden expresar actos de diálogo distintos dependiendo de su respectiva entonación.

El **capítulo 2** describe tres trabajos previos destacados acerca de la relación entre entonación y acto de diálogo: el de Shriberg *et al.*, VERBMOBIL y el de Fernández y Picard.

Shriberg *et al.* analizan y modelan la relación entre el acto de diálogo y la entonación para el idioma inglés americano usando un corpus de diálogos prácticos donde los participantes conversan por teléfono sobre

tópicos de interés general. La representación de la entonación está dada mediante parámetros estadísticos obtenidos con un enfoque ingenieril de procesamiento de señales. Los actos de diálogo se etiquetan con SWBD-DAMSL, una variante de DAMSL, adaptada para su uso en el corpus de diálogos Switchboard. La modelación se hace usando algoritmos de generación de árboles de decisión de estilo CART; adicionalmente, se usan modelos estadísticos de lenguaje como línea de base para comparar sus respectivos desempeños. Una limitación para los propósitos del presente trabajo es que usó una representación de la entonación que no toma en cuenta el aspecto perceptivo del fenómeno y utiliza un esquema de etiquetaje de actos de diálogo que no garantiza una consistencia satisfactoria, además de estar orientado a conversaciones telefónicas sobre tópicos generales y no a diálogos prácticos.

El proyecto VERBMOBIL, cuya meta es la construcción de sistemas de traducción automática de lenguaje hablado, se concentra en el análisis y la modelación de los tipos de actos de diálogo que ocurren dentro del dominio específico del agendamiento de reuniones de negocios. En este dominio, la cantidad de tipos de actos de diálogo que se observan es menor que en los diálogos prácticos genéricos. En VERBMOBIL, se estudia la entonación del habla como fuente de información para robustecer el reconocimiento automático de actos de diálogo. En ese proyecto se usaron otras técnicas en forma complementaria; principalmente redes neuronales artificiales con topología de Perceptrón Multicapa, así como estructuras de poligrama, una variante del modelo de n -grama para implementar modelos de lenguaje. Una limitación en el proyecto VERBMOBIL es la implementación de un esquema de etiquetaje exclusivo para un dominio especializado, lo cual dificulta o impide su uso en otros dominios.

En el trabajo de Fernández y Picard (2002) también se analiza y modela la relación entre la entonación del habla y el acto de diálogo, en particular, usando datos provenientes de un corpus en idioma Español, donde los participantes son familiares que desarrollan conversaciones informales por teléfono. El hecho de que las conversaciones sean de tipo informal y no diálogos prácticos, constituye una diferencia destacable respecto a los diálogos desarrollados en el Corpus Switchboard, usado en el trabajo de Shriberg *et al.* y respecto al corpus DIME, usado en el presente trabajo. El esquema que usan Fernández y Picard para etiquetar actos de diálogo es tomado de Levin *et al.* (1998), basado, a su vez, en DAMSL. El algoritmo de Aprendizaje Automático es el denominado Máquina de Soporte Vectorial (SVM, *Support Vector Machine*). El objetivo principal de su trabajo es evaluar la precisión del algoritmo SVM para el reconocimiento de actos de diálogo, comparándolo con otros algoritmos de Aprendizaje Automático, como los árboles de decisión y las redes neuronales artificiales. Al igual que en el trabajo de Shriberg, una limitación es el uso de una representación de la entonación que no toma en cuenta el aspecto perceptivo.

Al comparar los tres trabajos previos con el presente, algunas de las semejanzas más notorias son: 1) al igual que en Shriberg *et al.* y Fernández y Picard, se usa un esquema de etiquetaje basado en DAMSL; 2) al igual que en VERBMOBIL, se usa un corpus de diálogos prácticos; 3) al igual que Shriberg *et al.*, se usa un algoritmo de árbol de decisión. Las diferencias más visibles son: 1) el presente trabajo aplica un esquema de

etiquetaje de entonación y 2) no se aplican modelos de lenguaje, como lo hacen Shriberg *et al.* y el proyecto VERBMOBIL.

En el idioma Español, Navarro Tomás ha identificado que la región final del contorno entonativo contiene información de gran utilidad para dar instrucciones al oyente sobre la forma en la cual éste debe interpretar el contenido léxico del enunciado. Navarro Tomás ubica el inicio de esta región en la última sílaba tónica del enunciado, refiriéndola con el nombre de *tonema* y estableciendo una tipología en función de la forma de terminación del contorno, pudiendo ésta ser ascendente, descendente o lineal, con diversos grados de prolongación. En el capítulo se describen también los principales aspectos sobre la relación entre el *tonema* y la modalidad del enunciado.

El **capítulo 3** describe la teoría y la aplicación del esquema DIME-DAMSL (Pineda *et al.*, 2006 y 2006a), un esquema de etiquetación de actos de diálogo basado en DAMSL, de Allen y Core (1997), adaptado a las necesidades del proyecto DIME. La motivación para su desarrollo fue el hecho de que otros esquemas no ofrecen una consistencia satisfactoria en los etiquetajes producidos. Además de DAMSL, otros esquemas precedentes son el del proyecto VERBMOBIL y el de Levin *et al.* (1998). Se revisan los tres con el propósito de contrastarlos con DIME-DAMSL.

Los conceptos principales del esquema DAMSL son: las funciones comunicativas *hacia delante* y *hacia atrás*, el *estado comunicativo* y el *nivel informativo*. Las funciones *hacia delante* están constituidas por actos de diálogo que influyen en el desarrollo futuro de los eventos del diálogo; p. ej. la afirmación, la solicitud de información, el compromiso, etc. Las funciones *hacia atrás* están constituidas por actos que establecen vínculos con enunciados previos, tales como la aceptación, el rechazo, la espera, etc. El estado comunicativo y el nivel informativo aluden, respectivamente, a la inteligibilidad y al tópico tratado por un enunciado. Una limitación de DAMSL es que no toma en cuenta unidades comunicativas de mayor alcance que el enunciado, por lo cual la consistencia de los etiquetajes obtenidos tiende a ser medianamente satisfactoria.

En el esquema de VERBMOBIL se definen actos de diálogo particulares de un dominio altamente especializado: el agendamiento de reuniones de negocios, dentro de un contexto de traducción automática. Una limitación de este esquema es que, a pesar de tratar diálogos prácticos, su grado de especificidad para el dominio hace difícil su aplicación al etiquetado en otros dominios; además, al igual que DAMSL, tampoco considera unidades comunicativas mayores que el enunciado.

El esquema de Levin *et al.* (1998) se basa en DAMSL, teniendo como rasgo destacable la incorporación de estructuras comunicativas de mayor alcance que el enunciado, denominadas *juego de diálogo* (*dialogue game*) y *actividad* (*activity*). Una limitación para los propósitos del presente trabajo es que no fue creado para etiquetar diálogos prácticos, sino para conversaciones telefónicas sobre tópicos generales.

DIME-DAMSL se apoya principalmente en el conjunto de etiquetas y reglas de etiquetaje del esquema DAMSL, incorporando nuevos elementos teóricos y prácticos que permiten mayores niveles de consistencia en los etiquetajes. Los elementos nuevos son: la *transacción*, los *planos de comunicación*, los *cargos* y los *abonos* y las etiquetas para acciones gráficas. Dentro de un diálogo, la *transacción* es un subconjunto de enunciados que constituyen una unidad estructural donde los interlocutores colaboran para lograr una submeta que contribuye al logro de la meta global del diálogo. Los planos de comunicación son una organización de los actos de diálogo que distingue entre el *entendimiento* (*understanding*), el *acuerdo* (*agreement*) y las *obligaciones* (*obligations*), siendo los dos primeros agrupados en el *common ground*. Los planos pueden entenderse como capas superpuestas de la comunicación, sobre cada una de las cuales los diversos tipos de actos de diálogo contribuyen en mayor o menor medida. El plano del *entendimiento* se encontraría en la capa inferior; el del *acuerdo*, en la media y el de las *obligaciones* en la superior. En el del entendimiento contribuyen los actos que administran y reconstruyen el flujo de comunicación; p. ej. las señales de entendimiento o de no entendimiento. En el plano del acuerdo contribuyen los actos que establecen y refuerzan los conocimientos, creencias y presuposiciones compartidos por los interlocutores. Finalmente, en el de las obligaciones, contribuyen los actos que generan en el hablante o en el oyente la responsabilidad de realizar una acción, pudiendo ésta ser verbal o de otra índole. Los *cargos* y *abonos* constituyen un mecanismo que permite equilibrar pares de actos de diálogo, partiendo de un principio de conservación, de modo que un determinado acto de diálogo sobre cierto plano genera un cargo que es compensado por otro acto, generando su abono.

Para evaluar la consistencia de los etiquetajes de actos de diálogo se usa el estadístico Kappa, cuya aplicación en Lingüística Computacional fue propuesta por Carletta (1996), tomándola de las ciencias del comportamiento. Kappa toma en cuenta la posible sobreestimación del valor de la consistencia, a causa del sesgo probabilístico inducido por las interacciones entre el número de etiquetas disponibles, el número de instancias etiquetadas y el número de etiquetadores. Los valores Kappa obtenidos muestran que la consistencia es satisfactoria.

En el **capítulo 4** se aborda la representación de la entonación. Para los propósitos del presente trabajo se requiere que los datos de su etiquetaje sean alimentables a algoritmos de aprendizaje automático generadores de árboles de clasificación y regresión. Se exploran tres enfoques: el del procesamiento de señales, el de la Fonética y el de la Fonología. Se explica por qué se ha elegido el esquema INTSINT (*Internacional Transcription System for Intonation*), de Hirst *et al.* (2000), para la etiquetación de entonación y se describe cómo es adaptado a las necesidades del proyecto en particular.

En el enfoque del procesamiento de señales no se producen etiquetajes de la entonación; sino que ésta se representa mediante parámetros numéricos producidos mediante análisis estadísticos de las series numéricas que describen a la frecuencia fundamental (f_0), la amplitud (energía o intensidad) o la duración de la señal del habla. Por ejemplo, algunos parámetros son: la diferencia entre los valores máximo y mínimo de la f_0 de un

enunciado, la pendiente de la línea de declinación del contorno entonativo, la varianza de la intensidad, etc. Dado que cada hablante puede tener un rango entonativo y de intensidad y una velocidad de habla particulares, esta representación requiere que se realice una normalización estadística de la información para permitir que los datos provenientes de distintos hablantes sean comparables. Una ventaja de este enfoque es su completa independencia de anotador, lo cual le confiere una completa libertad de subjetividades en la interpretación de fenómenos entonativos. Por otra parte, una desventaja es que no incorpora la componente perceptiva en la representación, lo cual introduce el riesgo de producir representaciones cuya fidelidad perceptiva sería cuestionable.

Uno de los esquemas más conocidos para el etiquetaje de la entonación es ToBI, desarrollado desde el enfoque de la Fonología. ToBI, originalmente creado para el idioma inglés y con variantes posteriores para diversos idiomas, permite transcribir la entonación y la estructura prosódica de enunciados. Se basa en las nociones de tono (*H*, alto; *L*, bajo) e índices de disyunción (*break indices*). La versión de ToBI para Español, Sp-ToBI, ofrece una serie de lineamientos para desarrollar el etiquetaje; sin embargo, éstos no garantizan una consistencia de los etiquetajes suficientemente satisfactoria desde la perspectiva del presente trabajo. Otra limitación de Sp-ToBI es la falta de una herramienta de etiquetaje que facilite la tarea.

Bajo el enfoque de la Fonética, un esquema para el etiquetaje de la entonación es INTSINT. Éste se apoya en MOMEL (*Modelisation de Melodie*), de Hirst *et al.* (2000), un algoritmo de estilización de la *f0*, implementado en la herramienta de software MES (*Motif Environment for Speech*), de Espesser (1999). MOMEL produce una estilización de la *f0*, identificando los puntos de inflexión, denominados *targets*, que constituyen tonos relativos a lo largo del enunciado. Posteriormente, el etiquetador humano realiza una verificación perceptiva de la *f0* estilizada para determinar si ésta es perceptivamente fiel a la *f0* original; en caso necesario, el anotador puede modificar la presencia y ubicación de los *targets* estilizados hasta que la representación sea satisfactoria. Entonces, se generan automáticamente los etiquetajes INTSINT con un algoritmo que analiza la ubicación relativa de cada *target* respecto a su antecesor y su sucesor, asignando alguna de las ocho definidas en el *tagset*. Éste está constituido por: *M*, *medium* (medio absoluto); *T*, *top*, (máximo absoluto); *B*, *bottom* (mínimo absoluto); *H*, *higher* (máximo local); *L*, *lower* (mínimo local), *U*, *upstep* (ascendente), *D*, *downstep* (descendente), *S*, *same* (mismo que la etiqueta que le preceda). A cada etiqueta asignada a un *target* se le asocia la marca de tiempo correspondiente a su ubicación temporal sobre el contorno entonativo. Entre las ventajas que ofrece INTSINT está el hecho de que produce la representación entonativa a partir de la señal de habla usando algoritmos del enfoque de procesamiento de señales pero incorporando la componente perceptiva; además, ofrece una herramienta de software para realizar el etiquetaje en una modalidad semi-automática.

Los etiquetajes INTSINT producidos se transforman en una representación especial en la cual se elimina la información de marcas de tiempo con el propósito de lograr un nivel de abstracción más alto, resultando en la concatenación de las etiquetas INTSINT correspondientes al enunciado, lo cual produce una secuencia

alfabética. El incremento en el nivel de abstracción está determinado por el hecho de que se elimina la influencia de las velocidades de habla de los diversos participantes en los diálogos. A partir de la secuencia obtenida se pueden tomar subcadenas para implementar atributos que representan a tonos relativos de las regiones inicial o final del contorno entonativo.

El **capítulo 5** describe la configuración y los resultados de experimentos basados en Aprendizaje Automático para analizar y modelar las interacciones entre variables involucradas en el fenómeno del acto de diálogo. Se describen las variables elegidas con base en teorías sobre la entonación del Español y sobre el acto de diálogo, explicando por qué se usan en particular y por qué se descartaron otras frecuentemente utilizadas en trabajos previos del área. En ese capítulo se presentan los modelos de clasificación y predicción obtenidos.

Para poder implementar un experimento con métodos de Aprendizaje Automático se requiere conocer desde un punto de vista estadístico el fenómeno estudiado. Una vez determinado el atributo *target*; es decir, aquel del cual se desea reconocer o predecir su valor, se debe definir los atributos que se evaluarán como predictores de ese *target*. Por ello, se presenta una serie de análisis estadísticos útiles para comprender la naturaleza del fenómeno y para guiar la implementación de los experimentos. De especial interés son los análisis que muestran la correlación entre ciertas variables, tales como el rol del hablante con la modalidad del enunciado, o la modalidad con el tipo de acto de diálogo, o el rol del hablante con el tipo de acto; se observa, por ejemplo, que algunos actos de diálogo son típicos de uno de los dos roles de hablante, o que ciertos actos de diálogo se expresan más frecuentemente mediante ciertos tipos de modalidad. Estos análisis permiten también evaluar la consistencia de los modelos de clasificación y predicción obtenidos posteriormente. Con base en los análisis estadísticos se implementaron diversos atributos *target* y predictores, como se describe a continuación.

Dado que en el esquema DIME-DAMSL los actos de diálogo se etiquetan sobre dos planos, se implementaron dos atributos representativos, el etiquetaje tomado del plano de *obligaciones* y el del *common ground*.

Los atributos implementados a partir de los etiquetajes de entonación con INTSINT son subcadenas de estos etiquetajes, donde se toman las primeras 1, 2, 3, 4 ó 5 etiquetas, así como las últimas 5, 4, 3, 2 ó 1. Ésto permite representar diversos fragmentos de las regiones inicial o final, respectivamente, del contorno entonativo.

La representación de la duración del enunciado se implementa mediante dos atributos: la duración en milisegundos y el número de tonos INTSINT contenidos en el etiquetaje entonativo; el segundo atributo es una medida indirecta de la duración, ya que el número de tonos depende de ésta pero está libre de la influencia de la velocidad de habla del sujeto.

El rol del hablante es representado mediante el atributo identificado con el mismo nombre, cuyo valor puede ser de *Usuario* o de *Sistema*, dado que los diálogos son orientados a tareas y cada participante desempeña funciones preestablecidas.

Los atributos implementados para representar la modalidad del enunciado son, inicialmente, el etiquetaje manual de la modalidad; después, un etiquetaje automático de la misma generado por un modelo de Aprendizaje Automático que usa etiquetajes de entonación e información de duración del enunciado; finalmente, un tercer atributo se genera también mediante Aprendizaje Automático, usando además de la entonación y la duración, el rol del hablante. Se observa que al usar cuatro etiquetas genéricas de modalidad (*declarativa*, *interrogativa*, *imperativa* y *otra*) se obtienen mejores resultados que al usar etiquetas más especializadas, tales como *interrogativa sí/no*, *interrogativa pronominal*, etc.

Para evaluar la contribución de los diversos atributos al reconocimiento del acto de diálogo, los experimentos de Aprendizaje Automático se implementaron en tres configuraciones distintas usando diversos subconjuntos de los atributos como predictores. En la configuración más simple se evaluaron atributos que describen primordialmente la entonación, tales como los etiquetajes INTSINT, la duración y la modalidad del enunciado; en la segunda configuración, se usaron además de los entonativos, otros no entonativos, como el rol del hablante y el acto de diálogo del plano complementario; en la tercera se usaron, además de los atributos de la segunda, el acto de diálogo perteneciente al enunciado previo. Como resultado de los experimentos, se presenta una serie de árboles de clasificación producidos con el software WEKA (Witten y Frank, 2005), de los cuales se evalúa su precisión (en términos brutos y también mediante el estadístico Kappa para descontar la influencia del azar), computando también, el soporte (*support*) y la confianza (*confidence*) de sus respectivas reglas clasificatorias y los valores de *recall*, *precision* y medida *F* de las categorías clasificadas. De cada árbol se muestran las reglas con mayor capacidad descriptiva del fenómeno, resultando generalmente consistentes con las aseveraciones de las teorías lingüísticas. Los resultados muestran que la tercera configuración genera las mejores tasas de reconocimiento.

El **capítulo 6** presenta las conclusiones, donde se revisan todos los elementos del presente trabajo a la luz de los resultados empíricos obtenidos, se argumenta el valor teórico de los mismos y se esboza una aplicación potencial de los modelos propuestos para el reconocimiento automático del acto de diálogo y de la modalidad del enunciado.

Los resultados sugieren que en tareas de reconocimiento automático del acto de diálogo bajo un enfoque de dos planos (*obligaciones* y *common ground*), la información entonativa contribuye en menor medida que la de otras fuentes, tales como el tipo de acto de diálogo del plano complementario, el acto de diálogo del enunciado previo y el rol del hablante. Una posible explicación de este comportamiento es la presencia de *actos de habla indirectos*, altamente frecuentes en el corpus analizado, lo cual es indicio de un fenómeno común en los diálogos prácticos en general. Sin embargo, la entonación y la modalidad del enunciado

adquieran utilidad como recursos para desambiguar enunciados que tienen un mismo contenido léxico pero expresan actos de diálogo distintos.

Los modelos producidos tienen una utilidad tanto teórica como práctica, ya que permiten describir y explicar las relaciones entre fuentes de información que intervienen en el fenómeno del acto de diálogo y al mismo tiempo pueden servir como herramientas de etiquetación automática, susceptibles de ser incorporadas a sistemas de administración de diálogo basados en parseo del contenido léxico, con lo cual se podría mejorar el desempeño general.

La sección de **Apéndices** contiene tres tipos de información: la descripción del contenido del recurso empírico usado como fuente para desarrollar el presente trabajo; la colección de análisis estadísticos detallados de los etiquetajes y los árboles de decisión creados con algoritmos de Aprendizaje Automático.

Se describe a DIME (Pineda *et al.*, 2007), corpus de diálogos prácticos grabados en video y audio, usado para desarrollar el presente trabajo. Se describen las motivaciones que le dan origen y sus objetivos, así como las condiciones de su grabación. Con base en los objetivos se determinaron sus niveles de etiquetación, enumerados a continuación:

- a) alófonos
- b) sílabas fonéticas
- c) palabras
- d) *break indices* (índices de disyunción) de Sp-ToBI
- e) entonación, con INTSINT
- f) partes del habla, con el esquema de Moreno y Pineda (2006)
- g) marcadores del discurso, también con el esquema de Moreno y Pineda
- h) modalidad del enunciado, bajo un esquema de convenciones propias
- i) actos de diálogo, incluyendo transacciones, cargos-abonos y DIME-DAMSL

En la etiquetación del corpus tienen una gran utilidad el alfabeto fonético computacional MexBet (Cuétara, 2004) y la herramienta de etiquetación semi-automática Transcribemex (Cuétara y Villaseñor, 2004), ambos aplicados en los niveles de alófonos, sílabas fonéticas y palabras.

Se describe el control del proceso de etiquetación y de la calidad de los etiquetajes, presentando también algunas estadísticas generales de los datos producidos.

Los análisis estadísticos permiten una mejor comprensión del fenómeno estudiado y de los lineamientos que guiaron la implementación y evaluación de los experimentos de Aprendizaje Automático. Estos análisis incluyen, por ejemplo, Paretos de los tipos de actos de diálogo y análisis de la correlación entre diversos pares de atributos, tales como la modalidad y el acto de diálogo, entre el acto de diálogo y el rol de hablante, o entre

los actos de diálogo de los diversos planos. Las correlaciones encontradas revelan la capacidad discriminativa de los diversos atributos y permiten explicar su presencia o ausencia en los modelos de reconocimiento y predicción.

Se presentan los árboles de decisión en su forma original, generados con WEKA. Esto permite al lector verificar los resultados. A partir de los árboles originales, se muestran sus correspondientes reglas *if-then* en un formato que facilita su legibilidad; además, al presentarlas ordenadas con base en *soporte* y con base en *confianza*, se facilita su uso, tanto para estudios teóricos del fenómeno, como para su implementación en sistemas de administración de diálogo.

Finalmente, en la sección de **referencias** se enumeran más de cincuenta artículos de investigación, libros, reportes técnicos, herramientas y links de WWW, que permiten profundizar en los fundamentos teóricos, los algoritmos para la modelación y el software para su implementación.

RESUMEN

ESTE TRABAJO ABORDA EL ANALISIS Y LA MODELACION DE LA RELACION ENTRE LA ENTONACION DEL HABLA Y EL TIPO DE INTENCION DEL HABLANTE (ACTO DE DIALOGO). TAMBIEN SE ANALIZA Y MODELA LA CONTRIBUCION DE OTRAS FUENTES DE INFORMACION QUE CONTRIBUYEN AL RECONOCIMIENTO DEL ACTO DE DIALOGO, TALES COMO EL ROL DEL HABLANTE (USUARIO O SISTEMA) Y EL ACTO DE DIALOGO PREVIO. EL OBJETIVO ES AL MISMO TIEMPO TEORICO Y PRACTICO, YA QUE SE INTENTA DESCRIBIR Y COMPRENDER EL FENOMENO, ASI COMO CREAR MODELOS QUE CONDUZCAN A LA IMPLEMENTACION DE SISTEMAS EFICIENTES PARA LA ADMINISTRACION DEL DIALOGO HABLADO.

EL PROBLEMA A RESOLVER SE PLANTEA COMO LA EVALUACION EMPIRICA DE LA CONTRIBUCION DE LA INFORMACION ENTONATIVA DEL HABLA AL RECONOCIMIENTO DE TIPOS DE ACTOS DE DIALOGO USANDO ALGORITMOS DE APRENDIZAJE AUTOMATICO (MACHINE LEARNING) PARA CREAR ARBOLES DE CLASIFICACION CON DATOS DE UN CORPUS DE DIALOGOS PRACTICOS. UN DIALOGO PRACTICO ES AQUEL DONDE LOS INTERLOCUTORES INTENTAN ALCANZAR UNA META COMUN, DESARROLLANDO UNA TAREA EN FORMA COLABORATIVA.

SE APLICA UN ESQUEMA DE ETIQUETACION DE ACTOS DE DIALOGO QUE DEFINE DOS PLANOS DE EXPRESION: LAS OBLIGACIONES Y EL COMMON GROUND, ESTE ULTIMO FORMADO POR DOS SUBPLANOS: DEL ACUERDO Y DEL ENTENDIMIENTO. EL ESQUEMA ESTABLECE QUE UN DIALOGO PRACTICO ESTA ESTRUCTURADO EN TRANSACCIONES, MISMAS QUE SE ENCUENTRAN EN UN ESTADO DE EQUILIBRIO LOGRADO POR LA INTERACCION DE PARES DE ACTOS DE DIALOGO QUE FUNCIONAN CON UN MECANISMO DE CARGOS Y ABONOS INFORMACIONALES.

LOS RESULTADOS SUGIEREN QUE LA ENTONACION CONTRIBUYE AL RECONOCIMIENTO DE LOS ACTOS DE DIALOGO PERO NO ES DETERMINANTE, YA QUE EN EL HABLA ESPONTANEA ES ALTAMENTE FRECUENTE LA PRESENCIA DE ACTOS DE HABLA INDIRECTOS, MEDIANTE LOS CUALES SE EXPRESA UN TIPO DE INTENCION HACIENDO USO DE ENUNCIADOS CUYA MODALIDAD PODRIA CORRESPONDER A OTRO TIPO.

ADEMAS DE LAS CONTRIBUCIONES TEORICAS, ESTE TRABAJO PODRIA OFRECER MODELOS SUSCEPTIBLES DE IMPLEMENTACION EN SISTEMAS DE ADMINISTRACION DE DIALOGO.

1. El acto de diálogo y la entonación

En este capítulo se revisan los fundamentos teóricos para el análisis y la modelación de la relación entre el acto de diálogo y la entonación. Los antecedentes sobre el acto de diálogo se remontan a la noción de *acto de habla*, siendo las teorías de Austin y de Searle las más conocidas. En este contexto se describe la *implicatura conversacional* de Grice. Se revisa el *grounding*, de Clark y Schaefer, debido a que ofrece una teoría sobre la estructura del diálogo a partir de relaciones entre actos de diálogo. Se exploran de modo general algunos esquemas para el etiquetaje de los actos de diálogo. Respecto a la entonación, se revisan algunas ideas desarrolladas acerca de su relación con la modalidad del enunciado, específicamente para el idioma Español. Para este idioma, es mayor la cantidad de trabajos teóricos que estudian la relación entre la modalidad del enunciado y la entonación que la de los dirigidos a la relación entre ésta última y el acto de diálogo.

La teoría de Searle sobre los actos de habla (1969) establece que la producción o emisión de una instancia de un enunciado bajo ciertas condiciones constituye un acto de habla, y los actos de habla son las unidades básicas o mínimas de la comunicación lingüística. La noción de acto de diálogo es una extensión de la anterior e involucra al acto de habla en el contexto de un diálogo (Bunt, 1994) o un acto con estructura interna relacionada específicamente con su función de diálogo, como se asume en Allen y Core (1997), o una combinación del acto de habla y la fuerza semántica de un enunciado (Bunt, 1995). El presente trabajo se apoya principalmente en la noción de Allen y Core pero sintetiza una propia.

La entonación es una de las fuentes de información que los hablantes usan para reconocer el tipo de acto de diálogo en un enunciado. Algunos trabajos teóricos y empíricos (e.g. Wilson y Sperber, 1988) han sugerido que la modalidad del enunciado y por lo tanto la entonación proporcionan instrucciones al oyente sobre cómo procesar el contenido léxico; dichas instrucciones son muy útiles en tareas de desambiguación cuando el tipo de acto de diálogo no puede reconocerse solamente analizando el contenido léxico. Los análisis de la relación entre entonación y acto de diálogo podrían permitir la creación de modelos para mejorar el desempeño de los sistemas de administración de diálogo. La entonación también está relacionada con el tipo de modalidad de enunciado, p. ej. interrogativa, declarativa, imperativa, etc. La modalidad del enunciado es una caracterización de los patrones entonativos y sintácticos en el habla y podría proporcionar un medio para expresar el acto de diálogo; por ello se hipotetiza que el reconocimiento del acto de diálogo puede apoyarse en el reconocimiento de la modalidad del enunciado.

1.1 El acto de habla

Un antecedente del término *acto de diálogo* es el *acto de habla* (*speech act*), introducido por Austin (1962), quien afirma que un enunciado en un diálogo es en cierta forma una acción realizada por el hablante y por ello la denomina *acto de habla*. Según su visión, un enunciado conlleva tres tipos de actos: *locutivo*, *elocutivo* y *perlocutivo*; el primero es la propia emisión de la oración; el segundo involucra la intención del hablante, p. ej. preguntar, responder, prometer, etc. y el tercero es la producción, generalmente intencional, de algún efecto en el oyente. En otras palabras, un enunciado tiene una forma, una intención y un efecto determinados.

Searle afirma que hablar un lenguaje es participar en una forma de conducta altamente compleja, gobernada por reglas. Otra de sus ideas es que el acto de habla constituye la unidad mínima de comunicación lingüística, realizándose en el momento de la producción o emisión de un símbolo, palabra u oración. Usa el término *acto de habla* para referirse específicamente al denominado *acto elocutivo* de Austin; es decir, atañe primordialmente a la intención del hablante. Para Searle, el acto de habla es una función de la intención; pero esta función no constituye una relación biunívoca, ya que un mismo enunciado podría expresar diferentes actos de diálogo al darse en diferentes contextos.

En el análisis realizado por Searle, se tiene que: un hablante H expresa la intención $i-I$ mediante el enunciado E creado a partir de la oración S para producir en el oyente O el conocimiento (reconocimiento, conciencia) de que existe el estado de asuntos especificados mediante S . Searle denomina a esto *efecto elocutivo*, IE . H intenta que E produzca IE mediante el reconocimiento de $i-I$ por parte de O . H intenta que $i-I$ sea reconocida mediante el uso del conocimiento que O tiene de ciertas reglas que gobiernan a los elementos constituyentes de E . En este análisis es importante distinguir el uso que Searle da a los términos *oración* (*sentence*) y *enunciado* (*utterance*), ya que representan conceptos diferentes, pues éste último es una instancia de la realización verbal de una oración. La *oración* tiene propiedades independientes del contexto; es decir, una intención o significado invariante en virtud del significado de las palabras que la constituyen y de la forma en que están organizadas. Por su parte, el *enunciado* es una entidad concreta cuyo significado depende del contexto en que se expresa. Por ello, una oración podría ser expresada por cualquiera de los enunciados pertenecientes a un conjunto determinado por las reglas del lenguaje. El análisis de Searle se apoya en el hecho de que ambos interlocutores conocen y aceptan de antemano las reglas que determinan el uso del lenguaje, las cuales guían al hablante en la definición del enunciado que usa como medio para expresar una oración y al oyente en su respectiva interpretación.

Searle presenta una taxonomía de actos de habla en cinco clases, que se enumeran a continuación:

- a) *Asertivas*: acciones mediante las cuales el hablante expresa una postura comprometida con una idea, como al sugerir, proponer, jurar, alardear, obtener una conclusión, etc.
- b) *Directivas*: el hablante intenta que el oyente realice una acción, como al preguntar, dar un orden, solicitar, invitar, aconsejar, suplicar, etc.

- c) *Comisivas*: el hablante se compromete a realizar una acción en el futuro; p. ej. al prometer, planear, tomar votos, apostar, oponerse
- d) *Expresivas*: el hablante comunica cómo se siente acerca de un estado de asuntos; p. ej. agradecer, disculparse, dar la bienvenida, deplorar
- e) *Declaraciones*: introducen un estado diferente del mundo por el simple hecho de emitir el enunciado; éstas incluyen muchas de las performativas de Austin; p. ej. nombrar (dar nombre por primera vez a una instancia de una entidad), renunciar (a un empleo), despedir (de un empleo)

Respecto a las reglas que gobiernan al diálogo, Grice (1975) presenta la teoría denominada *implicatura conversacional* (*conversational implicature*), estableciendo que la interpretación de un enunciado se apoya en más información que sólo el significado literal de la oración. Esta información es obtenida con base en el denominado *principio de cooperatividad* y en una serie de *máximas* heurísticas. En general, se trata de un conjunto de suposiciones en el hablante y el oyente; por ejemplo, cuando el hablante dice algo asume que el oyente puede hacer una serie de inferencias para interpretar el mensaje en la forma correcta. A su vez, el oyente asume que el hablante está conversando guiado por ciertas reglas. Grice introdujo cuatro máximas heurísticas que se utilizan cuando se observa el *principio de cooperatividad*:

- a) *Máxima de cantidad*: se supone que la cantidad de información en un enunciado no es ni más ni menos informativa que la necesaria
- b) *Máxima de calidad*: se supone que lo que se dice es algo verdadero y que está disponible una cantidad suficiente de evidencia
- c) *Máxima de relación (o relevancia)*: se espera que la contribución del hablante se relacione y sea adecuada para satisfacer las necesidades de información del oyente
- d) *Máxima de manera*: se espera que la contribución sea clara, breve, ordenada y libre de ambigüedad

El *principio de cooperatividad* (*cooperative principle*) involucra que los dos interlocutores desean comunicarse en una forma eficiente y que aceptan cumplir un conjunto de reglas para lograr un propósito común.

En el análisis realizado por Grice se tiene que: el hablante *H* tiene la intención (*i-I*) de que el enunciado *E*, como instancia de la oración *X*, produzca ciertos efectos perlocutivos *PE* en el oyente *O*; así que *H* intenta que *E* produzca *PE* mediante el reconocimiento que *O* hace de *i-I*. Grice aborda el estudio del fenómeno considerando, además de la intención (*i-I*) del hablante, los efectos que el enunciado produce en el oyente. Hay un punto en común con el análisis planteado por Searle, en el sentido de que ambos consideran al enunciado como una instancia particular de una oración. Sin embargo, una diferencia destacable es el uso de la terminología para referir al *acto de habla*, ya que Searle lo refiere con el mismo término (*speech act*), mientras que Grice lo denomina *intención* (*meaning*). En opinión de Searle (1969, p. 62), Grice no hace una

distinción clara entre actos elocutivos y perlocutivos, por lo cual tampoco distingue claramente entre la intención del hablante y el efecto producido en el oyente.

La postura de la presente investigación respecto a las nociones de Grice y Searle se plantea en los siguientes términos: el acto de habla es abordado dentro del ámbito de los diálogos prácticos, en los cuales el análisis de un enunciado emitido por un hablante S_1 es insuficiente para la caracterización del tipo de acto, ya que se requiere observar también el efecto inducido en el oyente S_2 y además la reacción de S_1 ante estos últimos. El efecto producido en S_2 es verificable mediante sus acciones verbales (o de otra índole) y estos permiten a S_1 comprobar si su intención fue entendida plenamente. La conducta (verbal o de otra índole) observable en S_1 permite determinar si éste considera que su intención fue entendida. Por lo anterior, podría decirse que en el presente trabajo hay una mayor cercanía a la noción de Grice, porque se toma en cuenta el efecto producido en el oyente; pero una diferencia importante es que además se analiza la reacción del emisor inicial ante ese efecto.

1.1.2 El acto de diálogo

La noción de acto de habla ha sido útil para tipificar ciertos elementos de pragmática desde las perspectivas de la lingüística teórica y la filosofía del lenguaje. Sin embargo, la lingüística computacional requiere de una noción que pueda dar soporte al estudio empírico del fenómeno, incluyendo su análisis estadístico y su modelación computacional. Estos requerimientos no pueden ser satisfechos por el marco teórico del concepto de acto de habla; por ello, la propia lingüística computacional ha desarrollado el concepto de acto de diálogo.

Jurafsky (2001) afirma que en trabajos previos del área el término *acto de diálogo* se usa en forma ambigua, por lo cual no son claras las diferencias con el acto de habla y eventualmente llegan a confundirse. Algunas acepciones son, por ejemplo, la de Bunt (1994), quien entiende al acto de diálogo como un acto de habla en el contexto de un diálogo; o la del mismo Bunt (1995), para referirse a una combinación del acto de habla y la fuerza semántica de un enunciado; o la de Allen y Core (1997), quienes lo consideran un acto con estructura interna relacionado específicamente con su función de diálogo. Un término similar al acto de diálogo es la *movida conversacional*, abordada por Power (1979). Con el paso del tiempo, en el campo de la lingüística computacional parece prevalecer el término acto de diálogo sobre el de acto de habla y uno de los enfoques más aceptados es el de Allen y Core.

1.1.3 El *grounding* o aterrizamiento

El *grounding* o *aterrizamiento*, noción difundida por Clark y Schaefer (1989), consiste en incrementar el *common ground* (*aterrizamiento común*) de los participantes del diálogo en una forma ordenada. Stalnaker

(1978) refiere al *common ground* como un conjunto de presuposiciones compartidas entre el hablante y el oyente; dichas presuposiciones son proposiciones cuya verdad el hablante da por hecho. El hablante asume que su interlocutor (o su audiencia) presuponen todo lo que él presupone.

Clark y Schaefer se apoyan en Stalnaker e introducen nuevos conceptos respecto al grounding: la *contribución*, las fases de *presentación* y de *aceptación* y el *incrustamiento* (*embedding*). La contribución es una unidad de discurso creada por los interlocutores al intentar aterrizar (*ground*) lo que se dice; el aterrizamiento consiste en alcanzar la creencia mutua de que el oyente ha entendido suficientemente bien lo dicho por el hablante. La contribución difiere de la mayoría de las unidades lingüísticas típicamente consideradas por el hecho de que no se formula autónomamente por el hablante, sino que se construye colectivamente por los participantes del diálogo con base en sus acciones coordinadas.

La contribución está constituida por dos fases: la de *presentación* y la de *aceptación*, descritas a continuación:

En la *fase de presentación*, el hablante *A* presenta el enunciado *U* para que *B* lo considere. *A* supone que, si *B* ofrece cierta evidencia mínima *E* o mayor, *A* puede creer que *B* entendió lo que *A* quería decir mediante *U*.

En la *fase de aceptación*, *B* acepta el enunciado *U* al ofrecer la evidencia *E'* de que cree que entiende lo que *A* quería decir mediante *U*. Lo hace suponiendo que, una vez que *A* registre la evidencia *E'*, él también creerá que *B* entiende. Esta fase es inherentemente recursiva porque a cada aceptación se tiene que dar un acuse de recibo (*acknowledgment*) por el interlocutor; la recursión se detiene cuando los hablantes obtienen la mínima evidencia de entendimiento por parte de cada uno, usualmente constituida por la atención continuada.

La forma de contribución que los interlocutores prefieren consiste en expresar uno o más enunciados. La complejidad de las contribuciones puede conducir a una variedad de formas, una de las cuales es la contribución *incrustada* (*embedded*); así que la incrustación (*embedding*) consiste en la presencia de una o más contribuciones dentro de otra contribución de mayor jerarquía. La contribución incrustada también se denomina *installment*.

La incrustación puede ocurrir debido a que, en la fase de presentación, las instrucciones difíciles pueden necesitar ser divididas en *installments*, cada una con sus propias fases de presentación y aceptación y, en la fase de aceptación puede requerirse un proceso recursivo debido a problemas de entendimiento.

1.1.4 Esquemas de etiquetaje de actos de diálogo

Debido a los objetivos tanto teóricos como ingenieriles de la inteligencia artificial, el acto de diálogo ha sido analizado y modelado con base en datos tomados de *corpora* de diálogo; por ello, ha sido necesario crear

esquemas de etiquetaje de este fenómeno. Un esquema de etiquetaje es una taxonomía a la cual se asocia un conjunto de etiquetas y reglas para asignarlas a enunciados, generalmente tomados de un corpus. Entre estos esquemas, DAMSL (*Dialogue Act Markup in Several Layers* ó *Marcaje de Actos de Diálogo en Varias Capas*), de Allen y Core (1997), es uno de los primeros, más conocidos y aplicados; por ello, merece una especial atención por parte del presente trabajo.

El esquema DAMSL

DAMSL es un esquema para etiquetar actos de diálogo en diálogos prácticos. Éste incorpora algunos elementos de la teoría de Searle y de otras teorías provenientes de la filosofía del lenguaje, enriqueciéndolos para satisfacer necesidades de la modelación computacional. La mayoría de las teorías que le anteceden intentan capturar el propósito (o propósitos) de un enunciado al asignarle una sólo etiqueta elegida de un universo de etiquetas posibles. Esto ha sido identificado como un problema importante por parte de Allwood (1995), Cohen y Levesque (1990) y Hancher (1979), porque un enunciado puede involucrar varios propósitos al mismo tiempo, de modo que requeriría más de una etiqueta para caracterizar su acto de diálogo. El esquema DAMSL aborda el problema definiendo *varias capas (several layers)* de etiquetación, con lo cual se dispone de una amplia flexibilidad al caracterizar los diversos propósitos de un enunciado. Desde su perspectiva, el acto de diálogo se concibe como un flujo de información compuesto por diversas componentes. DAMSL define una taxonomía de los actos agrupándolos en dos conjuntos: las *funciones hacia delante* y las *funciones hacia atrás*. Las primeras corresponden, en general, con las diversas categorías de la teoría tradicional de actos de habla; por ejemplo, las *directivas de acción*, los *compromisos* o las *afirmaciones* en DAMSL corresponden con las *directivas*, las *comisivas* o las *representativas*, respectivamente, en el esquema de Searle. La denominación de *funciones hacia delante* se usa como opuesta a la de las *funciones hacia atrás*, que son los actos que expresan una reacción del hablante respecto a algún enunciado previo; p. ej. aceptar una propuesta, confirmar entendimiento, responder una pregunta, etc. Estas dos funciones son, hasta cierto punto, comparables con conceptos tales como los *pares adyacentes* de Schegloff (1968) y la *contribución* de Clark y Schaefer (1989).

Cada una de las dos funciones de DAMSL se divide en dimensiones; las *funciones hacia delante* se dividen en: *declarativa*, *solicitud de información*, *influencia sobre el oyente* e *influencia sobre el hablante*. Las *funciones hacia atrás* contienen, principalmente, a las dimensiones de *acuerdo*, *entendimiento* y *respuesta*. Un enunciado puede expresar actos correspondientes a las funciones hacia delante, o bien, a las funciones hacia atrás, a ambas o incluso a ninguna; un ejemplo del tercer caso sería una *afirmación* que al mismo tiempo fuera *respuesta* o, respecto al cuarto caso, un *monólogo* dentro de un diálogo. Dentro de cada función, un enunciado expresa un acto en alguna de las dimensiones de la función. La mayoría de las dimensiones contiene, a su vez, un conjunto de etiquetas de acto de diálogo que permiten el manejo de una granularidad fina; por ejemplo, la influencia sobre el oyente puede ser del tipo *directiva de acción* u *opción abierta*, una

declarativa puede ser *afirmación* o *reafirmación*, la influencia sobre el hablante puede ser *oferta* o *compromiso*, etc. Dentro de cada dimensión, sus correspondientes actos de diálogo son mutuamente excluyentes; p. ej. en la dimensión de influencia sobre el hablante, el acto expresado podría ser una *oferta* o un *compromiso*, pero no ambos.

DAMSL añade el *estado comunicativo* y el *nivel informativo*. El primero permite distinguir entre enunciados que son etiquetables, o no, desde el punto de vista de las dos funciones comunicativas; algunos tipos del estado comunicativo son *monólogo* e *ininteligible*, casos en los cuales no habría acto de diálogo que etiquetar porque el enunciado estaría dirigido hacia el mismo hablante sin la intención de comunicar, o bien, el enunciado sería inaudible o incomprensible. El nivel informativo indica el tipo de asunto que el enunciado trata; p. ej. la realización de la tarea, la administración de la comunicación, la administración de la tarea o combinaciones de éstos.

El esquema DIME-DAMSL

DIME-DAMSL (Pineda *et al.*, 2006 y 2006a) se apoya fuertemente en DAMSL, extendiéndolo mediante la identificación de estructuras dentro del diálogo que tienen un alcance mayor que el del enunciado. Dichas estructuras son:

- a) Los *planos de expresión: obligaciones* y *common ground*, éste último dividido en los subplanos de *acuerdo* y de *entendimiento*
- b) La *transacción*
- c) Los *cargos* y *abonos* en estado de equilibrio

Al identificar estas estructuras, la interpretación de los actos de diálogo adquiere una mayor consistencia, ya que se dispone de más fuentes de información para realizar la tarea.

DIME-DAMSL asume que los actos de diálogo ocurren en dos *planos de expresión: obligaciones* y *common ground*. Las *obligaciones* son los actos que generan en el hablante o en el oyente la responsabilidad de realizar una acción, pudiendo ésta ser verbal o de otra índole. El *common ground* son los actos que establecen, refuerzan o reconstruyen el flujo comunicativo y los conocimientos compartidos por los interlocutores. Cada tipo de acto de diálogo puede tener componentes en cualquier plano o aún en ambos; sin embargo, algunos tipos específicos contribuyen a alguno de los dos planos más que al otro. Por ejemplo, las directivas de acción contribuyen al plano de las *obligaciones* en mayor grado que al del *common ground*; las aceptaciones contribuyen más al *common ground* más que a las *obligaciones*, y los acuses de recibo (*acknowledgments*) contribuyen al *common ground* exclusivamente.

Cada plano incluye una tipología de actos de diálogo que, en lo general, es similar a la de DAMSL. De este modo, el plano de *obligaciones* contiene la componente más relevante de las directivas de acción, las solicitudes de información, los compromisos y las ofertas. Respecto al *common ground*, en el subplano del *acuerdo* los principales contribuyentes son las aceptaciones, los rechazos, las aceptaciones parciales, los rechazos parciales, los quizás y las puestas en espera; en el subplano del *entendimiento*, los contribuyentes primordiales son los acuses de recibo (*acknowledgments*), los *back-channels* (acuses de recibo simultáneos al enunciado del interlocutor), las repeticiones o refraseos y las señales de no entendimiento.

La *transacción* es un conjunto de actos de diálogo en los cuales los interlocutores colaboran para lograr una meta (o submeta) particular. Se caracteriza en términos de una fase de *especificación de la intención*, seguida por una fase de *satisfacción de la intención*; esto es comparable hasta cierto punto con el denominado *árbol de contribuciones* de Clark y Schaefer. La fase de *especificación* presenta, a su vez, una subfase de *presentación* y otra de *interpretación*; la fase de *satisfacción* se estructura en las subfases de *realización* de la acción e *interpretación* de la misma. Cada una de las subfases admite la presencia de procesos recursivos. Estos procesos son tomados en cuenta por DIME-DAMSL, aunque para propósitos prácticos ha incorporado la convención de que las transacciones se etiqueten sin considerar anidamientos ni traslapes, de modo que cada una sea manejada como autocontenida dentro de los límites de la tarea que especifica y satisface una sola instancia de una directiva de acción sobre una sola instancia de un tipo de objeto.

El sistema de *cargos* y *abonos* de DIME-DAMSL involucra la existencia de ciertos pares de actos de diálogo que generan un estado de equilibrio. Los respectivos actos pueden corresponder a enunciados consecutivos, o no, dentro del diálogo. Dentro del par, primero ocurre el acto correspondiente al cargo, generándose así un desequilibrio, que es equilibrado posteriormente por el segundo elemento del par, denominado abono. El equilibrio de los actos se basa en la existencia de ciertas reglas de la conversación, previamente aprendidas e implícitamente aceptadas por los interlocutores, que los guían y, hasta cierto punto, los fuerzan, a elegir un determinado tipo de acto entre un conjunto finito de éstos cuando se atiende a un acto previamente expresado por el interlocutor. Si el acto elegido *no* perteneciera al conjunto finito esperado por el interlocutor, éste último supondría que hubo un problema con la comunicación, verificaría su suposición y, eventualmente, reconstruiría el flujo comunicativo.

DIME-DAMSL identifica para cada plano comunicativo los pares de tipos de actos que producen estados equilibrados. Por ejemplo, en el de las *obligaciones*, un *cargo* producido por una *solicitud de información* se abona con una *respuesta*; en el plano del *acuerdo*, un par típico es la *afirmación*, como cargo, con la *aceptación*, como abono; finalmente, en el del *entendimiento*, un par es el del *acuse de recibo* (*acknowledgment*) para abonar el cargo generado por el enunciado previo del otro hablante. Los enunciados correspondientes a los actos de un par equilibrado pueden ser inmediatamente contiguos, o bien, estar separados por un número indefinido de enunciados.

DIME-DAMSL incorpora un subconjunto de etiquetas para anotar acciones gráficas. Esto es útil para el etiquetaje de diálogos multimodales; es decir, aquellos en los cuales se presentan eventos comunicativos de naturaleza no verbal, tales como los eventos gráficos en una interfaz de C.A.D.

Con el propósito de interpretar el acto de diálogo preponderante de un enunciado, DIME-DAMSL establece convenciones de etiquetación que describen la relación entre los tipos particulares de actos de diálogo y las estructuras de los planos comunicativos, la transacción y los cargos y abonos. Estas convenciones tienden a reducir el número de etiquetas posibles que un acto de diálogo puede recibir, lo cual favorece el acuerdo entre etiquetadores. Esto robustece la consistencia interna del modelo, aspecto importante al producir etiquetajes que se usen en análisis y modelaciones con enfoque empírico.

1.1.5 Concepto de acto de habla en el presente trabajo

Con base en las teorías sobre el acto de habla y en los esquemas de etiquetaje estudiados, se sintetiza un concepto propio de acto de habla para los propósitos particulares del presente trabajo. Más que aseverar una definición delimitada rígidamente, se enumeran a continuación algunos elementos que constituyen al concepto sintetizado, que resulta similar al de DIME-DAMSL.

El acto de habla es una forma de acción de los participantes de un diálogo, cuyo desarrollo está gobernado por reglas. Se aborda bajo la visión del acto de diálogo, es decir, como acto de habla dentro del contexto del diálogo y, particularmente, del diálogo colaborativo orientado a tareas, denominado *diálogo práctico*; por ello, en este trabajo se usa preferentemente el término *acto de diálogo* y no acto de habla.

Dada la menor complejidad del diálogo práctico, la cantidad de tipos de actos de diálogo es menor que en el diálogo de propósito general

El acto de diálogo se expresa mediante tres planos de comunicación que presentan una jerarquía de capas superpuestas: en la capa inferior, el plano del *entendimiento*; sobre ésta, el del *acuerdo* y en la superior, el de las *obligaciones*. El plano del *entendimiento* conduce la componente del acto de diálogo que establece, refuerza o reconstruye el flujo comunicativo. El plano del *acuerdo* conduce la componente que establece o refuerza los conocimientos, las creencias y las presuposiciones compartidos por los interlocutores. El plano de las *obligaciones* conduce la componente del acto que genera una responsabilidad, ya sea en el hablante o el oyente, de realizar una acción, pudiendo ésta ser verbal o de otra índole. El orden de los planos se debe a que, para que se pueda crear una obligación en cualquiera de los interlocutores, primero debe lograrse un acuerdo entre ambos y éste necesita un mínimo grado de entendimiento.

Cada uno de los tres planos recibe contribuciones en diferente medida de cada tipo de acto de diálogo. Enseguida se enumeran algunos ejemplos de los principales contribuyentes de cada plano. En el del entendimiento, las señales de entendimiento, tales como el *acknowledgment* y el *back-channel* y la *señal de no entendimiento*. En el plano del acuerdo, la afirmación, la aceptación o rechazo, la puesta en espera, etc. Finalmente, en el plano de las obligaciones, la *directiva de acción*, el *compromiso*, la *solicitud de información* y la *oferta*.

En cada uno de los planos, ciertos pares de actos de diálogo mantienen una relación de equilibrio, donde un elemento del par introduce un desequilibrio que necesita compensación por parte del otro elemento del par. Esta relación de equilibrio funciona como un mecanismo que permite a los interlocutores ejercer un seguimiento y control del diálogo.

La *transacción* está constituida por una serie de actos de diálogo equilibrados que tienen una meta (o submeta) común dentro de un diálogo. La transacción presenta dos fases: *especificación* de una intención y *satisfacción* de la misma. En general, la *especificación* involucra la expresión de una directiva de acción aplicada sobre una instancia particular de un objeto dentro de un contexto espacial; la *satisfacción* consiste en la realización de la acción solicitada. El reconocimiento de la fase en la cual ocurre un acto de diálogo contribuye a la detección e interpretación del tipo de acto. En DIME-DAMSL se considera que en la transacción pueden presentarse transacciones anidadas; es decir, transacciones de menor jerarquía mediante las cuales se va satisfaciendo progresivamente una intención compleja. En este caso la transacción contiene varias directivas simples que pueden aplicarse a una sólo instancia de objeto o a varias. Sin embargo, para la vertiente ingenieril del presente trabajo, en el etiquetaje de inicio y fin de transacciones se ha asumido la convención de que cada transacción es autocontenida y de que no existen transacciones anidadas, siendo su ámbito delimitado por el alcance dentro del cual se desarrolla un tipo particular de directiva de acción sobre un tipo de objeto específico.

La interpretación de un acto de diálogo por parte del oyente requiere información adicional al contenido léxico del enunciado; p. ej. la entonación, el conocimiento y las creencias compartidos por los interlocutores, los actos de diálogo que anteceden al enunciado interpretado, etc. Desde una perspectiva de tres planos comunicativos, la componente de un acto sobre alguno de los tres da indicios al oyente sobre las componentes que existen en los otros dos.

En un entorno multimodal (visual y auditivo), una acción mediante la cual se muestre información gráfica tiene una naturaleza más similar a la de la afirmación que a la de las acciones físicas que modifican el ambiente virtual. Esto se debe a que su propósito principal es dar información al solicitante. Algunos de los actos de diálogo expresados en modalidad visual son, por ejemplo, agregar un objeto al espacio virtual, reubicarlo o borrarlo.

1.2 La entonación

En la mayoría de los idiomas la entonación del habla transmite, voluntaria o involuntariamente, información que el oyente puede usar para conocer la intención de un hablante; incluso, puede obtener indicios sobre su estado emocional, su nivel sociocultural, su estado de salud, etc. Por ello, la entonación no es sólo un complemento del contenido léxico del enunciado, sino que constituye una fuente de información en sí misma. De ahí que el estudio de la entonación se haya abordado desde diversos enfoques, tales como el análisis de las emociones, el estudio de los factores socioculturales, etc. Wilson y Sperber (1993), en la denominada *Teoría de la Relevancia*, afirman que se puede esperar que un enunciado codifique dos tipos básicos de información: *representacional (conceptual)* y *computacional (procedural)*; es decir, información sobre las representaciones que serán manipuladas, e información sobre cómo manipularlas. Adicionalmente, introducen el denominado *principio de relevancia* (1986), según el cual el hablante espera *no* causar al oyente una cantidad injustificable de esfuerzo de procesamiento.

Entre las diversas componentes informativas de la entonación, una de las más importantes es la que proporciona al oyente instrucciones acerca del uso que debe dar al contenido léxico. Esta componente es la que atañe al presente trabajo, ya que en diversas teorías se le asocia con la modalidad del enunciado y el acto de habla.

Algunas de las principales corrientes que estudian la entonación son:

- a) la Escuela Británica
- b) la Escuela Americana
- c) la Escuela Holandesa
- d) el modelo de Aix-en-Provence

La Escuela Británica, denominada también *del análisis por configuraciones*, organiza la frase entonativa en una serie de componentes: el *núcleo*, la *cola (tail)*, la *cabeza (head)*, y la *pre-cabeza (pre-head)*, considerando tres tonos nucleares: *ascendente*, *descendente* y *fijo*.

La Escuela Americana, denominada también *del análisis por niveles*, considera que los contornos tonales *no* se dividen estructuralmente en partes distintas; el contorno se representa mediante una serie de niveles tonales (*bajo*, *medio*, *alto*). Identifica a la *juntura terminal* como elemento que caracteriza a la dirección tonal de una frontera. Esta corriente es precursora de las teorías más populares en la actualidad, tales como el modelo AM (*Métrico y Autosegmental*) y ToBI (*Tone and Break Indices*).

La Escuela Holandesa es creadora del modelo IPO (*Institute for Perception Research*), basado en el análisis fonético de los contornos melódicos. Con un enfoque similar al de esta escuela, el modelo de Aix-en-Provence, mediante el esquema de etiquetaje INTSINT (*Internacional Transcription System for Intonation*), de Hirst *et al.*, (2000), ofrece la valiosa posibilidad de analizar y sintetizar contornos entonativos mediante algoritmos y herramientas computacionales que permiten una verificación perceptiva de la representación; por ello, este modelo es de especial interés para el presente trabajo.

Una referencia muy útil es la compilación de Prieto (2003), titulada *Teorías de la Entonación*, donde expertos de diversas escuelas describen sus elementos fundamentales desde los puntos de vista del Español y el Catalán.

Sobre el estudio de la entonación del español, son clásicos los trabajos de Navarro Tomás (1948) y los de Quilis (1981). En ambos, el fenómeno se analiza primordialmente en forma teórica, debido a la nula o incipiente disponibilidad de herramientas computacionales.

En la década de 1990, Sosa (1991 y 1999) propone el sistema Sp-ToBI, una adaptación de ToBI al análisis de la entonación española y enfatiza que hasta ese momento no existía un acuerdo entre la comunidad académica acerca de los criterios de análisis y los esquemas de etiquetación. Sin este acuerdo, es prácticamente imposible establecer un modelo que garantice una etiquetación satisfactoriamente consistente. Este aspecto de Sp-ToBI podría considerarse una deficiencia importante si el objetivo es crear modelos empíricos, particularmente basados en métodos de aprendizaje automático.

1.2.1 Relación entre el acto de diálogo y la entonación

En diversos idiomas parece existir cierta preferencia de los hablantes por ciertos tipos de modalidad de enunciado para expresar determinados actos de diálogo. La modalidad es una caracterización basada en el patrón del contorno entonativo y en la forma sintáctica; por ello, puede sugerirse la existencia de una relación entre el tipo de acto y la entonación del enunciado a través de la modalidad. En este contexto, uno de los fenómenos más conocidos es, por ejemplo, la preferencia por la modalidad interrogativa al expresar solicitudes de información. Las investigaciones empíricas sobre la relación *entonación - acto de diálogo* son más numerosas en idiomas como el inglés y poco desarrolladas en Español, donde los análisis se han orientado más hacia la relación *entonación - modalidad*; por ello, se considera valioso profundizar los estudios para este idioma.

1.3 Niveles de representación lingüística para el reconocimiento del acto de diálogo

Diversas teorías lingüísticas han analizado la contribución de diversos niveles lingüísticos de representación de los enunciados con el propósito de detectar e interpretar los actos de diálogo. Los lingüistas, Navarro Tomás entre ellos, han establecido que el contorno entonativo de los enunciados en Español, especialmente su región final, denominada *tonema*, está estrechamente relacionada con la modalidad del enunciado, misma que lo caracteriza como declarativo, interrogativo, imperativo, etc. Diversos estudios teóricos y empíricos en el área parecen sugerir que la entonación podría estar relacionada con el acto de diálogo a través de la modalidad. Además, otros han argüido que la duración y la localización de las sílabas tónicas y de sus respectivos fonemas vocálicos también son fuentes informativas vinculadas al acto de diálogo. También la cantidad de energía (intensidad) y la presencia y duración de pausas dentro del enunciado contribuyen a la detección e interpretación.

Se necesitan análisis y resultados basados en datos empíricos para producir modelos útiles en sistemas computacionales con viabilidad aplicativa. Diversos trabajos empíricos previos en ciencia computacional, como (Shriberg et al., 1998), muestran que la detección e interpretación de actos de diálogo está basada en dos fuentes de información principales: contenido léxico y entonación, por lo cual sus resultados sugerirían que los niveles lingüísticos relevantes de la representación son los que involucran información léxica y entonativa. La información léxica está constituida por las palabras en la señal del habla. La entonación involucra parámetros fonéticos y fonológicos y está determinada principalmente por la frecuencia fundamental (f_0), la intensidad (energía) y la posición y duración de las pausas y de las sílabas tónicas.

Los fenómenos fonológicos están determinados por interacciones entre diversos elementos de información del habla, tales como las unidades léxicas del enunciado, la forma del contorno entonativo, la duración y la intensidad de los fonemas y de las sílabas. Es de esperar que algunos de los diversos niveles de la representación de la información lingüística del enunciado vinculadas con el acto de diálogo puedan estar estadísticamente correlacionados; es decir, que exista un cierto grado de redundancia entre las fuentes informativas. La expectativa surge de la observación de que los oyentes logran reconocer el tipo de acto de diálogo correspondiente a un enunciado a pesar de que algunas unidades léxicas sean ininteligibles. Además de ser relevante dentro de la teoría, este aspecto también lo es para las posibles implementaciones prácticas.

Otro fenómeno a etiquetar está constituido por los índices de ruptura (*break indices*), que permite representar las pausas dentro del enunciado. La presencia de pausas está relacionada con ciertos patrones entonativos, así que podría esperarse alguna relación entre los índices de ruptura y las etiquetaciones INTSINT. Entre los diversos tipos de pausa, uno en particular parece caracterizar a las denominadas *enumeraciones*, expresadas dentro de algunas afirmaciones o en opciones abiertas. Una enumeración es un enunciado que enlista una serie de elementos consecutivamente; es relativamente común en el corpus DIME. Típicamente, en este corpus las enumeraciones ocurren en las opciones abiertas expresadas por el sistema, así que la presencia de

una o más pausas de enumeración en un enunciado podría ser indicio de que el acto de diálogo correspondiente es una opción abierta.

El hallazgo empírico de correlaciones entre los diversos niveles de representación lingüística vinculadas al acto de diálogo en el presente trabajo puede constituir una evidencia de consistencia con diversas teorías lingüísticas y, además, podría guiar la implementación de sistemas de uso práctico.

1.3.1 Niveles de representación útiles para la modelación computacional del acto de diálogo

Un modelo computacional para la modelación del acto de diálogo debería intentar beneficiarse de todos los niveles lingüísticos posibles que contribuyan al incremento de la capacidad de reconocimiento. Una posible restricción a considerar al usar los diversos niveles sería la impuesta por la redundancia de los atributos predictores; es decir, se requiere detectar y evaluar las posibles correlaciones estadísticas entre los niveles representacionales antes de ser incluidos en un modelo reconocedor. El propósito de esta evaluación es descartar aquellos atributos que no contribuyan (o que afecten negativamente) a la tarea de reconocimiento. Los algoritmos de aprendizaje automático, tales como CART (Breiman *et al.*, 1983) y J48 (Witten y Frank, 2005) pueden seleccionar y jerarquizar automáticamente los atributos predictores de acuerdo a su respectivo poder clasificatorio. En otras palabras, el atributo con el mayor poder clasificatorio es señalado como el primero que debe ser analizado cuando se use el modelo en una aplicación del mundo real; después el segundo atributo con mayor poder clasificatorio, luego el tercero y así sucesivamente. Esta jerarquización es valiosa por, al menos, dos razones: primero, porque la jerarquía de los atributos representa indirectamente la importancia que cada uno de ellos tiene en el fenómeno estudiado y además constituye una forma de explicación del mismo; segundo, porque la jerarquización de los atributos predictores optimiza el desempeño del proceso clasificador, al establecer que se hagan primero las preguntas clasificatorias que tengan mayor poder discriminador.

Cuando se diseña un experimento de aprendizaje automático, los atributos predictores se analizan estadísticamente para hallar correlaciones, con el propósito de guiar la implementación y evaluación de modelos clasificadores o predictivos. Si existen correlaciones, significa que algunos de los atributos no contribuirán significativamente al mejoramiento del poder discriminante del modelo. Por ello, algunos predictores correlacionados podrían ser descartados sin reducir la precisión del modelo, ya que su contribución está representada implícitamente por otro atributo.

Con base en las teorías lingüísticas y en la intuición, es de esperar que algunas de las denominadas *palabras clave* (*cue words*) estén presentes en ciertos tipos de actos de diálogo que tienen contornos entonativos típicos; por ejemplo, algunos pronombres interrogativos, tales como *qué*, *cuándo*, *dónde*, aparecen comúnmente en las solicitudes de información, teniendo éstas un contorno característico. Es por esto que

podría esperarse una relación entre la presencia de algunas palabras clave (en el nivel de representación léxica) y las etiquetaciones INTSINT (del nivel del etiquetaje entonativo).

En experimentos piloto para el presente trabajo, reportados en (Coria y Pineda, 2005, 2005a y 2006), al usar una pequeña muestra del corpus DIME (Pineda *et al.*, 2007) se observa que algunas clases de actos de diálogo están relacionadas con algunas de la modalidad del enunciado; p. ej. la solicitud de información está estrechamente relacionada con la modalidad interrogativa. Además, la modalidad del enunciado puede reconocerse a partir de la región final del contorno entonativo (f_0), representado como secuencias de etiquetas INTSINT, lo cual involucra una relación entre estos dos niveles representacionales; esto podría considerarse una evidencia de consistencia con la noción de *tonema* de Navarro Tomás. Al realizar experimentos de Aprendizaje Automático sobre mayores volúmenes de datos y con más atributos predictores, el presente trabajo produce modelos que muestran cuáles son los tipos de acto de diálogo que requieren un análisis intensivo del contenido léxico. Se esperaría que en implementaciones prácticas, el análisis del contenido léxico se simplifique y, en algunos casos, incluso no se requiera, al tomar ventaja de los demás niveles representacionales involucrados en la detección e interpretación del acto de diálogo.

1.3.2 Necesidad del parseo para la interpretación de los actos de diálogo

El *parseo* es el proceso de reconocer una cadena de entrada y asignarle una estructura para reconocerla (o rechazarla) como perteneciente a un conjunto (un lenguaje); la cadena en cuestión puede ser una sílaba, una palabra, una oración, etc. El logro de un parseo exitoso, es decir, el reconocimiento correcto de sílabas, palabras u oraciones, no garantiza una precisión satisfactoria en el reconocimiento del correspondiente tipo de acto de diálogo; esto se debe, en parte, a que el acto de diálogo se comunica mediante elementos de información complementaria, tales como la entonación y el contexto dentro del diálogo. Eventualmente, bajo determinadas condiciones, puede existir algún acto de diálogo que se comunique usando exclusivamente fuentes de información no léxica, o bien, que se apoye más en esta última que en la otra. En casos como éste, quizá el parseo no sería indispensable para reconocer el tipo de acto de diálogo, o bien, sería mínimamente requerido. Sobre esta base, el grado de necesidad del parseo sería determinado por el grado de incertidumbre del reconocimiento (clasificación o predicción) obtenido mediante modelos alternativos que usen la información no léxica. Shriberg *et al.* (1998) observan que al integrar árboles de decisión basados en información entonativa con modelos de lenguaje estadísticos de actos de diálogo específicos se mejora la tasa de reconocimiento del tipo del acto, en comparación con el uso de modelos de lenguaje solamente; en particular si se trata de una configuración donde la información léxica se basa en medios de reconocimiento automático.

Los resultados de Shriberg *et al.* también sugieren que en la conversación natural algunos tipos de actos de diálogo están marcados simultáneamente sobre la fuente de información léxica y también sobre la entonativa,

por lo cual una variedad de atributos entonativos extraíbles automáticamente podrían apoyar al procesamiento del diálogo en aplicaciones de habla. En el presente trabajo se hipotetiza que quizá la contribución de la información entonativa podría ser suficiente para reconocer algunos tipos de actos de diálogo sin requerir un proceso de parseo, o bien, aplicando un parseo sencillo.

Cuando se necesite un análisis de contenido léxico para reconocer algunos tipos específicos de actos de diálogo, en lugar de realizar una tarea de parseo podría hacerse la búsqueda de palabras clave (*clue words* o *cue words*) o frases clave (*clue phrases*) que típicamente se relacionen con los actos de diálogo en cuestión. Por ello, esta búsqueda constituiría una forma de resolver ambigüedades causadas por similitud entonativa; p. ej. la modalidad de enunciado interrogativo es altamente frecuente en las solicitudes de información, pero puede serlo también en las directivas de acción. La ambigüedad de origen entonativo podría resolverse a nivel léxico buscando pronombres interrogativos, tales como *qué, dónde, cómo*, etc. en el nivel representacional de las palabras. Este enfoque es comparable con el de Hirschberg y Litman (1993).

En análisis preliminares de algunas muestras del corpus DIME, presentados en (Coria y Pineda, 2005, 2005a y 2006) se observa que ciertos tipos de actos de diálogo se reconocen con mayor precisión que otros; p. ej. usando información entonativa, la *solicitud de información* es más fácil de reconocer que el *compromiso*. Quizá el segundo tipo de acto necesitaría más del parseo que el primero. Al implementar experimentos sobre una mayor cantidad de datos podrían producirse modelos reconocedores más robustos que muestren cuáles son los tipos de actos de diálogo que requieran un parseo más minucioso y la medida en que se requerirá.

Para concluir este capítulo, recuérdese que el análisis y la modelación de la relación entre el acto de diálogo y la entonación en esta investigación tienen, principalmente, el objetivo práctico que consiste en explorar la viabilidad de su aplicación para mejorar la robustez de los sistemas de administración de diálogo hablado. En los capítulos posteriores se evalúan las ventajas y desventajas que las diversas teorías sobre el acto de diálogo y la entonación presentan para la modelación empírica a la luz de los objetivos del presente trabajo.

2. Estado del arte

El estado del arte del presente trabajo está determinado por tres tópicos principales: (1) teorías y esquemas de etiquetación de actos de diálogo, (2) teorías y esquemas de representación de la entonación y (3) estudios sobre la relación entre acto de diálogo, modalidad de enunciado y entonación. A continuación se exploran estos tópicos.

Tres esquemas de etiquetación de actos de diálogo que resultan interesantes para los propósitos del presente trabajo son: DAMSL, de Allen y Core (1997), el del proyecto VERBMOBIL (Jekat *et al.*, 1995) y el de Levin *et al.* (1998), que se describen a continuación.

La noción de acto de diálogo puede considerarse una especialización de la noción de acto de habla con el propósito de ser implementada y analizada en una base computacional. Searle (1969) introdujo el término *acto de habla* refiriéndose a la unidad fundamental de comunicación lingüística y como una función de la intención de un enunciado. Bunt (1994) usa el término *acto de diálogo* para referirse a unidades funcionales usadas por el hablante para cambiar el contexto en un diálogo. Allen y Core (1997) DAMSL (*Dialogue Act Markup in Several Layers*), un esquema de etiquetación para actos de diálogo que ha sido usado en varios trabajos previos en el área.

DAMSL ofrece cuatro dimensiones de análisis para los actos de diálogo: *estado comunicativo*, *nivel informativo*, y funciones *hacia delante* y *hacia atrás*. Cada dimensión tiene un conjunto de etiquetas. El estado comunicativo describe si el enunciado es inteligible y si fue completado exitosamente; el nivel informativo clasifica el contenido semántico de un enunciado en *tarea*, *administración de la tarea*, o *administración de la comunicación*. La función hacia delante incluye actos de diálogo que delimitan las creencias y acciones de los participantes del diálogo. Finalmente, la función hacia atrás incluye actos de diálogo que relacionan a un enunciado con el discurso que precede.

Otro enfoque para la etiquetación de actos de diálogo es el del proyecto Verbmobil (Wahlster, 1993), del cual se presenta una taxonomía de actos de diálogo en (Jekat *et al.*, 1995). Una diferencia importante entre éste y DAMSL es que Verbmobil aborda los actos de diálogo de un dominio específico: la planeación (agendamiento) de citas de negocios y, por lo tanto, el conjunto de etiquetas está restringido a este dominio; en cambio, DAMSL tiene un alcance más amplio y puede usarse para etiquetar diálogos prácticos en dominios arbitrarios.

Levin *et al.*, (1998) proponen un esquema para etiquetar actos de diálogo basado en DAMSL, que incorpora una estructura de tres niveles: (1) *acto de habla*, que consiste en un conjunto de etiquetas extendido a partir de DAMSL y adaptado a las características del corpus Switchboard (Godfrey *et al.*, 1992); (2) *juego de diálogo*, definido como conjuntos de enunciados limitados por el cambio de interlocutor que tiene la *iniciativa* y por cambios en la *intención* del hablante y (3) *actividad*, definida como un conjunto de juegos de diálogo que tratan un mismo tópico.

Además de las teorías de actos de diálogo y sus respectivos esquemas de etiquetación, las teorías sobre la entonación, particularmente del idioma español, son una componente relevante en el presente trabajo. La noción de *tonema*, introducida por Navarro Tomás (1948) y la de *juntura terminal*, por Quilis (1981), refieren a la región final del contorno entonativo del enunciado, considerándola una fuente de información lingüística para determinar la modalidad del enunciado así como el correspondiente tipo de acto de habla.

Uno de los esquemas más conocidos para etiquetación de la entonación desde el punto de vista fonológico es ToBI, *Tone and Break Indices* (Beckman y Ayers-Elam, 1997), cuya implementación para Español, Sp-ToBI, se presenta en (Beckman *et al.*, 2002). Desde el punto de vista fonético, INTSINT es otro esquema para representar la entonación; está basado en un enfoque más perceptivo que lingüístico para analizar la señal de habla.

Respecto a la investigación sobre la relación entre entonación, modalidad del enunciado y acto de diálogo, algunas propuestas relevantes han sido presentadas por Wilson y Sperber (1988), Mast *et al.* (1996), Garrido (1991 y 1996), Shriberg *et al.* (1998) y Fernández y Picard (2002). Wilson y Sperber analizan la modalidad del enunciado desde un punto de vista semántico y su argumento es que las características lingüísticas típicas de las formas declarativa, imperativa e interrogativa contienen instrucciones que guían al oyente para la interpretación del contenido léxico. Mast *et al.*, dentro del proyecto VERBMOBIL, usan datos empíricos para implementar modelos de reconocimiento de acto de diálogo donde algunas piezas de información extraídas de la entonación se usan entre los datos de entrada. Garrido aborda un análisis empírico de los patrones entonativos del Español tomando teorías clásicas como línea de base para analizar la modalidad del enunciado en datos empíricos obtenidos de un corpus de habla; sus resultados extienden los de sus predecesores. Finalmente, Shriberg *et al.* usan datos de un corpus etiquetado con DAMSL, así como técnicas de aprendizaje automático, particularmente árboles de clasificación, para producir modelos de reconocimiento de actos de diálogo; ellos representan la entonación como señal *en crudo*, es decir, sin aplicarle ningún esquema de etiquetación y sus resultados muestran que en el inglés hablado la entonación puede contribuir al reconocimiento de actos de diálogo.

2.1 Algunos actos de diálogo cuya detección podría beneficiarse de la información prosódica

Los casos más evidentes donde la detección del acto de diálogo puede beneficiarse de la información prosódica son aquellos donde la información léxica y sintáctica no es suficiente para reconocer el tipo de acto correspondiente, como aquéllos que involucran una entonación típica; p. ej. la solicitud de información expresada como afirmación con entonación interrogativa, como se muestra a continuación:

Afirmación: *Así está bien.*

Solicitud de información: *¿Así está bien?*

Por ejemplo, una directiva de acción, una afirmación y una solicitud de información pueden ser indistinguibles entre sí si el contenido léxico es igual. Particularmente en español, el rol del orden de las palabras para marcar actos de diálogo en los enunciados no es tan relevante como en inglés u otros idiomas; en cambio, la entonación parece ser más relevante. Éstos son algunos ejemplos: *tú me das agua*, afirmación; *¿tú me das agua?*, solicitud de información; *¡tú me das agua!*, directiva de acción. Ésta es otra motivación para analizar tanto los atributos prosódicos como los léxicos al hacer clasificación automática de los tipos de actos de diálogo.

En resultados exploratorios obtenidos para el presente trabajo, la duración del enunciado fue un atributo útil para clasificar algunos actos de diálogo de common ground (Coria y Pineda, 2006). La duración también fue útil para reconocer algunos actos de diálogo de obligaciones, siendo usada con otros atributos no prosódicos, tales como la modalidad del enunciado; esta última fue reconocida a partir de datos de la entonación.

Una observación general de los datos empíricos muestra que el acto de diálogo *opción abierta* frecuentemente se presenta en forma de enumeración, la cual incluye una o más pausas. Por ello, se esperaría que la detección de pausas en un enunciado fuera indicio de la ocurrencia de una opción abierta. Por otra parte, algunos actos de diálogo parecen estar relacionados con la presencia de alargamientos vocálicos; p. ej. las solicitudes de información enunciadas con entonación interrogativa parecen presentar alargamiento vocálico en la última sílaba del enunciado. Este fenómeno se presenta también en algunas afirmaciones, directivas de acción o respuestas cuando el hablante duda antes de especificar objetos, ubicaciones, trayectorias o zonas. Una expectativa intuitiva es que las directivas de acción enunciadas con entonación interrogativa parecen no presentar alargamiento en la última vocal del enunciado, mientras que las solicitudes de información del tipo *sí/no* sí lo presentan. Se requerirían análisis estadísticos más amplios para confirmar estas expectativas.

2.2 El trabajo de Shriberg *et al.*

Shriberg *et al.* (1998) usan datos empíricos del corpus Switchboard (Godfrey *et al.*, 1992), constituido por conversaciones telefónicas en inglés americano. Su enfoque general involucra el uso tanto de algoritmos de aprendizaje automático para clasificar actos de diálogo a partir de atributos prosódicos, como modelos de lenguaje basados en características léxicas.

Para aprovechar la información proveniente de la entonación y del contenido léxico, Shriberg *et al.* combinan *multiplicativamente* las probabilidades obtenidas de modelos de árbol clasificatorio con las de modelos de lenguaje. La probabilidad proveniente del árbol se estiman con $P(U|F)$, donde U es el tipo de acto y F es un conjunto de atributos prosódicos; en el modelo de lenguaje se estiman las probabilidades (*likelihoods*) del tipo de acto mediante $P(W|U)$, donde W es la secuencia de palabras observadas. De este modo, se escoge el tipo de acto que tenga mayor probabilidad, basándose tanto en la entonación como en el contenido léxico, resultando $P(F, W|U)$. Por razones prácticas, se asume la simplificación de que los atributos entonativos son independientes de los léxicos y con base en ello se construyeron los modelos clasificadores; aunque bastaría observar algunos casos típicos para mostrar que no son independientes. Por ejemplo, al hablar sobre el equipo deportivo favorito se esperaría una entonación distinta que al hablar sobre impuestos.

$$\begin{aligned} P(F, W|U) &= P(W|U) P(F|W, U) \\ &\approx P(W|U) P(F|U) \\ &\propto P(W|U) P(U|F) \end{aligned}$$

Es pertinente la última línea porque, con base en el Teorema de Bayes, $P(F|U)$ es: $P(U|F) P(F) / P(U)$. El segundo factor, $P(F)$, es el mismo para todos los tipos de U y, dado que todos los tipos de acto están igualmente representados tras la aplicación del procedimiento de submuestreo, $P(U)$ también es la misma para todos los tipos. Por ello, la probabilidad $P(U|F)$, estimada por el árbol, es proporcional a la probabilidad $P(F|U)$.

Las probabilidades basadas en palabras se calculan mediante la producción de 42 modelos de lenguaje, uno para cada tipo de acto de diálogo considerado.

2.3 El proyecto VERBMOBIL

De acuerdo a Niemann *et al.* (1997), Verbmobil es el primer sistema para entendimiento del habla donde se integra el análisis de la prosodia, obteniéndose mejoras drásticas en el desempeño. Es un sistema para traducción de habla a habla en el dominio de diálogos para agendamiento de citas de negocios; en los cuales

dos interlocutores intentan fijar una fecha, hora y lugar de reunión. El énfasis radica en la traducción de enunciados del alemán al inglés. En octubre de 1996 se presentó al público un prototipo de investigación en forma exitosa. Otras referencias útiles al respecto son (Block, 1997) y (Wahlster *et al.*, 1997).

Verbmobil calcula probabilidades para límites de cláusula, acentuación y diferentes tipos de modalidad del enunciado para cada una de las hipótesis de palabras calculadas por un reconocedor de palabras. Estas probabilidades guían la búsqueda del análisis lingüístico. La desambiguación se logra durante el análisis y no por una verificación prosódica de hipótesis lingüísticas. La información prosódica más útil es la proporcionada por los límites de cláusula. Éstos se detectan con una tasa de reconocimiento del 94%. Para el parseo de las gráficas de hipótesis de palabras el uso de probabilidades de límite de cláusula proporciona una aceleración del 92% y una reducción del 96% de lecturas alternativas.

Los módulos de sintaxis usan la información de límites de cláusula del módulo de prosodia junto con los resultados de acústica de las hipótesis de palabras y modelos de lenguaje estocásticos de n-gramas para preseleccionar entre las muchas rutas combinatoriamente posibles a través de la gráfica de palabras. Estas cadenas preseleccionadas de palabras, las cuales contienen información de límites de oración, se analizan después usando una gramática TUG (*Trace Unification Grammar* o *Gramática de Unificación de Rastreo*) en el módulo de sintaxis desarrollado por Siemens (Block, 1997; Block y Schachtel, 1992) y una gramática HPSG (*Head-driven Phrase Structure Grammar* o *Gramática de Estructura de Frase Dirigida al Núcleo*) en el módulo de sintaxis desarrollado por IBM (Pollard y Sag, 1987).

El módulo semántico recibe del módulo de sintaxis un árbol de parseo, la cadena de palabras subyacente y los resultados prosódicos para acentuación. Basándose en éstos, se crean Estructuras de Representación del Discurso (*Discourse Representation Structures, DRS*) subespecificadas.

De acuerdo a Eberle (1996), el módulo de transferencia del sistema Verbmobil traduce DRSs que representan la información semántica subyacente al enunciado, en DRSs correspondientes a oraciones en inglés. Esta tarea podría involucrar análisis pragmático y desambiguación que es parcialmente hecha por el módulo de evaluación semántica. El módulo de transferencia usa información de acentuación y de modalidad oracional para algunas tareas. La información de modalidad oracional se usa para distinguir entre preguntas y no-preguntas si no hay indicadores gramaticales; p. ej. las preguntas y las declarativas con elisión de tópico pueden tener un orden de palabras idéntico. La información de acentuación desambigua principalmente la interpretación de las partículas.

Según Lieske *et al.* (1997), la integración de información prosódica en los análisis sintáctico y semántico introduce problemas tanto prácticos como conceptuales; por ello, se tuvieron que integrar datos descriptivos de la prosodia en un parser de cadenas modificado. Además, los valores probabilísticos provistos por el

análisis prosódico tuvieron que ser adaptados a la lingüística basada en unificación, procesando principalmente símbolos discretos.

Algunos de los parámetros que modelan a la prosodia en Verbmobil son las probabilidades del acento oracional y de la entonación, que se codifican a manera de aristas dentro de la representación gráfica constituida por una *WHG* (*Word Hypothesis Graph* o Gráfica de Hipótesis de Palabras), como lo describe Block (1997). Un módulo de reconocimiento prosódico proporciona valores que describen al acento y también valores para describir la entonación (*ascendente*, *descendente* y *medio*). Para transferir esta información a los módulos de procesamiento sintáctico y semántico, durante la búsqueda lingüística se hace un mapeo para generar la denominada *forma de palabra prosódica* (*PROSWOF*, *Prosodic Word Form*). Existen $10 \times 10 \times 3$ (=300) *PROSWOFS* diferentes, correspondientes a 10 valores diferentes de acento y 10 de límite de cláusula (calculados a partir de sus probabilidades) y los valores *ascendente* (*rise*), *descendente* (*fall*) y *medio* (*prog* ó *mid*) para la entonación. Así que una *PROSWOF* tiene la forma $a[0-9]_g[0-9]_i\{rise, fall, prog\}$. En las entradas léxicas para estas *PROSWOFS* los valores se copian a atributos gramaticales, por ejemplo:

```
lexicon(a3_g4_iris, proswoff:0) |
    0:accent = 3,
    0:boundary = 4,
    0:intonation = rise.
```

En la gramática, para cada categoría terminal se introduce una regla que expande la categoría sintáctica a la categoría léxica y a la *PROSWOF*; por ejemplo:

```
advgr:0 → adv:1, proswoff:2 |
    0:accent = 2:accent,
    0:boundary = 2:boundary,
    0:intonation = 2:intonation.
```

Por medio de estas reglas la información prosódica se vuelve accesible para un procesamiento semántico ampliado. El componente semántico decide, por ejemplo, que una oración ambigua (en alemán) como *kommen sie in mein Buro* será interpretada como imperativa si el valor del atributo de entonación es *descendente* y como *pregunta sí/no* si su valor es *ascendente*. De acuerdo a esto, la traducción sería *Venga a mi oficina* o *¿Usted viene a mi oficina?*.

Mast *et al.* (1996) describen el uso de tres métodos combinados para clasificar actos de diálogo con la ayuda de la información prosódica: Perceptrón Multicapa (MLP, *Multi Layer Perceptron*), modelos de lenguaje (LM, *language models*) de poligrama (no confundir con *n*-grama) y árboles de clasificación semántica (SCT,

Semantic Classification Trees). Estos métodos se usaron para reconocer límites de cláusula, acentuación y modalidad oracional; los poligramas son una mezcla de n -gramas con tamaño variante de n . Los árboles de clasificación semántica pueden verse como sistemas basados en reglas donde éstas no son codificadas a mano sino entrenadas sobre un corpus de texto.

2.4 Trabajo de Fernández y Picard

El objetivo general del trabajo de Fernández y Picard (2002) consiste en evaluar la utilidad (entendida como precisión) del algoritmo de la Máquina de Soporte Vectorial (*Support Vector Machine, SVM*), comparándola con otros algoritmos de aprendizaje automático como las redes neuronales y los árboles de decisión, para clasificar actos de diálogo usando información entonativa. Sus resultados sugieren que la SVM ofrece una mejor alternativa sobre los otros algoritmos para este propósito.

Se usan datos del corpus CallHome (Finke, Lapata *et al.*, 1998), en Español, para analizar y modelar la relación entre información entonativa y actos de diálogo. Por tratarse de este idioma, resulta de especial interés para los objetivos del presente trabajo. Es atractivo también el uso de un esquema basado en DAMSL para el etiquetaje de los actos de diálogo (Levin *et al.*, 1998). Sus diálogos no son orientados a tareas, sino conversaciones telefónicas más bien informales.

2.5 Comparación entre Shriberg *et al.* y VERBMOBIL

El proyecto Verbmobil usa atributos prosódicos para la clasificación de actos de diálogo mediante la aplicación de una serie de etapas intermedias de la clasificación: límites de cláusula, acentuación y modalidad oracional. Shriberg *et al.* (1998) no realizan una tarea intermedia como ésta; en cambio, usan datos “en crudo” de la señal del habla para extraer una serie de parámetros numéricos que se usan en el clasificador basado en prosodia.

Verbmobil usa Perceptrones Multicapa y poligramas (estructuras basadas en n -gramas) para analizar los atributos numéricos de la señal del habla y para producir modelos clasificadores de límite de cláusula, acentuación y modalidad oracional. Shriberg *et al.* usan el algoritmo CART (Breiman *et al.*, 1983) sobre atributos prosódicos numéricos, lo cual se complementa después con modelos de lenguaje (LM, *language models*) para cada tipo de acto de diálogo. Una diferencia es que Verbmobil usa LM en etapas tempranas del proceso clasificatorio, mientras que Shriberg *et al.* los usan en las etapas finales. El segundo enfoque produce en realidad dos clasificaciones independientes del acto de diálogo, la basada en prosodia y la basada en LM, y luego las compara entre sí, mientras que el primero produce una sólo clasificación.

2.6 Comparación entre los enfoques previos y el del presente trabajo

En la tabla 2.1 se presenta un cuadro donde se contrastan los enfoques previos y el del presente trabajo. Primeramente, tiene que abordarse una diferencia fundamental: la representación de la frecuencia fundamental (f_0) en el proyecto DIME al usar un esquema discreto (INTSINT, *International Transcription System for Intonation*) en lugar de uno numérico. Las etiquetaciones INTSINT permiten un alto grado de abstracción en la representación, evitando así la necesidad de normalizar los valores de la f_0 , y aún permitiendo una representación de la entonación en forma de cadena alfabética. Por otra parte, esta abstracción representacional de la entonación permitiría usar una serie de técnicas de análisis que los trabajos previos en el área no usaron; sin embargo, no permitiría desarrollar otros análisis que aquéllos sí hicieron.

Algunas experiencias obtenidas a lo largo del presente trabajo muestran que pueden ser aplicables diversos aspectos de ambos enfoques: una tarea intermedia de clasificación de la modalidad del enunciado, como en *Verbmobil*, y un modelo para clasificación preliminar de actos de diálogo a partir solamente de atributos prosódicos, enriquecido después con LM como en *Shriberg et al.* Algunos resultados preliminares (Coria y Pineda, 2005, 2005a, 2006) han mostrado que la clasificación de modalidad del enunciado, como paso intermedio, es altamente relevante en la clasificación de actos de diálogo de obligaciones.

En los experimentos preliminares no se usaron atributos léxicos para el reconocimiento del acto de diálogo; sin embargo, la necesidad de dichos atributos parece evidente tanto para la evaluación del modelo basado en prosodia como para incrementar la precisión de la clasificación.

En el trabajo de Fernández y Picard los diálogos de su corpus no son orientados a tareas, mientras que en los de *Shriberg et al.*, *VERBMOBIL* y *DIME* sí lo son; por ello, sería difícil comparar sus resultados con los de estos tres últimos. En el presente trabajo se usan árboles de decisión al estilo de *CART*, mientras que Fernández y Picard usan *SVM*. Otra diferencia radica en los respectivos objetivos: en el primero, es evaluar la contribución de la entonación al reconocimiento automático del tipo de acto de diálogo; en el segundo, es la evaluación de la *SVM* como algoritmo para reconocerlo.

2.7 Importancia de la relación entre el contorno entonativo y la modalidad del enunciado

En el Español se ha estudiado la relación entre el contorno entonativo y la modalidad del enunciado. Dichos estudios son de gran utilidad para el presente trabajo porque constituyen una guía en la implementación y evaluación de modelos de reconocimiento automático de la modalidad del enunciado y éstos, a su vez, podrían contribuir al reconocimiento del acto de diálogo. En algunos trabajos empíricos previos, como el

proyecto Verbmobil, se ha observado y aprovechado la correlación entre el contorno entonativo y otras variables para reconocer automáticamente el acto de diálogo.

Autor	Tipo de diálogo	Etiquetaje de actos de diálogo	Representación de la entonación	Algoritmo de Aprendizaje Automático
Shriberg <i>et al.</i>	Conversación sobre tópicos generales	SWB-DAMSL	Parámetrica	Árbol de decisión (CART) y modelos estadísticos de lenguaje
VERBMOBIL	Práctico	Propio	Parámetrica	Perceptrón multicapa y poligramas
Fernández y Picard	Conversación sobre tópicos generales	Levin et al. (basado en DAMSL)	Parámetrica	Máquina de soporte vectorial (SVM)
Presente	Práctico	DIME-DAMSL	INTSINT	Árbol de decisión (J48)

Tabla 2.1 Cuadro comparativo de los enfoques de trabajos previos y del presente

En análisis preliminares (Coria y Pineda, 2005, 2005a, 2006) al presente trabajo se ha observado la correlación entre contorno entonativo y modalidad del enunciado, siendo esto consistente con lo estipulado en diversas teorías de la entonación, como las de Navarro Tomás (1948) y Quilis (1981).

La relación entre los dos atributos en cuestión involucra que la entonación es un medio para transmitir información acerca de, entre otras cosas, la categoría gramatical a la cual pertenece un enunciado (*declarativa, interrogativa, imperativa, etc.*) y acerca de qué uso debe darse a la información léxica contenida en éste.

2.8 El tonema en español: noción y tipología

El *tonema* del idioma español ha sido ampliamente analizado por Navarro Tomás (1948). El tonema es la región final del contorno entonativo de un enunciado. Inicia en la última sílaba tónica e incluye las siguientes sílabas átonas hasta el final del enunciado y constituye la parte más significativa del contorno entonativo.

Las diferencias relativas entre las frecuencias fundamentales (f_0) de los tonos que ocurren a lo largo del contorno entonativo son el vehículo de expresión del tonema. Las magnitudes de dichas diferencias están acotadas por el rango entonativo del hablante.

García-Lecumberri (2003) analiza la teoría de Navarro Tomás y la compara con la de la Escuela Británica. Deben considerarse dos diferencias importantes: primero, el análisis de Navarro Tomás se ubica dentro de las denominadas *análisis de tonos*; en cambio, la Escuela Británica, dentro del análisis de configuración; segundo, la Escuela Británica no considera a las sílabas átonas finales como componente integrante del tonema.

Es necesario explicar algunos conceptos clave de la teoría de Navarro Tomás antes de describir la estructura del tonema. Aguilar (2000) explica algunos de ellos, entre los cuales destacan: la *unidad melódica*, las *ramas inicial y final* y el *cuerpo*.

Como lo define Navarro Tomás (1948), *la unidad melódica es la parte más corta de discurso con sentido propio y forma entonativa específica*. Sus límites pueden coincidir con pausas o con variaciones de tono, aún dentro de un mismo enunciado, dependiendo de factores tales como el hablante o la velocidad de enunciación. El *grupo fónico* refiere a la parte del discurso localizada entre dos pausas sucesivas y el *grupo entonativo* refiere a aquellos fragmentos de habla que no están necesariamente delimitados por silencios y que, sin embargo, reciben una entonación determinada (Quilis, 1993). Si se producen pausas en el enunciado, entonces no hay diferencia entre grupos entonativos y grupos fónicos. Sin embargo, una misma secuencia puede enunciarse mediante la realización de cambios de tono sin detener la producción; de este modo, los entonativos no coinciden con grupos fónicos pero sí coinciden con unidades melódicas.

Aguilar apunta, de acuerdo a Navarro Tomás, que la unidad melódica puede dividirse en tres partes: *rama inicial*, denominada también *rama tensiva* o *tema*; *cuerpo* y *rama final*, llamada también *rama distensiva* o *rema*. La rama inicial está constituida por todas las sílabas desde el inicio hasta la última sílaba tónica. El cuerpo de la unidad está formado por la primera sílaba tónica, adicionada con las sílabas restantes incluyendo hasta la que precede al último acento. La rama final está formada por la última sílaba tónica y las siguientes, en caso de que estas últimas existan.

Quilis (1981) presenta la noción de *morfema de entonación*, equiparable al tonema de Navarro Tomás. El morfema de entonación es la abstracción fonológica del contorno entonativo. Su tipología se basa en tres conceptos: *fonema tonal*, *fonema acentual* y *juntura terminal*. El fonema *tonal* es el tono relativo de una sílaba comparado con los tonos de las otras sílabas del enunciado. El fonema *acentual* es el acento definido por las reglas de acentuación para las palabras del idioma. La juntura terminal describe cómo es la forma de la región final del contorno entonativo. Se identifican tres fonemas tonales: *bajo* (representado con 1), *medio* (representado con 2) y *alto* (representado con 3). También, se describen dos fonemas acentuales: *fuerte* (el que es marcado en el enunciado) y *débil* (el que no lo es). Finalmente, existen tres tipos de junturas terminales: *ascendente*, *descendente* y *suspensión*, aunque Quilis afirma que la suspensión realmente no existe en Español. Las combinaciones de estos tres parámetros producen una serie de tipos de morfemas de entonación.

Una de las tipologías más conocidas del tonema fue propuesta por Navarro Tomás (1948), presentando cinco tipos: cadencia (*falling*), anticadencia (*rising*), semicadencia (*half-falling*), semianticadencia (*half-rising*) y suspensión (*level*). García-Lecumberri (2003) interpreta esta tipología en forma gráfica como se muestra en la fig. 2.1.



Figura 2.1 Representación gráfica de la tipología de tonemas del Español de Navarra Tomás según la interpretación de García-Lecumberri.

Con base en las interpretaciones de García-Lecumberri (*op. cit.*), se describe a continuación la tipología de tonemas de Navarro Tomás:

La *cadencia* (fig. 2.1a) es el tonema con el cual terminan las declarativas en grupos entonativos finales absolutos; es decir, en el rema del enunciado; también se usa, con algunas variaciones, en muchas interrogativas (p. ej. las interrogativas pronominales), exclamativas e imperativas. La cadencia marca el punto más bajo de la unidad melódica.

La *semicadencia* (fig. 2.1b) es el tonema en grupos finales no absolutos; por ejemplo, en las enumeraciones y en cierto tipo de interrogativas (preguntas aseverativas). La semicadencia se realiza como un cambio en la dirección del tono en el sentido descendente, pero menos grave que la cadencia.

La *anticadencia* (fig. 2.1c) expresa continuidad, información incompleta; por ello, se usa en la rama tensiva para señalar el tema, típicamente precediendo al grupo que termina en cadencia; está presente en algunas interrogativas (preguntas reiterativas) y en exclamativas. La anticadencia marca el punto más alto del contorno entonativo de una frase.

La *semianticadencia* (fig. 2.1d) separa grupos entonativos con una diferenciación menor que la de grupos marcados por anticadencia. La semianticadencia es un movimiento ascendente del contorno entonativo menos agudo que la anticadencia.

La *suspensión* (fig. 2.1e) se usa entre grupos estrechamente relacionados; por ejemplo, precediendo a una aposición o a un vocativo. La suspensión es la ausencia de cambio en el contorno entonativo.

Por su parte, Navarro Tomás (1948) también describe un patrón de anticadencia-cadencia (*rise-fall*).

Para Aguilar, los contornos finales descendente o ascendente son los únicos que cumplen una función distintiva; las otras terminaciones (semicadencia, semianticadencia, suspensión) cumplen una función delimitativa y producen un contraste interno en el enunciado. Es decir, semicadencia y semianticadencia aparecen dentro de los enunciados y no marcan su final.

Solamente tres tipos de tonemas son considerados por Quilis (1981): cadencia, anticadencia y suspensión.

2.9 Principales estudios sobre la relación entre el tonema y la modalidad del enunciado

El fenómeno de mayor interés en el estudio de la relación entre el fonema y la modalidad es que ciertos tipos de tonema están estrechamente relacionados con modalidades de enunciado específicas. Como lo observa Aguilar (*op. cit.*), *los contornos melódicos suelen presentar una forma típica de acuerdo al tipo de enunciado.*

Existen algunas diferencias en la terminología y la representación del contorno entonativo entre los trabajos de investigación que abordan este tópico; sin embargo, la mayoría de ellos concuerdan en algunos patrones que relacionan a los tonemas con las modalidades del enunciado.

Un fenómeno comúnmente observado es que algunos de los tipos de la modalidad interrogativa están asociados con un contorno entonativo final ascendente; también, otras modalidades están definitivamente excluidas de presentar este comportamiento y, en cambio, presentan contorno descendente o ascendente-descendente.

Modalidad del enunciado	Navarro Tomás (1948)	Canellada y Kuhlmann (1987)	Quilis (1981, 1993)
Declarativa	Descendente	Descendente	***
Pregunta sí/no	Ascendente	Ascendente	Ascendente
Pregunta pronominal	Ascendente o descendente	***	Ascendente o descendente
Pregunta relativa	Ascendente-descendente	Descendente	***
Exclamativa	Descendente o Ascendente o Ascendente-descendente	***	***
Imperativa	Descendente o Ascendente-descendente	***	***

Tabla 2.2 Patrones entonativos de diversas modalidades del enunciado, resumido por Garrido (1996).

Garrido (1996) resume algunos indicios (*cues*) entonativos para marcar seis tipos de enunciado en Español: declarativas, preguntas *sí/no*, preguntas pronominales, preguntas relativas, exclamativas e imperativas. Los indicios entonativos considerados son: movimiento final de la *f0* (región final del contorno), forma global y otros indicios. El autor incluye referencias a análisis previos de Navarro Tomás, Quilis, Garrido, y Canellada y Kuhlmann y describe algunas coincidencias entre sus respectivos resultados, como se muestra en la tabla 2.2.

Hay cuatro categorías de entonación principales identificadas por Navarro Tomás: enunciativa, interrogativa, volitiva y emotiva. Dentro de cada categoría se definen tipos de modalidad de enunciado, cada uno asociado a patrones entonativos específicos, representados como combinaciones y variaciones en la forma de la rama inicial, el cuerpo y la rama final de la unidad melódica. La tabla 2.3 presenta las modalidades de enunciado analizadas para cada categoría entonativa; en la segunda columna se muestra el número de tipos de modalidad de enunciado que considera. Por razones de espacio, los patrones específicos de las modalidades no se describen aquí.

Categoría Entonativa	Modalidad del Enunciado
Enunciativa	Aseveración (4)
	Enumeración (13)
	Complemento (4)
	Locución (2)
	Coordinaciones (4)
	Otras enunciativas (4)
Interrogativa	Preguntas (13)
Volitiva	Diferentes tipos (12)
Emotiva	Énfasis (6)
	Exclamaciones (3)
	Uniformidades (2)
	Otras emotivas (3)

Tabla 2.3 Modalidades de enunciado tipificadas por Navarro Tomás (1948).

Quilis (1981) presenta una descripción del patrón entonativo de cada modalidad del enunciado. La tabla 2.4 presenta sus categorías y subcategorías de modalidades. Los patrones entonativos se representan usando una simbología para niveles tonales, fonemas acentuales y junturas terminales.

En su trabajo, Garrido (1996, p. 146) observa que *los movimientos finales son el principal indicio entonativo para el tipo [modalidad] de enunciado en Español y en otros idiomas*. El *slope* (declinación), la altura global y el rango de la frecuencia fundamental no fueron indicios útiles (pp. 173-178).

Garrido analizó patrones entonativos en cuatro tipos de modalidad del enunciado: declarativas, preguntas *sí/no* (equiparables con las preguntas absolutas de Navarro y de Quilis), preguntas pronominales (comparables con las de la misma categoría de Navarro y de Quilis) y exclamativas. Una de sus conclusiones establece que solamente los movimientos ascendentes parecen ser claramente excluidos de los enunciados declarativos. Se observan algunas tendencias y se pueden establecer los patrones más preferidos, pero pueden usarse otros tipos en lugar de los típicos (p. 147). En la tabla 2.5 se presentan algunas cifras de Garrido respecto al contorno final, quien aclara que los valores para las exclamativas no se consideran confiables, debido al escaso número de éstas en el corpus (p. 147).

Categoría	Subcategoría
Declarativa	***
Interrogativa	Absoluta
	Pronominal
	Disyuntiva y copulativa
	Pregunta pronominal con cortesía
	Pregunta con cortesía
	Pregunta relativa
	Pregunta confirmativa
	Pregunta imperativa
Enumeración	Completa
	Incompleta
Complemento	Hiperbatizada
Coordinación	Coordinación de primer grado
	Coordinación de segundo grado
Enunciación parentizada	***
Subordinación	***
Exclamación	***

Tabla 2.4 Modalidades de enunciado presentadas en (Quilis, 1981).

	Descendente	Ascendente	Ascendente-descendente
Declarativa	82.61	---	17.38
Pregunta sí/no	19.99	46.67	33.34
Pregunta pronominal	33.33	50	16.67
Exclamativa (*)	40	40	20

Tabla 2.5 Porcentajes de contornos entonativos finales por tipo de enunciado (Garrido, 1996, p. 147)

Las cifras de Garrido parecen confirmar los resultados generales presentados en los trabajos más conocidos del área; sin embargo, debe señalarse una diferencia: *el uso de los patrones ascendente-ascendente en grupos fónicos de final del enunciado (...) emerge como indicio alternativo para los patrones descritos en estudios clásicos* (p. 148); esto fue reportado para las declarativas y las preguntas sí/no.

Otros hallazgos reportados para el español son, por ejemplo, los de Alcoba y Murillo (1998), quienes describen patrones de la relación entre etiquetajes INTSINT y modalidades declarativa e interrogativa del enunciado. En las declarativas, generalmente existe juntura terminal descendente. Las preguntas *sí/no* son opuestas a las preguntas pronominales, ya que estas últimas presentan el patrón declarativo normal y las primeras coinciden con el patrón no terminal, marcado por una juntura terminal ascendente. Por su parte, Aguilar (2000) revisa, además de las teorías de Navarro Tomás, las de Quilis y otras también destacables y Ávila (2003) analiza el caso particular de la entonación del enunciado interrogativo de la ciudad de México.

Para concluir, la exploración del estado del arte nos ha permitido revisar una serie de trabajos teóricos y prácticos en materias de acto de diálogo, entonación y análisis y modelación de su relación; de este modo se pueden sentar las bases para elegir esquemas de etiquetaje adecuados a las necesidades de la presente investigación y para el planteamiento de experimentos de aprendizaje automático que conduzcan a la generación de modelos con propósitos explicativos y clasificatorios.

3. Teoría de DIME-DAMSL

Un esquema de etiquetaje es una serie de reglas para clasificar instancias en categorías predefinidas, constituidas por una serie de etiquetas; el uso de las reglas se basa en la observación de características descriptivas de las instancias. Ejemplos de esquemas de etiquetaje son las metodologías para clasificar especies de seres vivos en las ciencias biológicas, los métodos de diagnóstico de enfermedades usados en la medicina, la tipificación de delitos en las ciencias jurídicas, etc.

3.1 Motivación y objetivos

En Lingüística Computacional es común el uso de esquemas de etiquetaje para analizar y modelar diversos fenómenos lingüísticos con el apoyo de herramientas computacionales. Ejemplos de estos esquemas son los diversos alfabetos fonéticos computacionales desarrollados para etiquetar corpus de habla y los esquemas para etiquetar la entonación, como ToBI (Beckman y Ayers-Elam, 1997) o INTSINT (Hirst *et al.*, 2000).

Al permitir la representación computacional de fenómenos lingüísticos, los esquemas de etiquetaje sirven para crear modelos reconocedores, clasificadores o predictivos de variables involucradas en dichos fenómenos. Por ejemplo, los alfabetos fonéticos permiten la creación de sistemas de reconocimiento automático del habla, y los esquemas para representar la entonación sirven para dotar de una mayor naturalidad al habla sintetizada en sistemas TTS (*Text-To-Speech*).

Del mismo modo, un esquema de etiquetación de actos de diálogo permite su análisis y modelación, con lo cual se pueden enriquecer y robustecer los sistemas automáticos de administración de diálogo.

En Lingüística Computacional, el análisis y la modelación del diálogo se han concentrado en los denominados *diálogos prácticos* (*practical dialogues*), identificados por Allen *et al.* (2000), en los cuales la conversación está enfocada en la realización de una tarea con metas definidas. Este tipo de diálogo presenta un grado de complejidad menor que la conversación general y su estructura es independiente del dominio. Por ello, se han desarrollado diversos esquemas de etiquetaje para los actos de diálogo correspondientes a este paradigma; uno de los más conocidos es DAMSL, *Dialogue Act Markup in Several Layers*, (Allen y Core, 1997) y otro es el implementado dentro del proyecto VERBMOBIL (Wahlster, 1993). Uno más es el propuesto por Levin *et al.* (1998), usado en el trabajo de Fernández y Picard (2002). El esquema DIME-DAMSL (Pineda *et al.*, 2006 y 2006a), descrito más adelante, fue desarrollado tomando a DAMSL como punto de partida debido a su amplia difusión y uso en diversos trabajos previos del área.

3.2 Esquema DAMSL

DAMSL está constituido por etiquetas agrupadas en cuatro conjuntos: 1) *estado comunicativo*, 2) *nivel informativo*, 3) *función hacia delante* y 4) *función hacia atrás*.

El *estado comunicativo* describe si un enunciado es ininteligible, si fue dejado inconcluso (abandonado) por el hablante o si es un monólogo; para ello se usan las etiquetas: *uninterpretable*, *abandoned* y *self-talk*, respectivamente. Se considera ininteligible si el sujeto habló usando una amplitud muy baja, en forma confusa o si hubo problemas causados por el medio de comunicación que impidió que el oyente pudiera entender. Un enunciado se considera abandonado si el hablante interrumpe bruscamente su realización. Se considera monólogo cuando el sujeto habla en forma audible y entendible, dirigiéndose a sí mismo, sin la intención de que su interlocutor atienda lo dicho.

El nivel informativo describe el tipo de asunto tratado en un enunciado, pudiendo éste ser alguno (o varios) de los siguientes: 1) *tarea*, 2) *administración de la tarea*, 3) *administración de la comunicación* ó 4) *otro*. El nivel de *tarea* incluye a los enunciados mediante los cuales se está realizando la tarea y acercan a los participantes hacia la meta; p. ej. *¿Puedes mover la estufa hacia la izquierda?* Los enunciados cuyo propósito es planear, organizar, dirigir o controlar la actividad y no tanto desarrollarla en sí, se consideran parte de la *administración de la tarea*; un ejemplo es *Hemos terminado la tarea*. La *administración de la comunicación* consiste en mantener, reforzar o reestablecer la comunicación entre los participantes; p. ej. *¿Perdón?* Con la etiqueta de *otro* se marca cualquier otro nivel informativo que no sea etiquetable con los primeros tres.

La *función hacia delante* es una colección de actos que constituyen una extensión a la noción de *acto de habla* de Searle; son actos que introducen información y generan efectos e interacciones hacia el futuro del diálogo. En cambio, la *función hacia atrás* agrupa a los tipos de actos que generan relaciones de un enunciado con otros anteriores. La tabla 3.1 presenta las etiquetas de DAMSL para los actos de diálogo de la *función hacia delante* y la tabla 3.2 las de la *función hacia atrás*.

DAMSL ha sido usado en trabajos previos del área para la etiquetación de diversos corpora; p. ej. de TRAINS (Allen *et al.*, 1996) y Switchboard (Jurafsky *et al.*, 1997; Godfrey *et al.*, 1992); a su vez, los etiquetajes de éste último son usados por Shriberg *et al.* (1998) para analizar y modelar la relación entre entonación y actos de diálogo.

Categoría	Etiqueta	Descripción
<i>Statement</i>	<i>assert</i>	Se asevera una proposición
	<i>reassert</i>	Se reafirma una aseveración previamente expresada
	<i>other-statement</i>	Cualquier otro tipo de aseveración que no pueda considerarse <i>assert</i> o <i>reassert</i>
<i>Influencing-addressee-future-action</i>	<i>open-option</i>	El hablante intenta producir una acción futura del oyente pero sin que éste esté obligado a cumplirla
	<i>action-directive</i>	Se ordena al oyente la realización de una acción no verbal
<i>Info-request</i>	<i>info-request</i>	Se obliga al oyente a proporcionar información, pudiendo ésta darse en modalidad hablada o gráfica
<i>Committing-speaker-future-action</i>	<i>offer</i>	Si el oyente la acepta, el hablante se obliga a realizar una acción (verbal o gráfica)
	<i>commit</i>	El hablante se obliga a realizar una acción no verbal
<i>Conventional</i>	<i>opening</i>	Se da inicio a un diálogo o a una tarea
	<i>closing</i>	Se da terminación a un diálogo o a una tarea
<i>Explicit-performative</i>	<i>perform</i>	Se ejecuta una acción en virtud de expresar un enunciado (como agradecer, disculparse, etc.)
<i>Exclamation</i>	<i>exclamation</i>	Se expresa un estado emotivo
<i>Other-forward-function</i>	<i>other-forward-function</i>	Cualquier otro acto de la <i>función hacia delante</i> que no sea etiquetable con las etiquetas anteriores

Tabla 3.1 Actos de diálogo de la *función hacia delante* en DAMSL (Allen y Core, 1997).

Categoría	Etiqueta	Descripción
<i>Agreement</i>	<i>accept</i>	Aceptación total respecto a algún enunciado previo del interlocutor
	<i>accept-part</i>	Aceptación parcial respecto a algún enunciado previo del interlocutor
	<i>maybe</i>	Incertidumbre acerca de algún enunciado previo del interlocutor
	<i>reject-part</i>	Desacuerdo parcial respecto a algún enunciado previo del interlocutor
	<i>reject</i>	Desacuerdo total respecto a algún enunciado previo del interlocutor
	<i>hold</i>	Se pospone la retroalimentación a algún acto de diálogo previo del interlocutor
<i>Understanding</i>	<i>Signal-non-understanding</i>	El hablante expresa que no escuchó o no entendió el enunciado emitido previamente por su interlocutor
	<i>acknowledge (Signal-understanding)</i>	Se expresa entendimiento de un enunciado previo del interlocutor
	<i>repeat-rephrase (Signal-understanding)</i>	Se repite o se parafrasea algún enunciado previo del interlocutor
	<i>completion (Signal-understanding)</i>	Se completa o complementa algún enunciado previo del interlocutor
	<i>correct-misspeaking</i>	Se corrige algún enunciado previo del interlocutor que, desde el punto de vista del oyente, contenga algún error
<i>Answer</i>	<i>answer</i>	Se responde una solicitud de información del interlocutor
<i>Information-relation</i>	***	Nota: no es un acto de diálogo, sino una representación de la relación de un enunciado con alguno o varios previamente expresados

Tabla 3.2 Actos de diálogo de la *función hacia atrás* en DAMSL (Allen y Core, 1997).

DAMSL es uno de los esquemas pioneros en el área y ofrece la ventaja de ser altamente flexible y adaptable a necesidades particulares de etiquetación. Su desventaja principal es que no toma en cuenta estructuras informativas de mayor alcance que el enunciado, lo cual dificulta la obtención de consistencias satisfactorias en los etiquetajes.

3.3 Esquema de VERBMOBIL

Uno de los objetivos principales del proyecto VERBMOBIL fue la creación de un sistema de reconocimiento y traducción automáticos del habla y de administración de diálogo, aplicable a la planeación de citas de negocios; por ello, el alcance de su esquema de etiquetaje de actos de diálogo se limitó a los actos que ocurren con una alta frecuencia en este dominio específico. En la tabla 3.3 se presentan algunas de las etiquetas de mayor jerarquía de este esquema, jerarquizadas por Jurafsky (2006).

Etiqueta	Ejemplo
<i>thank</i>	<i>Thanks</i>
<i>greet</i>	<i>Hello Dan</i>
<i>introduce</i>	<i>It's me again</i>
<i>bye</i>	<i>Allright bye</i>
<i>request-comment</i>	<i>How does that look?</i>
<i>suggest</i>	<i>From thirteenth through seventeenth June</i>
<i>reject</i>	<i>No Friday I'm booked all day</i>
<i>accept</i>	<i>Saturday sounds fine</i>
<i>request-suggest</i>	<i>What is a good day of the week for you?</i>
<i>init</i>	<i>I wanted to make an appointment with you</i>
<i>give_reason</i>	<i>Because I have meetings all afternoon</i>
<i>feedback</i>	<i>Okay</i>
<i>deliberate</i>	<i>Let me check my calendar here</i>
<i>confirm</i>	<i>Okay, that would be wonderful</i>
<i>clarify</i>	<i>Okay, do you mean Tuesday the 23rd?</i>
<i>digress</i>	<i>[we could meet for lunch] and eat lots of ice cream</i>
<i>motivate</i>	<i>We should go to visit our subsidiary in Munich</i>
<i>garbage</i>	<i>Oops, I-</i>

Tabla 3.3 Los 18 actos de diálogo de mayor nivel de VERBMOBIL-1 (Jekat *et al.*, 1995), jerarquizados por Jurafsky (2006).

Dado que el proyecto VERBMOBIL tiene un dominio de aplicación muy especializado, su esquema de etiquetaje también lo es y no resulta aplicable a otros dominios; sin embargo, resultó de gran utilidad para la implementación de sistemas de administración de diálogo y de traducción automática de lenguaje hablado.

3.4 Esquema de Levin *et al.*

En el trabajo de Fernández y Picard (2002), se usó el esquema desarrollado por Levin *et al.* (1998), que maneja tres niveles: 1) *acto de habla* (basado en DAMSL y Switchboard); 2) *juego de diálogo*, definido por iniciativa e intención y 3) *actividad*, definido dentro de unidades de tópico. Los actos de habla se agrupan en 8 categorías, presentadas en la tabla 3.4, que se dividen en subcategorías. Fernández y Picard usan las 8

categorías para analizar y modelar la relación entre entonación y actos de diálogo en Español utilizando datos del corpus CallHome (Finke, Lapata *et al.*, 1998).

Categoría	Descripción
<i>Questions</i>	Describe actos que siguen la forma o la intencionalidad de una pregunta
<i>Answers</i>	Describe primariamente respuestas a preguntas <i>sí-no</i>
<i>Agreement / Disagreement</i>	Describe actos usados para aceptar o rechazar afirmaciones hechas por el otro hablante
<i>Discourse Markers</i>	Subsuma <i>backchannels</i> (actos usados primariamente para expresar entendimiento o atención)
<i>Forward Functions</i>	Incluye actos tales como exclamaciones, disculpas, fórmulas de cortesía, agradecimientos, etc.
<i>Control Acts</i>	Incluye actos que esperan una acción por parte del oyente o del hablante; p. ej. órdenes, peticiones, prohibiciones, etc.
<i>Statements</i>	Incluye afirmaciones que pueden expresar o no una opinión
<i>Other</i>	Otros actos de diálogo distintos de los anteriores

Tabla 3.4 Categorías de actos de diálogo en el esquema de Levin *et al.* (1998).

Desde el punto de vista lingüístico, Levin *et al.* toman en cuenta estructuras de mayor alcance que el enunciado; es decir, *el juego de diálogo* y *el segmento de tópico y la actividad*. Esto permite un etiquetaje más consistente que el obtenido al analizar solamente el enunciado, ya que se aprovecha información inherente de la estructura del diálogo. También se logra representar y analizar una mayor variedad de fenómenos del diálogo. Desde el punto de vista computacional, un esquema que produce etiquetajes consistentes facilita la creación de modelos confiables para aplicaciones en reconocimiento, clasificación y predicción.

3.5 DIME-DAMSL: Teoría

DIME-DASML es un esquema de etiquetaje de actos de diálogo basado en DAMSL. Éste último fue tomado como base y modificado para satisfacer las necesidades particulares del proyecto DIME, *Diálogos Inteligentes Multimodales en Español* (Pineda *et al.*, 2002), analizando diálogos del corpus del mismo nombre (Pineda *et al.*, 2007).

El esquema DAMSL ofrece un marco conceptual y una serie de reglas para etiquetar actos de diálogo; sin embargo, es difícil lograr una consistencia satisfactoria de los etiquetajes para usarlos confiablemente en experimentos de aprendizaje automático. Una fuente de inconsistencia de sus etiquetajes es la falta de una estructura de alto nivel que permita restringir las posibles etiquetas asignables a un enunciado. En otras palabras, el alcance de las reglas de DAMSL está limitado al análisis de enunciados individuales, con un cierto margen para analizar los enunciados previos o siguientes dentro del diálogo, pero sin incluir mayores

criterios formales en la selección de etiquetas. Esto permite un amplio espacio para elegir y combinar etiquetas, favoreciendo la flexibilidad del esquema; pero, por otra parte, se introduce un alto riesgo de baja consistencia entre los etiquetadores al permitirles una gran subjetividad.

Evolucionando de DAMSL, DIME-DAMSL (Pineda *et al.*, 2006 y 2006a) toma su conjunto de etiquetas organizadas en dimensiones y lo extiende al definir tres nociones adicionales, como se describe a continuación:

- (1) Transacción
- (2) Planos de comunicación: las *obligaciones* y el *common ground*
- (3) Cargos y abonos en equilibrio

3.5.1 Transacción

Una *transacción* está definida por un conjunto de instancias de actos de diálogo que intentan cumplir una (sub) meta dentro del diálogo. Al analizar el corpus DIME, se distinguen generalmente cuatro fases en la transacción: 1) *especificación de la intención*, 2) *interpretación de la intención*, 3) *satisfacción de la intención* y 4) *interpretación de la acción*. En la *especificación de la intención* el hablante generalmente expresa una directiva de acción que el oyente interpreta; una vez establecido el acuerdo entre los interlocutores acerca de la intención especificada, el responsable realiza una acción (generalmente gráfica), cuyo resultado es evaluado y, en su caso, aceptado, por el interlocutor. El resultado de la acción debe cumplir también las restricciones definidas en el dominio. Cada una de las fases puede contener subdiálogos si se requiere eliminar la ambigüedad o la vaguedad, o bien, para lograr que la intención se satisfaga incrementalmente y que se cumplan las restricciones del dominio.

La mayoría de las transacciones en el corpus DIME inician con una directiva de acción o con una oferta y terminan con la aceptación de la acción realizada. También es frecuente que en la transacción se manipule solamente un tipo de objeto (mueble de cocina) mediante un tipo especial de directiva de acción (agregarlo al espacio virtual, reubicarlo, eliminarlo, etc.). Esto permite identificar con relativa facilidad el final de una transacción y el inicio de la siguiente; sin embargo, en algunos casos esto no es muy evidente, sobre todo si la tarea particular involucra la manipulación de dos o más tipos de objetos (p. ej. estufa y refrigerador). Esto podría representarse como anidamiento de transacciones; pero se ha asumido la convención de que una transacción contiene acciones donde se procesa solamente un tipo de objeto, con una directiva de acción principal a la cual pueden estar subordinadas otras pero actuando sobre el mismo tipo de objeto.

Entre las formas de terminación de una transacción, una frecuente es la cancelación de la misma por parte de cualquiera de los interlocutores. Esto puede obedecer a que la intención especificada no satisface sus

preferencias, o bien, a que la intención especificada o el resultado de la acción violan restricciones predefinidas del dominio.

3.5.2 Los planos de comunicación de DIME-DAMSL

Para que un participante de un diálogo orientado a tareas pueda asumir una obligación, antes debe lograr un acuerdo con su interlocutor y previamente debe existir el grado suficiente de entendimiento entre ambos. Por ello, en DIME-DAMSL las *obligaciones* conversacionales, el *acuerdo* entre los hablantes y el *entendimiento* entre ambos se conciben como capas superpuestas que constituyen el diálogo. De las tres, el nivel básico es el del *entendimiento*, al cual se sobrepone el del *acuerdo* y sobre éste último, el de las *obligaciones*. Así surge la noción de plano de comunicación en DIME-DAMSL y se asume que los actos de diálogo ocurren sobre tres niveles: *obligaciones*, *acuerdo* y *entendimiento*, agrupando éstos dos últimos en el plano del *common ground*. De este modo, la mayoría de los tipos de actos de diálogo contribuyen con una componente en el subplano del *entendimiento*, algunos menos en el subplano del *acuerdo* y muchos menos en el plano de las *obligaciones*.

El plano de las *obligaciones* es el conjunto de actos de diálogo que generan una responsabilidad, ya sea en el hablante o en el oyente, de realizar una acción que puede ser verbal o de otra modalidad; p. ej. la obligación de responder a una pregunta o la de mover un objeto. Los actos de diálogo que tienen presencia preponderante en el plano de las *obligaciones* son: el *compromiso*, la *oferta* (si ésta ha sido aceptada por el interlocutor), la *directiva de acción* y la *solicitud de información*.

En DIME-DAMSL el *common ground* es el conjunto de conocimientos, creencias y suposiciones compartidos entre los interlocutores, que se establecen, refuerzan y reconstruyen mediante tipos de actos de habla específicos, organizados en dos subplanos del plano del *common ground*: el del *acuerdo* y el del *entendimiento*. El subplano del *acuerdo* está constituido por actos de diálogo que aportan conocimiento o creencias para ser compartidos en el *grounding* de los interlocutores. El subplano del *entendimiento* está definido por actos que mantienen, refuerzan o reconstruyen el canal de comunicación. Los actos de diálogo que tienen presencia primordialmente en el *acuerdo* son: la *opción abierta*, la *oferta*, la *afirmación*, la *aceptación*, el *rechazo*, la *aceptación parcial*, el *rechazo parcial*, la *espera* y el *quizá*. Los actos de diálogo que contribuyen en el subplano del *entendimiento* son: el acuse de recibo (*acknowledgment*), el *backchannel*, la *repetición o parafraseo*, la *señal de no entendimiento*, la *corrección* y la *complementación*.

Los dos planos de DIME-DAMSL redefinen la estructura de actos de la *función hacia delante* y la *función hacia atrás* de DAMSL, apoyándose en el conjunto original de etiquetas, introduciendo algunas propias y descartando otras.

Los actos de diálogo mantienen relaciones que establecen estados de equilibrio sobre cada plano (o subplano). Estas relaciones son abordadas en la sección de *Cargos y abonos*.

Plano de las obligaciones

La tabla 3.5 presenta las etiquetas de actos de diálogo del plano de las *obligaciones*, mostrando algunos ejemplos con enunciados tomados del Corpus DIME.

Etiqueta	Acto	Descripción	Ejemplo
<i>dir-accion</i>	Directiva de acción	Se ordena al oyente a realizar una acción no verbal	<i>¿Puedes mover la estufa hacia la izquierda?</i>
<i>sol-info</i>	Solicitud de información	Se obliga al oyente a proporcionar información, pudiendo ésta darse en modalidad hablada o gráfica	<i>¿Cuáles son los tipos de muebles que tengo?</i>
<i>oferta</i>	Oferta	Si el oyente la acepta, el hablante se obliga a realizar una acción verbal o gráfica	<i>¿Quieres que desplace o traiga algún objeto a la cocina?</i>
<i>compromiso</i>	Compromiso	El hablante se obliga a realizar una acción no verbal	<i>Okey.</i>

Tabla 3.5 Etiquetas DIME-DAMSL del plano de *obligaciones*, con ejemplos del Corpus DIME.

La respuesta (*resp*) es un tipo de acto de diálogo que satisface a la solicitud de información. En el contexto particular del Corpus DIME, las acciones gráficas desarrolladas por los interlocutores en el ambiente virtual son también actos que pueden satisfacer a actos tales como directivas, ofertas o compromisos. Esta relación entre los actos de *obligaciones*, la *respuesta* y las acciones gráficas se describe ampliamente en la sección de *cargos y abonos*.

Plano del common ground

El plano del *common ground* se divide en dos subplanos: el de *acuerdo* (*agreement*) y el de *entendimiento* (*understanding*). Aunque en DAMSL existen dos dimensiones nombradas con estos mismos términos, en DIME-DAMSL su noción es distinta, ya que en DAMSL el *agreement* es estrictamente un conjunto de etiquetas que representan el grado de aceptación (o de rechazo) del hablante respecto a enunciados previos de su interlocutor y el *understanding* son las etiquetas que representan el correspondiente grado de entendimiento. En cambio, en DIME-DAMSL el *acuerdo* y el *entendimiento* se conciben como dos de las tres capas que constituyen la comunicación en el diálogo (siendo la tercera el plano de las *obligaciones*).

Subplano del *acuerdo*

La tabla 3.6 presenta las etiquetas del subplano del *acuerdo*, mostrando algunos ejemplos tomados del Corpus DIME.

Etiqueta	Acto	Descripción	Ejemplo
<i>opcion-abierta</i>	Opción abierta	El hablante intenta producir una acción futura del oyente pero sin que éste esté obligado a cumplirla	<i>Estos son los tipos de muebles que tenemos</i>
<i>oferta</i>	Oferta	El hablante ofrece realizar una acción si el oyente acepta la propuesta	<i>¿Quieres que desplace o traiga algún objeto a la cocina?</i>
<i>afirm</i>	Afirmación	Se asevera algo acerca del estado del mundo	<i>Sí... ah... es lo único que está en la pared izquierda sobre el piso</i>
<i>reafirm</i>	Reafirmación	Se expresa una aseveración previamente dicha por el mismo hablante	<i>Para que esté en el centro del del espacio de <sil> la alacena y el fregadero</i>
<i>acepta</i>	Aceptación	Se expresa aceptación respecto a algo dicho previamente por el interlocutor	<i>Sí</i>
<i>rechaza</i>	Rechazo	Se expresa desacuerdo respecto a algo dicho previamente por el interlocutor	<i>No.</i>
<i>acepta-parte</i>	Aceptación parcial	Se expresa aceptación parcial respecto a algo dicho previamente por el interlocutor	<i>Un poco más separadita</i>
<i>rechaza-parte</i>	Rechazo parcial	Se expresa desacuerdo parcial respecto a algo dicho previamente por el interlocutor	<i>mm bueno am <sil> tal cual está en la posición no</i>
<i>espera</i>	Espera	Se pospone la aceptación a algún acto previo del interlocutor	<i>¿Quieres que borre este refrigerador?</i>
<i>quizá</i>	Quizá	Se expresa incertidumbre acerca de algo dicho previamente por el interlocutor	<i>bueno par-parece que está...</i>
<i>conv-abre</i>	Convencional-abre	Se indica el inicio de un diálogo o de una tarea	<i>¿Quieres que traiga algún mueble a la cocina?</i>
<i>conv-cierra</i>	Convencional-cierra	Se indica el final de un diálogo o de una tarea	<i>Hemos terminado la tarea</i>

Tabla 3.6 Etiquetas DIME-DAMSL del plano de *common ground*, subplano del *acuerdo*, con ejemplos del Corpus DIME.

Aunque la *opción abierta* intenta inducir una acción en el oyente, no conlleva una carga de obligatoriedad para éste, de modo que puede atenderla o simplemente soslayarla; es más bien una sugerencia que orden, por lo que se considera propia del subplano del *acuerdo* y no de las *obligaciones*.

La *oferta* es un acto de diálogo que forma parte del subplano del *acuerdo* y también del plano de las *obligaciones*, ya que la generación de la obligación en el hablante depende de que el interlocutor acepte el ofrecimiento, mismo que se da en el *acuerdo*. Si el oyente rechaza la oferta, no se genera obligación en el hablante.

Las etiquetas *acepta-parte* y *rechaza-parte* podrían considerarse equivalentes y quizá subsumirse en una sola, puesto que al aceptar parcialmente una proposición, implícitamente una parte de ella está siendo rechazada y viceversa. En DAMSL, esquema antecesor de DIME-DAMSL, existen ambas etiquetas para diferenciar los

respectivos actos de diálogo con base en la forma del enunciado; es decir, si la forma es más similar a una aceptación se usa *acepta-parte* y si es más similar a un rechazo se usa *rechaza-parte*. Por ello, se decidió preservarlas en el presente esquema.

Mediante el acto de *espera* el hablante pospone la aceptación hasta que haya suficiente información; eventualmente, plantea contrapropuestas o preguntas que buscan clarificar información como reacción a enunciados previos de su interlocutor. En el corpus DIME es alta la ocurrencia de *esperas* donde el hablante solicita confirmar información antes de comprometerse a realizar una acción.

La etiqueta *quizá* se usa cuando el hablante tiene la intención de expresar incertidumbre respecto a algún hecho; es decir, cuando el contenido del enunciado es equivalente a expresiones tales como *quizá, tal vez*, etc.

Las etiquetas *conv-abre* y *conv-cierra* incluyen fórmulas de cortesía o convencionales para indicar el inicio o el final de un diálogo o de una tarea.

Los actos de diálogo que ocurren en el subplano del *acuerdo* pueden tener también componentes en el del *entendimiento* y en el plano de las *obligaciones*; los detalles se describen en la sección de *cargos y abonos*.

Subplano del *entendimiento*

El subplano del *entendimiento* sirve para administrar el flujo comunicativo del diálogo; por ello constituye su nivel básico. La tabla 3.7 presenta las etiquetas de actos de diálogo de este subplano, con enunciados típicos tomados del Corpus DIME.

Etiqueta	Acto	Descripción	Ejemplo
<i>back-channel</i>	Back-channel	Se muestra atención al enunciado del interlocutor al mismo tiempo que éste último lo está expresando	<i>Ajá</i>
<i>ack</i>	Acknowledgment	Se expresa entendimiento de un enunciado recién expresado por el interlocutor	<i>Okey</i>
<i>SNE</i>	Señal de no entendimiento	El hablante expresa que no escuchó o no entendió el enunciado emitido previamente por su interlocutor	<i>¿Perdón?</i>
<i>repet-refr</i>	Repetición o parafraseo	Se repite o se parafrasea un enunciado expresado previamente por el interlocutor	<i>¿Este mueble... aquí?</i>
<i>complement</i>	Complementación	Se complementa algo dicho previamente por el interlocutor	<i>...a l[a] izquierda</i>
<i>correccion</i>	Corrección	Se corrige algo dicho previamente por el interlocutor	<i>No hay extractores para los refrigeradores</i>

Tabla 3.7 Etiquetas DIME-DAMSL del plano de *common ground*, subplano del *entendimiento*, con ejemplos del Corpus DIME.

El *back-channel* y el *acknowledgment* expresan que el sujeto está poniendo atención a su interlocutor; tienen el propósito de hacer evidente que el sujeto está entendiendo lo expresado por su interlocutor. Se distingue entre uno y otro por el hecho de que el *back-channel* ocurre al mismo tiempo que el interlocutor está hablando, mientras que el *acknowledgment* se da cuando éste ha concluido su enunciado.

Las interacciones de los actos del subplano del *entendimiento* con los del *acuerdo* y los del plano de las *obligaciones* se describen en la sección de *cargos* y *abonos*.

3.5.3 Cargos y abonos

Al participar en un diálogo orientado a tareas, los hablantes aceptan implícitamente cumplir ciertas reglas generadas por convenciones sociales preexistentes; p. ej. toda solicitud de información debe ser atendida con una respuesta, toda directiva de acción debe atenderse con la realización de la acción especificada, toda afirmación debe atenderse con alguna señal de entendimiento o de acuerdo, etc. Esto sugiere la existencia de obligaciones conversacionales que constituyen un protocolo que puede modelarse mediante un conjunto de pares de actos de diálogo; en cada par, un tipo de acto genera un desequilibrio y otro tipo de acto reestablece el equilibrio. En DIME-DAMSL este modelo se define como la estructura de *cargos* y *abonos*, existiendo una para cada plano: para el de las *obligaciones* y para el del *common ground*. Las tablas 3.8 y 3.9 presentan estas estructuras, tomadas de Pineda *et al.* (2006a).

Cargo	Abono
<i>sol-inf</i>	<i>resp</i>
<i>dir-accion</i>	<i>acción</i>
<i>compromiso</i>	<i>acción</i>
<i>oferta</i>	<i>acción</i>

Tabla 3.8 Pares equilibrados en el plano de las *obligaciones* (Pineda *et al.*, 2006a).

El término *acción* refiere a los posibles actos de diálogo con modalidad gráfica que pueden realizarse en el ambiente virtual del Corpus DIME; p. ej. agregar un nuevo objeto (mueble de cocina), reubicarlo, eliminarlo, etc. Estos actos son descritos en la sección de DIME-DAMSL *Detallado*. La *dir-accion* obliga a una acción por parte del interlocutor, mientras que el *compromiso* y la *oferta* obligan al mismo hablante. En un entorno que maneja la modalidad gráfica, algunas solicitudes de información pueden parecer directivas de acción porque la información solicitada es de modalidad gráfica y porque se requiere que el interlocutor realice acciones gráficas para entregarla. La terminación de una transacción por cancelación, generalmente mediante un *rechaza*, constituye también una forma de abono para el cargo constituido por una *dir-accion*, un *compromiso* o una *oferta*.

Cargo	Abono
<i>sol-inf</i>	<i>Acto de acuerdo + afirm</i>
<i>dir-accion</i>	<i>Acto de acuerdo</i>
<i>oferta</i>	<i>Acto de acuerdo</i>
<i>opcion-abierta</i>	<i>Acto de acuerdo</i>
<i>afirm</i>	<i>Acto de acuerdo</i>
<i>reafirm</i>	<i>Acto de acuerdo</i>
<i>Acto de entendimiento</i>	<i>Enunciado siguiente (o actual)</i>
<i>Señal de no entendimiento</i>	<i>Enunciado siguiente que atiende a la señal</i>

Tabla 3.9 Pares equilibrados en el plano del *common ground* (Pineda *et al.*, 2006a).

El término *acto de acuerdo* refiere a cualquiera de los seis pertenecientes a la dimensión del mismo nombre (*agreement*) de DAMSL de Allen y Core; es decir, *acepta*, *rechaza*, *acepta-parte*, etc. *Acto de entendimiento* refiere a cualquiera de los pertenecientes al subplano del *entendimiento*, exceptuando a la señal de no entendimiento.

En el *common ground*, un cargo generado por una solicitud de información se abona con una combinación de *acto de acuerdo* y afirmación porque en una situación ideal se da primero la aceptación (explícita o implícita) de la solicitud y después se da la información solicitada. En otras situaciones, en lugar de la aceptación pueden ocurrir otros tipos de actos que pospongan o rechacen la solicitud. Las *dir-accion* y las *ofertas* deben ser acordadas entre los interlocutores antes de generar alguna obligación; la *opcion-abierta* es una sugerencia que puede ser aceptada o no por el interlocutor; la *afirm* y la *reafirm* necesitan una aceptación (explícita o implícita) del oyente. La mayoría de las acciones con modalidad gráfica introducen información al cuerpo de conocimiento compartido por los interlocutores, por lo cual su naturaleza es comparable a la de la *afirmación* y generan cargos y abonos en forma similar. Por otra parte, los *actos de entendimiento* se abonan por el simple hecho de ser atendidos por el interlocutor.

Para simplificar el etiquetaje, se estableció la convención de etiquetar sólo los actos que estén marcados explícitamente en los enunciados; esto ocurre con etiquetas tales como el *ack*, el *acepta* de un interlocutor para responder una *sol-inf* o el *acepta* que se da a una afirmación. Otra convención es la noción de *segmento*, que consiste en el agrupamiento de dos o más elocuciones que pertenecen a un turno de un hablante y que constituyen un mismo acto de diálogo.

Asimismo, en el proceso de etiquetaje también se pueden usar las etiquetas del *estado comunicativo*; en caso de que el enunciado sea inaudible o ininteligible para el oyente, se aplica la etiqueta *uninterp*. Si el enunciado constituye un monólogo, donde el sujeto habla para sí mismo sin la intención de que su interlocutor tome en cuenta lo expresado, se usa la etiqueta *mono* (p. ej. *Eh <ruido> <sil> <no-vocal> el lavatrastes; no hay lavatrastes*). Si el hablante deja inconcluso un enunciado, se usa la etiqueta *abandonada* (p. ej. *Que me traiga no que me m[uevas] que nuevas el...*). Al usar estas etiquetas no se anota cargo ni abono.

Con base en las nociones de inicio y fin de transacción, segmento, cargos y abonos, planos de comunicación y actos de diálogo de DIME-DAMSL, se produce un etiquetaje denominado *DIME-DAMSL Preliminar* usando un formato predefinido sobre una hoja de cálculo de Excel. La tabla 3.10 presenta un ejemplo de etiquetaje de un fragmento de un diálogo tomado del Corpus DIME. La columna # indica el número identificador de enunciado, la columna *T* especifica el turno del hablante (*U* de *usuario* y *S* de *sistema*), la columna *Enunciado* es la transcripción ortográfica (transliteración) del mismo; las columnas denominadas *Cgos* y *Abs* especifican los *cargos* y *abonos*, respectivamente, correspondientes a los planos de las *obligaciones* y del *common ground*, ésta última dividida en *acuerdo* y *entendimiento*. En la columna de *Tipo de acto de diálogo*, dividida en *obligaciones* y *common ground* se muestran las etiquetas de actos de diálogo de DIME-DAMSL.

#	T	Enunciado	Obligaciones		Common Ground				Tipo de acto de diálogo	
			Cgos	Abs	Acuerdo		Entendim.		Obligaciones	Common Ground
					Cgos	Abs	Cgos	Abs		
81	U	Eh <sil> bien eh quisiera volver a ver el catálogo de refrigeradores	81		81				<i>sol-inf</i>	<i>sol-inf</i>
82	S	Okey				81				<i>acepta</i>
83		<no-vocal> Esos son los refrigeradores con que contamos		81	83				<i>resp</i>	<i>opción-abierta, despliega</i>
84	U	Bien				83				<i>acepta</i>
85		Okey[y] va[mos] eh vamos a cambiar el refrigerador por favor	85		85				<i>dir-acción</i>	<i>dir-acción</i>
86		Necesitaría quitar el refrigerador éste <sil> y cambiarlo por éste	85		85				<i>dir-acción</i>	<i>dir-acción</i>
87	S	Okey								<i>ack</i>
88		¿Quieres que borre este refrigerador?	88		88		88		<i>sol-inf</i>	<i>espera, repet-refr</i>
89	U	Sí		88		88		88	<i>resp</i>	<i>acepta</i>
90	S	¿Así está bien ?	90	85	90	85			<i>sol-inf, acción-gráf</i>	<i>acepta, acción-gráf</i>
91	U	Sí		90		90			<i>resp</i>	<i>acepta</i>
92	S	Okey								<i>ack</i>
93	U	Ahora coloca éste por favor	93		93				<i>dir-acción</i>	<i>dir-acción</i>
94	S	<no-vocal> Okey				93				<i>acepta</i>
95		¿Dónde quieres que lo coloque?	95						<i>sol-inf</i>	<i>espera</i>
96	U	Aquí		95	96				<i>resp</i>	<i>afirm, apt-zona</i>
97	S	Okey	97			96			<i>compromiso</i>	<i>acepta</i>
98		<no-vocal> ¿Así está bien?	98	97, 93	98				<i>sol-inf, acción-gráf</i>	<i>acción-gráf</i>
99	U	Sí		98		98			<i>resp</i>	<i>acepta</i>

Tabla 3.10 Etiquetaje DIME-DAMSL Preliminar de una transacción del Corpus DIME.

Esta transacción, constituida por 19 enunciados, tiene el propósito de sustituir un mueble, específicamente un refrigerador, ubicado en una localidad de la cocina virtual, por otro “nuevo” elegido del catálogo. La mayoría de las transacciones en el corpus son más simples que ésta, ya que se enfocan en la manipulación de una sola instancia de un mueble en la cocina virtual, siendo claramente reconocibles los límites de las cuatro fases internas de la transacción (especificación, interpretación y satisfacción de la intención e interpretación de la acción); sin embargo, la mostrada aquí hace evidente el hecho de que los límites de una transacción no siempre son claros y también hace suponer que las nociones de *transacción anidada* o *transacción compleja* podrían ser necesarias para abordar estos casos. Entre los enunciados 81 y 84 el *usuario* elige del catálogo un refrigerador que sustituirá al otro. En 85 y 86 expresa al *sistema* su intención de realizar la sustitución; los dos enunciados constituyen un mismo acto de diálogo, es decir, una directiva de acción, aunque involucrará la realización de dos acciones gráficas por parte del *sistema*: borrar un objeto y agregar otro. En particular, 86 expresa la intención de borrar el refrigerador existente para dejar libre su espacio; entre 87 y 92 se da satisfacción a esa componente de la intención. Entre 93 y 96 se desarrolla la segunda componente de la intención del *usuario*, es decir, elegir el nuevo objeto y precisar su ubicación. Finalmente, entre 97 y 99, el *sistema* coloca el refrigerador nuevo en la ubicación especificada, solicitando confirmación del resultado de su acción y recibiendo la aprobación del *usuario*.

El etiquetaje *DIME-DAMSL Preliminar* da soporte a una etiquetación de granularidad más fina de los actos de diálogo en una etapa subsecuente, denominada *DIME-DAMSL Detallado*, que se describe en la sección de apéndices.

Diálogo	Cantidad de transacciones	
	V.E.	S.C.
d01	6	6
d03	8	10
d12	10	9
d13	16	19
d14	9	10
d15	10	9
d17	27	28
d19	9	7
d21	7	7
d22	16	13
d23	8	8
d26	13	13
Totales	139	139

Tabla 3.11 Cantidades de transacciones identificadas por 2 anotadores en 12 diálogos del corpus DIME.

3.6 Consistencia de los etiquetajes

Una vez obtenidos valores de consistencia satisfactorios en la etapa de implementación del esquema DIME-DAMSL, inició la fase de aplicación del esquema, en la cual dos anotadores expertos etiquetaron 12 diálogos del corpus DIME en sus aspectos de inicio/fin de transacción, segmentos, cargos y abonos y acto de diálogo de DIME-DAMSL Preliminar. La tabla 3.11 presenta las cantidades de transacciones identificadas por cada uno de los dos anotadores. Las tablas 3.12 y 3.13 presentan los correspondientes valores Kappa. De estos diálogos no se hizo etiquetaje DIME-DAMSL Detallado porque no era necesario para los propósitos del presente trabajo.

Diálogo	Cant. enuncs.	Inicio / Fin	Segmento	Obligaciones		Common Ground			
				Cargos	Abonos	Acuerdo		Entendimiento	
						Cargos	Abonos	Cargos	Abonos
d01	116	1.00000	0.81161	0.94062	0.93668	0.89502	0.64709	1.00000	1.00000
d03	168	0.82746	0.66768	0.83901	0.75263	0.71957	0.47822	0.71238	0.77568
d12	117	0.85956	1.00000	0.96480	0.92321	0.93653	0.88224	0.88396	0.88396
d13	191	0.70285	0.89995	0.97198	0.92887	0.96220	0.85978	0.77187	0.81907
d14	137	0.94385	0.71399	0.97788	0.93654	0.90291	0.83086	1.00000	0.95443
d15	90	0.87998	0.88173	0.92694	0.87490	0.89717	0.79911	0.93004	0.91975
d17	237	0.76015	0.55671	0.90128	0.76234	0.77071	0.61389	0.66395	0.66395
d19	105	0.69534	0.70801	0.92235	0.86642	0.76835	0.71456	0.56250	0.48812
d21	69	0.86977	-0.01471	0.96098	0.87491	0.78769	0.70648	0.91868	0.91868
d22	181	0.76485	0.88544	0.88540	0.88450	0.82190	0.78490	0.79525	0.79525
d23	81	1.00000	1.00000	0.89417	0.94112	0.88230	0.82744	1.00000	1.00000
d26	210	0.60932	0.40528	0.77417	0.62250	0.71648	0.58334	0.57020	0.57016
Total	1,702	0.82609	0.70964	0.91330	0.85872	0.83840	0.72733	0.81740	0.81575
Promedios									

Tabla 3.12 Valores Kappa de etiquetajes de inicio/fin de transacción, ocurrencia de segmento, y cargos/abonos (2 anotadores, fecha de cómputo: octubre de 2006).

La tabla 3.13 presenta los valores de Kappa del etiquetaje de actos de diálogo con DIME-DAMSL Preliminar realizado por los mismos etiquetadores expertos y los correspondientes valores Kappa. Los porcentajes son la proporción de enunciados que fueron etiquetados con la misma etiqueta por ambos anotadores y los valores Kappa son la medida de la consistencia entre ambos descontándole el sesgo introducido por el azar.

Los valores de *Kappa* de las tablas 3.12 y 3.13 fueron computados mediante aplicaciones desarrolladas *ad hoc* sobre hojas de cálculo Excel y bases de datos en Access, ambos de Microsoft.

Tomando en cuenta los valores de *Kappa* para la etiquetación de los 12 diálogos, se concluye que los etiquetajes son consistentes y confiables para usarse en los análisis y modelaciones del presente trabajo.

Diálogo	Cant. Enuncs.	Porcentajes		Kappas	
		Obligs.	Common Ground	Obligs.	Common Ground
d01	116	91.4	75.9	0.89714	0.73185
d03	168	83.3	65.5	0.78998	0.60691
d12	117	94.0	88.9	0.92526	0.86814
d13	191	92.7	85.9	0.91155	0.82246
d14	137	91.2	85.4	0.89170	0.82514
d15	90	88.9	75.6	0.86806	0.71002
d17	237	84.8	70.9	0.82209	0.66321
d19	105	85.7	74.3	0.85714	0.74286
d21	69	92.8	79.7	0.91105	0.75523
d22	181	86.7	79.6	0.83928	0.74826
d23	81	90.1	87.7	0.88339	0.85026
d26	210	74.3	72.9	0.70097	0.69304
Total	1,702				
Promedios		88.0	78.5	0.85813	0.75145

Tabla 3.13 Porcentajes básicos de acuerdo y valores *Kappa* en etiquetajes de acto de diálogo con DIME-DAMSL Preliminar (2 anotadores, fecha de cómputo: octubre de 2006).

4. Representación de la entonación

Aunque en diversas teorías sobre la entonación del Español se establece que está determinada por las variaciones en la frecuencia fundamental (f_0), la amplitud (denominada también energía o intensidad) y la duración de los segmentos y suprasegmentos de la señal del habla, una serie de trabajos empíricos se han limitado a abordar las variaciones en f_0 como la variable más importante para representar los contornos entonativos. Uno de estos trabajos es el de Garrido (1991). Algunos trabajos anteriores a éste, como Rossi *et al.* (1981), Thorsen (1979, 1980) y 't Hart *et al.* (1990), se concentran únicamente en el análisis de la forma y variaciones de los contornos entonativos. En diversas definiciones que se han dado para la entonación se considera a la f_0 la única variable que la determina. Por ejemplo: *línea melódica con que se produce un mensaje*, de Alcina y Blecua (1975, pág. 452); o la que da León (1970, pág. 58): *variations mélodiques de l'enoncé, perçues par l'auditeur*. Además, los trabajos más influyentes sobre la entonación del español, como los de Navarro Tomás y Quilis se concentran en la información descriptiva de las variaciones de la f_0 . Por ello, en el presente trabajo no se usa información de las variaciones de la amplitud de la señal del habla.

4.1 Requerimientos de la representación

Para los propósitos del presente trabajo, la información entonativa necesita representarse en forma tal que permita realizar tanto análisis estadísticos como modelos de aprendizaje automático. La representación debería permitir una fácil discretización de los contornos entonativos, lo cual involucra la posibilidad de abstraer el contorno como una pieza de información manejable como un dato simple, ya sea numérico o categórico. También, la representación debería permitir un modo sencillo de hallar regularidades; es decir, evidencias que muestren correlaciones entre entonación y acto de diálogo. Además, el esquema debería ser tan libre como sea posible de subjetividades introducidas por los etiquetadores.

Los modelos para representar la entonación en lingüística y en ciencias computacionales han evolucionado a lo largo de tres líneas que se traslapan: (1) procesamiento de señales desde el punto de vista ingenieril, (2) fonética y (3) fonología. El procesamiento de señales ofrece una representación puramente física sin incorporar ningún conocimiento lingüístico; la frecuencia fundamental (f_0) y la intensidad, ambas representadas como series de valores numéricos a lo largo del tiempo son dos de las principales fuentes de información consideradas por este enfoque. La fonética acústica incorpora una componente perceptiva que intenta asegurar que la representación produzca la sensación de que se está escuchando la señal original; esta representación tampoco incluye conocimiento lingüístico. El enfoque en fonología enfatiza el conocimiento lingüístico acerca de la entonación y se apoya en los dos enfoques previamente descritos. Los modelos de los

incisos 2 y 3 pueden ser más útiles que el del 1 para los propósitos del presente trabajo porque proporcionan abstracciones que facilitan la definición de representaciones discretas.

4.2 Enfoque del procesamiento de señales

El procesamiento de señales es la metodología básica para representar y analizar la entonación, ya que permite transformar el sonido del habla en señales electrónicas y éstas en series de datos numéricos, lo cual permite la automatización de su análisis. Las series de datos numéricos describen principalmente las variaciones de la f_0 y de la amplitud de la señal a lo largo del tiempo y a partir de esas series se realizan análisis estadísticos que generan parámetros descriptivos, tales como promedios, máximos, mínimos, rangos, desviaciones estándar, etc. Otros parámetros frecuentemente usados son los que describen la presencia y duración de pausas y otros para la pendiente del contorno entonativo; éstos últimos son útiles para representar si el contorno tiene final ascendente, descendente o plano. En este enfoque no se producen etiquetajes, sino que se usan los parámetros que describen las características de los segmentos o suprasegmentos.

Si el corpus de habla del cual se obtienen los audios para analizar la entonación incluye a más de un hablante, es común la práctica de identificar y segmentar los enunciados de cada uno de ellos. Cada hablante tiene sus propios rangos de f_0 y de amplitud, y un promedio de velocidad de habla que le caracteriza. Por ello, los datos de f_0 , amplitud y duración de los enunciados se someten a un proceso de normalización, que consiste en transformarlos aritméticamente para que ajusten dentro de una escala numérica estandarizada. De esta forma, la normalización elimina las variaciones individuales de los locutores y se logra que parámetros resultantes sean estadísticamente comparables. Una vez obtenidos los parámetros estadísticos descriptivos, pueden usarse en diversas herramientas automáticas de análisis y modelación.

Algunos trabajos previos en el área aplican el enfoque del procesamiento de señales para representar la entonación. Por ejemplo, en la tabla 4.1 se presentan algunos parámetros usados por Shriberg *et al.* (1998) para describir f_0 , energía, duración, pausas, y velocidad de habla con el propósito de modelar la relación entre entonación y actos de diálogo en Inglés. En la tabla 4.2 se presentan algunos usados para el Español por Garrido (1991), quien se concentra en f_0 y duración. La tabla 4.3 muestra otros usados por Fernández y Picard (2002), también para Español, usando información de f_0 , energía y duración. En el proyecto VERBMOBIL se representa la entonación mediante un enfoque acústico-prosódico (Kompe *et al.* 1995), pues basándose en la transcripción alineada de fonemas, se computan diversos parámetros para cada sílaba, considerando a la sílaba bajo análisis, a las seis que le preceden y a las seis que le siguen; la tabla 4.4 presenta algunos de los principales parámetros implementados, que condujeron a la producción de aproximadamente 242 atributos, siendo éste el trabajo previo del área donde se usa la mayor cantidad de éstos.

Nombre del parámetro	Descripción
<i>end_f0_mean</i>	Promedio de f_0 en la región final del contorno
<i>abs_nrg_diff</i>	Diferencia entre promedios de energía RMS de las regiones final y penúltima
<i>ling_dur</i>	Duración del enunciado
<i>min10pause_count_n_ldur</i>	Número de pausas de al menos 10 <i>frames</i> (100 ms) en el enunciado, normalizado por duración de enunciado
<i>stdev_enr_utt</i>	Desviación estándar de valores de velocidad de habla en el enunciado

Tabla 4.1 Algunos parámetros usados por Shriberg *et al.* (1998)

Descripción
<i>Tiempo de inicio de la curva</i>
<i>Valor de f_0 en el inicio de la curva</i>
<i>Localización de los diferentes puntos de inflexión a lo largo del tiempo</i>
<i>Valor de f_0 en cada punto de inflexión</i>
<i>Tiempo al final de la curva</i>
<i>Valor de f_0 en el final de la curva</i>

Tabla 4.2 Algunos parámetros usados por Garrido (1991).

En Garrido (1991) se considera punto de inflexión cualquier punto de la curva en el que se produce un cambio en el signo de la pendiente de la misma, siempre que la diferencia de frecuencia con respecto a la última inflexión sea de 10 Hz o más.

Descripción
<i>Estimado sin sesgo (bias) de la varianza de la f_0</i>
<i>Diferencia entre f_0 máxima y la media muestral de f_0</i>
<i>Estimado sin sesgo (bias) de la distorsión (skewness) del contorno de energía</i>
<i>Media muestral del contorno de energía</i>
<i>Rango intercuartil del contorno de energía</i>

Tabla 4.3 Algunos parámetros usados por Fernández y Picard (2002).

Descripción
<i>Duración normalizada del núcleo de sílaba</i>
<i>F0 mínima y F0 máxima</i>
<i>Tiempo de inicio y de fin</i>
<i>Energía máxima y su posición sobre el eje de tiempo relativo a la sílaba actual</i>
<i>Promedio de energía y promedio de F0</i>
<i>Banderas para indicar si la sílaba contiene el acento léxico de la palabra o si está en posición de final de palabra</i>

Tabla 4.4 Algunos parámetros usados por Kompe *et al.* (1995).

Al no hacerse verificación perceptiva ni etiquetaje en el enfoque de procesamiento de señales, se evita la subjetividad y por lo tanto, posibles sesgos, que los anotadores podrían introducir. Por otra parte, una desventaja importante es que no se tiene la certeza de que la señal representada produzca el mismo efecto perceptivo que la señal original.

4.3 Enfoque de la Fonética: INTSINT

La etiquetación basada en fonética se produce sobre una base perceptiva y no requiere que el etiquetador interprete el fenómeno entonativo usando conocimientos lingüísticos amplios; produce una representación que considera un nivel de abstracción más alto que el del procesamiento de señales pero sin llegar al nivel del conocimiento lingüístico aplicado en los esquemas fonológicos. Generalmente la primera etapa consiste en analizar automáticamente la señal acústica para producir una estilización que posteriormente se etiqueta. En este campo existen pocos esquemas disponibles; uno de los más conocidos es INTSINT, *International Transcription System for Intonation*, de Hirst *et al.* (2000), que se describe a continuación.

INTSINT ofrece una herramienta de software, *Motif Environment for Speech*, MES (Espesser, 1999), para realizar la tarea de etiquetaje en forma semiautomática. Ésta se desarrolla en cuatro etapas: primero, se obtiene la frecuencia fundamental (f_0) de la señal acústica mediante tres de los algoritmos más conocidos para extracción de f_0 : *AMDF* (Average Magnitude Difference Function), *autocorrelación* y función *comb*; segundo, el algoritmo *MOMEL*, *Modeling Melody*, (Hirst *et al.*, 2000) busca puntos de inflexión (denominados *targets*) en el contorno entonativo y produce un contorno estilizado de la f_0 . Tercero, un etiquetador humano realiza una verificación perceptiva escuchando una versión sintetizada del contorno estilizado y observando su contorno en una representación gráfica; al etiquetador se le permite modificar manualmente la ubicación de los *targets* para ajustar el contorno estilizado de modo que éste represente a la f_0 original tan fielmente como sea posible. En la cuarta etapa una herramienta automática compara la posición relativa de cada *target* con el previo y el siguiente; la posición determina la etiqueta INTSINT que la herramienta asignará al *target* en cuestión. La verificación perceptiva aumenta las probabilidades de que la representación sea consistente con la f_0 original y de que refleje el efecto perceptivo que se produce en el oyente; además, evita la subjetividad que los esquemas fonológicos tienden a inducir en el etiquetaje entonativo.

Como ejemplo, la fig. 4.1 presenta la f_0 original del enunciado *Eh... ¿me puedes mostrar los tipos de muebles que tengo?* (d12, utt30). MES permite elegir el mejor contorno posible entre los obtenidos por los tres algoritmos. La figura muestra el obtenido mediante *AMDF*.

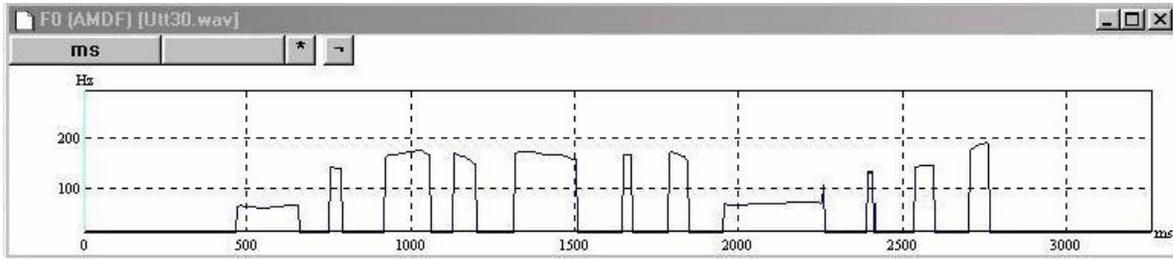


Figura 4.1 Frecuencia fundamental original.

La fig. 4.2 presenta el contorno estilizado generado por el algoritmo MOMEL (línea gruesa), sobrepuesto al contorno original (línea delgada). Los círculos representan a los *targets*.

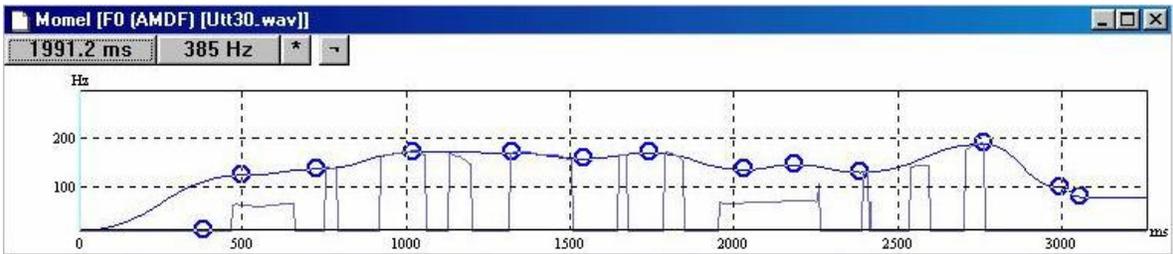


Figura 4.2 F_0 estilizada (contorno grueso) con puntos de inflexión (círculos).

MOMEL no puede garantizar una estilización perfecta y podría producir un contorno ligeramente diferente del original, como puede observarse en la fig. 4.3. Las regiones de la curva estilizada marcadas con 1, 2, 3 y 4 son distintas de la f_0 original, así que requieren un ajuste manual; la diferencia se percibe tanto en la gráfica como en la síntesis automática de habla producida por MES aplicando la curva estilizada al contenido léxico original. Es de especial importancia atender a la región final del contorno, ya que ésta contiene información de gran utilidad para el análisis de la entonación en español. En la zona 1 se observa un *target* ubicado en una región sin sonido de la f_0 original; además, la curva estilizada tiene una frecuencia superior a la de la original. En la región 2 existe un *target* por encima de la f_0 original y aparentemente se necesita un *target* adicional para lograr una buena coincidencia con la señal original. En la región 3 existen dos *targets* por encima de la f_0 y, finalmente, en la región 4 se observan dos *targets* ubicados en una zona donde la f_0 original tiene silencio.

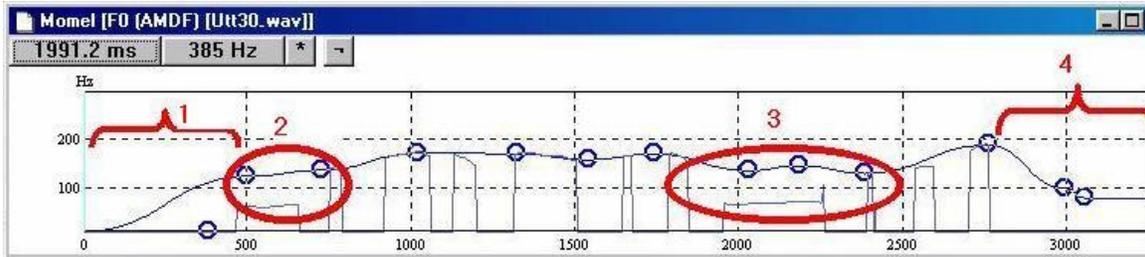


Figura 4.3 Regiones del contorno estilizado que requieren modificación.

El anotador humano hace la verificación perceptiva analizando visual y auditivamente la curva estilizada y comparándola con la original. El resultado de la curva estilizada puede ser escuchado mediante las funcionalidades de síntesis de habla que tiene MES, de modo que el enunciado puede escucharse tanto con su entonación original como con la estilización creada por MOMEL. Si la curva estilizada no suena como la original, el anotador usa las funcionalidades de MES para reubicar el (los) *target(s)* generados por MOMEL, pudiendo colocarlos arriba, abajo, a la izquierda o a la derecha de su posición; también puede eliminarlos o insertar otros nuevos. Cada vez que se hace un cambio en un *target* el anotador evalúa el resultado perceptivamente (escuchando la síntesis) para garantizar que se está logrando una estilización similar a la curva original.

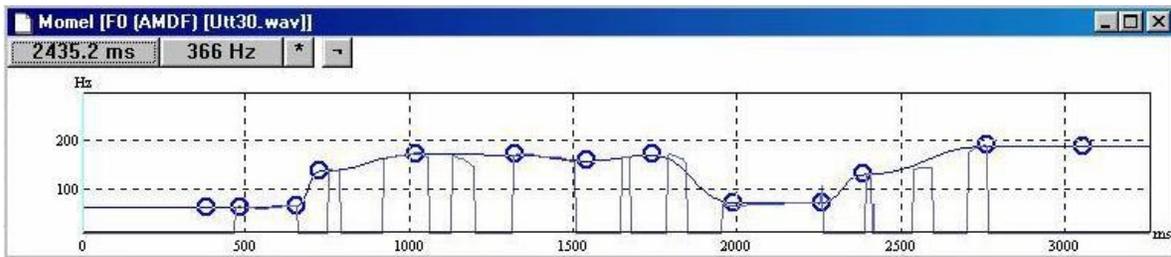


Figura 4.4 FO estilizada modificada en la verificación perceptiva.

Se usan algunas convenciones formalizada para hacer los ajustes de los *targets* en la verificación perceptiva. A continuación se enumeran las más importantes:

- 1) Identificar la primera región sonora del contorno original y verificar la alineación de los *targets* que correspondan a ésta. Por ejemplo, en la región 2 de la fig. 4.3 se debe mover hacia abajo el primero de los dos *targets* y también se debe insertar uno nuevo; los efectos de este cambio se observan en la fig. 4.4.
- 2) Identificar la última región sonora del contorno original y verificar la alineación de los *targets* que correspondan a ésta. En la fig. 4.3 no se requiere ninguna modificación para este propósito porque en su última región sonora el *target* del contorno estilizado es adecuado para representar al original.

3) Verificar la alineación de los *targets* correspondientes a las regiones sonoras internas (de la segunda a la penúltima) del contorno original. Por ejemplo, en la región 3 de la fig. 4.3 los primeros dos *targets* tienen que moverse hacia abajo.

4) Si al inicio del contorno estilizado existen uno o más *targets* que correspondan a una región de silencio del contorno original, como en la región 1 de la fig. 4.3, se conserva solamente el que se encuentre más cerca de la región sonora del contorno original y se borran los restantes. Ese *target* se debe alinear a la altura del *primero* de la región sonora del contorno original.

5) Si al final del contorno estilizado existen uno o más *targets* que correspondan a una región de silencio del contorno original, se conserva solamente el último de ellos y se borran los restantes. Por ejemplo, en la región 4 de la fig. 4.3, existen dos *targets*, de los cuales se elimina el primero. El *target* restante se debe alinear a la altura del *último* de la región sonora del contorno original.

6) Cuando se requiera insertar *targets*, la cantidad de éstos debe ser la mínima indispensable.

INTSINT tiene un conjunto de ocho etiquetas, correspondientes a ocho tonos relativos: *T* (*top*), *B* (*bottom*), *M* (*medium*), *H* (*higher*), *L* (*lower*), *U* (*upstep*), *D* (*downstep*) y *S* (*same*). Los tres primeros se denominan *tonos absolutos* y los cinco restantes, *tonos iterativos*. *T* y *B* se aplican a los *targets* más alto y más bajo, respectivamente, del contorno estilizado; *M* es el *target* ubicado en la altura media; cada una de estas tres etiquetas aparece generalmente sólo una vez a lo largo del contorno. *H* y *L* son máximos o mínimos locales, respectivamente; *U* y *D* son *targets* localizados en regiones ascendentes o descendentes, respectivamente; *S* es un *target* localizado a la misma altura que su predecesor y describe una meseta, es decir, una prolongación del tono que le antecede; cada una de estas cinco etiquetas puede aparecer más de una vez a lo largo del contorno. A cada etiqueta se le asocia una marca de tiempo (*timestamp*) que indica el instante donde se ubica el *target* a lo largo del contorno.

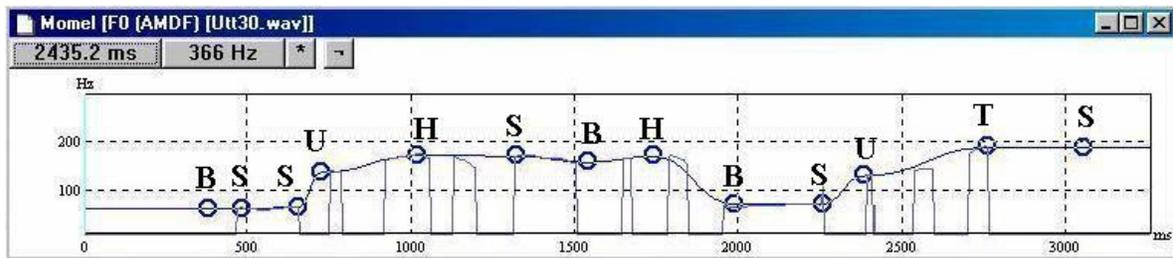


Figura 4.5 Etiquetaje INTSINT de los puntos de inflexión.

El etiquetaje obtenido se guarda en un archivo de texto en formato ETI, del cual se muestra un ejemplo en la fig. 4.6. La primera columna contiene las etiquetas INTSINT de los *targets* y la segunda los tiempos (medidos

en *frames* de 16 milisegundos) en los cuales éstos ocurren con relación al inicio del enunciado; por ello, la medida de tiempo es un dato puntual que refiere a un instante, y no a un intervalo continuo.

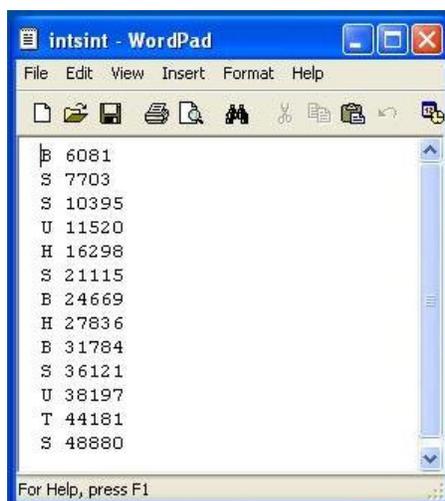


Figura 4.6 Contenido del archivo de etiquetaje INTSINT en formato *eti*.

Dado que en INTSINT la etiqueta *S* se asigna al *target* cuya *f0* sea igual (con un cierto margen de semejanza) a la del *target* precedente, la presencia de *S* indica la prolongación del tono inmediato anterior; por ejemplo, en la fig. 4.5 los primeros tres *targets* se encuentran a una altura (*f0*) muy similar entre sí, por lo cual la segunda y la tercera etiquetas son *S*.

4.4 Enfoque de la Fonología: ToBI

Desde el enfoque fonológico, uno de los esquemas más conocidos es ToBI, *Tones and Break Indices* (Silverman et al., 1992), creado para el etiquetaje prosódico del inglés norteamericano. A pesar de ser el esquema más usado para etiquetación prosódica, es frecuente que dentro de los equipos de transcripores se obtengan inconsistencias considerables en el etiquetaje; por ello, diversos autores lo consideran un sistema difícil de aplicar. ToBI maneja una serie de niveles o estratos de etiquetación, siendo los principales: las palabras, los índices de disyunción (*break indices*) y los tonos (*tones*). El proceso de etiquetación se apoya en software para observar gráficas que describen el contorno entonativo del enunciado, de modo que los estratos se alínean a lo largo del eje de tiempo. Sobre el estrato de *palabras* el enunciado se segmenta en estas unidades lingüísticas y se transcribe ortográficamente. En el estrato de *índices de disyunción*, se etiqueta, a juicio de la percepción del anotador, el grado de juntura que existe entre pares de palabras consecutivas; las etiquetas disponibles son: 0 (cero) para indicar reducción silábica por contacto de vocales entre palabras, como en casos de sinalefa; 1 (uno) para cualquier otra juntura ordinaria entre palabras y 4 (cuatro) para frase entonativa (grupo melódico). Los números 2 y 3 de los índices de disyunción dan libertad al anotador para

marcar fenómenos de disyunción de grado intermedio entre 1 y 4, tales como la frase intermedia, el grupo tónico o el grupo clítico. En el estrato de *tonos* se anotan etiquetas que describen la forma de regiones particulares del contorno entonativo del enunciado; las etiquetas disponibles son dos: *L* (*Low*) para tono bajo y *H* (*High*) para tono alto; la primera corresponde a un valle y la segunda a un pico en el contorno. El símbolo *, escrito a la derecha de alguna de las dos etiquetas, indica que el valle o el pico coinciden con una región del contorno que contiene un acento léxico. Para algunos idiomas, ToBI usa pares de estas etiquetas con el propósito de anotar tonos compuestos, por ejemplo: L*+H, L+H*, H+L*, etc. Existen también los denominados *tonos de frontera*, que describen fenómenos entonativos que ocurren en el inicio o el final de las frases entonativas (grupos melódicos); dichos tonos se etiquetan con L% y H%, donde el primero marca la caída a una frecuencia fundamental menor después de un L+H* y el segundo, la subida a una frecuencia mayor después de cualquier acento tonal.

La adaptación de ToBI al español, llamada Sp-ToBI (*Spanish ToBI*), de Beckman *et al.* (2002), maneja también otros tres estratos: *sílabas*, *misceláneo* y *de código*. En el primero se segmentan y transcriben fonéticamente las sílabas que puedan tener alofonía segmental potencialmente reveladora; por ejemplo, el ensordecimiento final de algunos segmentos. El estrato *misceláneo* sirve para etiquetar fenómenos diversos, tales como pausas dubitativas, disfluencias, risas, etc. que compliquen el análisis. El estrato de *código* sirve para marcar el dialecto y sociolecto del hablante. Sp-ToBI fue concebido como una herramienta de investigación para la prosodia del español, más que como un esquema terminado de etiquetaje; carece de un estándar que sea aceptado y usado mayoritariamente por la comunidad de investigación y a ello podrían atribuirse los bajos niveles de acuerdo inter-anotador que se han obtenido en las transcripciones de diversos corpora. Se requiere un alto grado de conocimiento lingüístico y experiencia en etiquetaje por parte del transcriptor, ya que la etiquetación resulta afectada por la subjetividad en la interpretación de los eventos entonativos; esto tiende a reducir la consistencia. Finalmente, otro motivo que dificulta la aplicación de Sp-ToBI es la falta de un software que guíe al anotador en el desarrollo de la tarea o que la automatice total o parcialmente. La baja consistencia que tiende a presentarse en los etiquetajes hace poco confiable su uso en análisis y modelaciones basados en estadística y Aprendizaje Automático.

4.5 Adaptación de INTSINT al presente trabajo

Tomando en cuenta la facilidad para generar etiquetaciones consistentes, se eligió INTSINT como esquema de representación de la entonación. Una ventaja adicional es la disponibilidad de su herramienta de software, *MES*, para realizar la tarea de etiquetaje en forma semiautomática.

El contorno entonativo de un enunciado está determinado por el rango entonativo (la diferencia entre las *f0* máxima y mínima), por la velocidad de habla del sujeto y por las posiciones relativas que éste da a los tonos a lo largo del enunciado. En el presente trabajo se requiere que los contornos entonativos de los diversos

hablantes sean estadísticamente comparables entre sí. Para ello se debe tomar en cuenta que cada sujeto tiene un rango entonativo y una velocidad de habla propios. Por ejemplo, una misma oración puede ser realizada por diversos hablantes que le impriman una entonación que se perciba igual por un oyente, y sin embargo, las realizaciones podrían presentar contornos distintos en f_0 y en duración de sus tonos. Esto sería visible al graficar las f_0 de los correspondientes enunciados. Por ello, se debe garantizar que el etiquetaje INTSINT utilizado en los experimentos esté libre de la influencia de ambos factores. Una solución frecuentemente aplicada en trabajos previos del área (e.g. Shriberg, 1998 y Garrido, 1991) es la normalización tanto de los valores de la f_0 como del tiempo; sin embargo, en este trabajo se aborda otro enfoque, mismo que se describe a continuación.

Al ser INTSINT un esquema basado en tonos relativos, se logra una abstracción del contorno que resulta independiente del rango entonativo de hablantes específicos. De esta forma, una misma oración realizada por hablantes que tengan distinto rango entonativo generaría la misma serie de etiquetajes INTSINT para los diversos hablantes.

Por otra parte, en lugar de normalizar la información de tiempo, en este trabajo se ha optado por eliminar esa información de la representación entonativa; es decir, descartar las marcas de tiempo asociadas a los etiquetajes INTSINT. De este modo, únicamente se preserva la información sobre el orden en que ocurren los tonos y, aunque se pierde la información de la duración de cada tono, se gana generalidad en la representación del contorno.

Los etiquetajes INTSINT son transformados en cadenas de etiquetas sin información de tiempo. Por ejemplo, en la fig. 4.7 se presenta la cadena correspondiente al etiquetaje presentado en la fig. 4.6. La cadena es denominada *secuencia INTSINT* o *INTSINT word*.

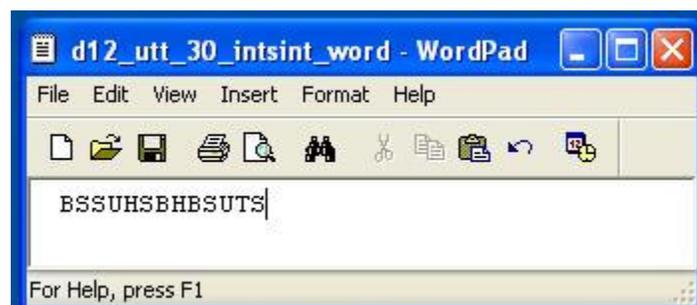


Figura 4.7 Etiquetaje INTSINT transformado en secuencia alfabética de etiquetas sin información de tiempo.

Dado que la secuencia INTSINT no contiene información de la duración de los tonos y considerando que la etiqueta *S* representa la prolongación de un tono inmediato anterior, la presencia de dos o más *S* consecutivas dentro de la secuencia puede reducirse a una sola *S*. Por ello, la secuencia INTSINT es transformada en una

nueva secuencia donde las múltiples *S* consecutivas han sido reducidas. Por ejemplo, la fig. 4.8 presenta la secuencia reducida correspondiente a la de la fig. 4.7.

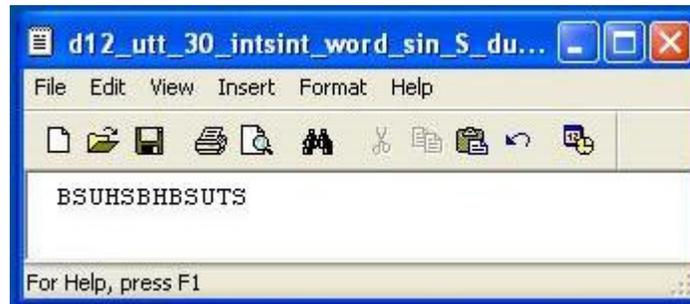


Figura 4.8 Etiquetaje INTSINT sin etiquetas *S* múltiples.

En la fig. 4.9 se presenta una muestra de secuencias INTSINT del corpus DIME. En la columna *D* se observan las secuencias después de realizar la reducción de etiquetas *S*. En la columna *A* se presentan las secuencias antes de la reducción (con información auxiliar concatenada por la izquierda para su formateo y procesamiento).

The image shows a Microsoft Excel spreadsheet with the following data:

	A	B	C	D	E	F	G	H	I
1	string_by_perl_scripts	dialog_nr	utt_number	intsint_with_unified_s					
2	d01/utt4.eti:MSTSDDDDB	1	4	MSTSDDDDB					
3	d01/utt5.eti:BSTSLH	1	5	BSTSLH					
4	d01/utt7.eti:BUTS	1	7	BUTS					
5	d01/utt8.eti:MBUT	1	8	MBUT					
6	d01/utt11.eti:BSUUSLUTLTDHLUSLHSHLSTDDLT	1	11	BSUUSLUTLTDHLUSLHSHLSTDDLT					
7	d01/utt14.eti:MTDB	1	14	MTDB					
8	d01/utt15.eti:MTBS	1	15	MTBS					
9	d01/utt17.eti:MTDBS	1	17	MTDBS					
10	d01/utt20.eti:MTLHDDDB	1	20	MTLHDDDB					
11	d01/utt27.eti:MSHBUSDLHLTS	1	27	MSHBUSDLHLTS					
12	d01/utt28.eti:MSBTLs	1	28	MSBTLs					
13	d01/utt30.eti:MB	1	30	MB					
14	d01/utt34.eti:MHDHLUSBST	1	34	MHDHLUSBST					
15	d01/utt36.eti:MMSUHLsDBUTDS	1	36	MMSUHLsDBUTDS					
16	d01/utt39.eti:MTB	1	39	MTB					
17	d01/utt40.eti:MBT	1	40	MBT					
18	d01/utt44.eti:MSSTsBS	1	44	MSTsBS					
19	d01/utt47.eti:MBTL	1	47	MBTL					
20	d01/utt48.eti:MSULHSDLUHDBUSULMTLUS	1	48	MSULHSDLUHDBUSULMTLUS					
21	d01/utt49.eti:MTDB	1	49	MTDB					
22	d01/utt50.eti:MBTS	1	50	MBTS					
23	d01/utt51.eti:BSUUTLHLHL	1	51	BSUUTLHLHL					
24	d01/utt59.eti:MTDSSLHDSDB	1	59	MTDSSLHDSDB					
25	d01/utt60.eti:BT	1	60	BT					
26	d01/utt62.eti:BTS	1	62	BTS					
27	d01/utt64.eti:BSUT	1	64	BSUT					
28	d01/utt66.eti:MUTSDBH	1	66	MUTSDBH					
29	d01/utt67.eti:BUtDLS	1	67	BUtDLS					
30	d01/utt68.eti:MR	1	68	MR					

Figura 4.9 Muestra de secuencias INTSINT de algunos enunciados del corpus DIME.

4.6 Ventajas de la representación

La representación entonativa finalmente obtenida es distinta de las usadas en trabajos previos del área. La elección y adaptación de INTSINT satisfacen los requerimientos del presente trabajo, ofreciendo también otras ventajas; la principal es que se logra una abstracción de mayor nivel que al usar una representación numérica, pues las etiquetas de tonos relativos describen al contorno de modo nominal y discreto. Esto evita la necesidad de normalizar los valores de la f_0 . Adicionalmente, al descartar la información de la duración de los tonos, se elimina la influencia de las diferencias entre las velocidades del habla de los diversos sujetos.

Al estar la secuencia INTSINT constituida por una cadena de etiquetas, es posible seleccionar y procesar diversas subcadenas que representan fenómenos entonativos ocurridos en diversas regiones del contorno entonativo. Por ejemplo, pueden producirse subcadenas para representar la región final, asociada a la noción de *tonema*. También, se tiene la posibilidad de invertir el orden de las etiquetas de la cadena (o subcadena), lo cual produce una representación que emula al inverso simétrico del contorno.

Capítulo 5. Experimentos y resultados

Dado que una de las metas de este trabajo es encontrar y representar patrones que describan la relación entre entonación y acto de diálogo, se diseñaron experimentos de aprendizaje automático como un medio de crear modelos para describir esta relación y para reconocer el tipo de acto de diálogo en instancias específicas de enunciados. Los sistemas de administración de diálogo hablado pueden aprovechar tales modelos para usar el tipo de acto de diálogo como una entrada adicional además del contenido léxico del enunciado.

Se usa J48 (Witten y Frank, 2005), un algoritmo de aprendizaje automático supervisado, para crear árboles de clasificación y regresión para reconocer el tipo de acto de diálogo a partir de la entonación, la modalidad del enunciado y otros elementos de información no léxica. J48 se asemeja al algoritmo CART (*Classification and Regression Trees*) de Breiman *et al.* (1983). Se necesita un conjunto de datos (*dataset*) para crear y probar los modelos; el *dataset* debe contener atributos, es decir variables o campos (*features*), que representen al fenómeno que será analizado, organizados en dos grupos: atributos predictores (generalmente más de un atributo) y atributo objetivo (*target*) (generalmente un sólo atributo). Los predictores son los que determinan el valor de un atributo objetivo. Algunos trabajos previos en el área muestran que el acto de diálogo está correlacionado con la entonación y, a su vez, ésta última está correlacionada con la modalidad del enunciado; por lo tanto, estos son los tres atributos principales que fueron considerados para especificar la configuración de los experimentos.

La implementación específica de atributos objetivo a partir de los datos de la etiquetación de los diálogos es un aspecto importante porque determina lo que puede ser reconocido mediante modelos automatizados y también puede influir en la confiabilidad del reconocimiento. El uso de datos de etiquetación *DIME-DAMSL detallada* produciría un *dataset* demasiado heterogéneo debido a su alto grado de detalle, reduciendo de esta forma las probabilidades de hallar patrones generales. Los modelos producidos con este *dataset* mostrarían una capacidad de generalización pobre y no serían suficientemente capaces de describir el fenómeno en forma satisfactoria. Por lo tanto, se usan los datos de la etiquetación *DIME-DAMSL preliminar* en lugar de la detallada, abstrayendo de este modo los actos de diálogo más prominentes expresados por el enunciado.

Los niveles de etiquetación del Corpus DIME que se usan para crear atributos predictores y objetivo en el *dataset* son: etiquetación *DIME-DAMSL preliminar*, etiquetación de la entonación con INTSINT, duración del enunciado, rol del hablante que emitió el enunciado, modalidad del enunciado y número de etiquetas INTSINT en su etiquetaje de entonación. Con los datos de las etiquetaciones se produjo una serie de *datasets* alternativos con diversos subconjuntos de atributos; las diferencias entre dichos *datasets* consisten en ligeras variaciones de la manera de representar la modalidad oracional y el acto de diálogo. La implementación de los atributos en los *datasets* se describe a continuación.

5.1 Atributos del acto de diálogo

La información de etiquetación DIME-DAMSL seleccionada es la de DIME-DAMSL preliminar, tanto de las obligaciones como del common ground, la cual se organiza en tres atributos categóricos, es decir, no numéricos, denominados *obligaciones*, *acuerdo* y *entendimiento*, cada uno conteniendo las etiquetas DIME-DAMSL del enunciado correspondientes al plano respectivo. Si el enunciado no expresa un acto de diálogo en alguno de los planos, las convenciones de etiquetación establecen que no se le asigne etiqueta en ese plano. En el *dataset* la ausencia de etiqueta se representa mediante el valor *x*; esto se hace para evitar que existan atributos con valor indefinido al realizar los experimentos. Cuando un enunciado está etiquetado con más de una etiqueta en un plano, p. ej. *dir-acción* y *respuesta* en las obligaciones o *afirma* y *rechaza* en el acuerdo, el valor del atributo se representa como una concatenación de las dos etiquetas: *dir-acción_respuesta* o *afirma_rechaza*. Esta representación es adecuada porque la presencia de etiquetas múltiples en un plano involucra que el enunciado expresa múltiples actos de diálogo, o bien, un acto de diálogo “mixto” o “híbrido” sobre ese plano, lo cual difiere de la ocurrencia de un acto de diálogo simple (etiqueta única); además, esta manera es más útil que usar dos atributos separados porque permite mantener un número fijo de atributos simples sin importar el número de etiquetas asignadas.

Dado que el acuerdo y el entendimiento son dos subplanos del plano de common ground, éste se representa en dos implementaciones: uno de los *datasets* contiene las etiquetaciones de common ground como un atributo simple donde las etiquetas de acuerdo y entendimiento se concatenan en forma de un sólo dato; otro *dataset* contiene los dos atributos en forma separada. Estas dos variantes permiten comparar cómo influye cada subplano en la tarea de reconocimiento del acto de diálogo.

5.2 Atributos de la etiquetación de entonación

La implementación de los atributos de la información entonativa en el *dataset* se basa en los criterios descritos a continuación. Cada hablante presenta una velocidad de habla particular, así que la forma general del contorno entonativo y el tiempo entre dos tonos consecutivos son, a su vez, determinados por este factor. Las diferencias entre las velocidades del habla pueden producir una serie de contornos de diferente longitud para un mismo enunciado emitido por diferentes hablantes, lo cual impide una comparación directa de los contornos. Algunos trabajos previos en el área han abordado este aspecto calculando una normalización de la información de tiempo, es decir, ajustando la longitud de cada contorno mediante el uso de una regla estadística que toma en cuenta los valores máximo, mínimo y promedio de las duraciones de los enunciados para producir una representación que mantiene la forma del contorno original y restringe su longitud dentro

de límites estandarizados para el *dataset*. De este modo, la normalización produce una abstracción discretizada del contorno y ésta es valiosa para buscar patrones de relación entre la entonación y el acto de diálogo.

En este trabajo no se computa la normalización del tiempo porque la representación implementada es una abstracción del contorno entonativo en un nivel más alto, cuyos fundamentos se describen a continuación. La duración de los tonos INTSINT a lo largo de un contorno, es decir, el tiempo transcurrido entre pares de puntos de inflexión, así como los tonos mismos, son los componentes que describen la forma genérica de un contorno. Un patrón entonativo es la abstracción de la forma descartando el sesgo inducido por la velocidad del habla y los rangos de f_0 de hablantes particulares. Del mismo modo que en trabajos previos del área, este trabajo asume que existe un número finito de categorías de contornos entonativos. Otra suposición, introducida por este trabajo, es que estas categorías pueden representarse como secuencias de etiquetas de INTSINT sin la información de etiquetas de tiempo. Al descartar las marcas de tiempo, preservando sólo el orden entre las etiquetas, se preserva una vista de nivel más alto del contorno; como una consecuencia de esto, la etiquetación de entonación se representa en el *dataset* como secuencias (cadenas) de caracteres alfabéticos formadas por etiquetas de INTSINT.

Las teorías de la entonación del Español establecen que la región final del contorno entonativo es la más influyente para determinar la modalidad del enunciado y el acto de habla; la región inicial también contribuye pero no tanto como la final. Dado que en el *dataset* los límites de las regiones inicial y final no están determinados, el número de etiquetas INTSINT correspondientes a esas regiones tampoco lo está; por ello, se usan subcadenas obtenidas de la secuencia de etiquetación entonativa con el propósito de determinar cuáles son las cantidades de etiquetas adecuadas para representar a cada una de las dos regiones. Son diez atributos nominales (no numéricos), cada uno en forma de cadena alfabética de hasta cinco caracteres, es decir, hasta cinco etiquetas de INTSINT. Un atributo contiene las últimas 5, otro las últimas 4, otro las últimas 3, etc.; otro atributo la primera etiqueta, otro las primeras 2, otro las primeras 3, etc.

También el número total de etiquetas INTSINT del contorno se considera como predictor y se implementa como un atributo numérico entero en el *dataset*. Parece evidente que está altamente correlacionado con la duración del enunciado, ya que a mayor duración corresponde un mayor número de tonos (etiquetas INTSINT) en el contorno y viceversa. Esta correlación se comenta a continuación.

5.3 Atributos de la duración del enunciado

La duración del enunciado, medida automáticamente en la señal del habla, se evalúa como uno de los predictores. Se implementa en el *dataset* como valor numérico con punto decimal en milisegundos.

Una medida indirecta de la duración del enunciado es el número de etiquetas INTSINT en el contorno entonativo. Ésta unidad podría ser más confiable que una unidad de medida estrictamente de tiempo como lo es el milisegundo, ya que la velocidad del habla podría producir diferentes duraciones de una misma oración enunciada por diferentes hablantes; sin embargo, el número de tonos INTSINT probablemente sería el mismo en los diferentes casos.

5.4 Atributos del rol del hablante

Aunque el rol desempeñado por el hablante en un diálogo no es un atributo entonativo, podría contribuir al reconocimiento del acto de diálogo, así que se evalúa su posible contribución y se implementa como un atributo nominal con dos valores posibles: *usuario* o *sistema*.

5.5 Atributos de la modalidad del enunciado

Algunos trabajos previos describen la relación entre la modalidad del enunciado y el contorno entonativo, presentando una alta correlación entre sus componentes, por ello se incluye la modalidad del enunciado en el presente *dataset*, primero como un atributo etiquetado manualmente y después como un dato reconocido automáticamente, es decir, como un atributo *target*. Sus valores son nominales: *interrogativa*, *declarativa*, *imperativa* u *otra*; en una implementación alternativa del *dataset* las interrogativas fueron, a su vez, divididas en *pregunta pronominal*, *pregunta sí-no* e *interrogativa genérica*.

5.6 Análisis estadístico

En los 12 diálogos que se tomaron del corpus DIME hay 116 minutos de grabación con aproximadamente 1,702 enunciados. La tabla 5.1 muestra la información detallada de cada uno de los diálogos seleccionados. A partir de ellos y dada la disponibilidad de etiquetaciones se tomaron 1,043 enunciados.

Antes de la modelación con aprendizaje automático se desarrollaron análisis estadísticos para conocer las distribuciones de los atributos y las posibles correlaciones entre ellos. Las distribuciones se analizan computando las frecuencias absolutas y relativas, el promedio, el histograma, el Pareto, el máximo y el mínimo de cada atributo.

A continuación se presentan algunas cifras obtenidas en los análisis: el *sistema* emite el 46.7% y el *usuario* el 53.3% de los enunciados. Respecto a la modalidad del enunciado, 67.2% son declarativas, 25.2% interrogativas, 1.4% imperativas y 6.2% de otras modalidades; así que la declarativa es la modalidad más

frecuente. La mayoría (80.3%) de las etiquetaciones de actos de diálogo contienen una o más de las siete etiquetas presentadas en la tabla 5.2. Los análisis muestran que la ocurrencia de actos de diálogo de entendimiento tiene una baja frecuencia, así que no es factible construir un modelo para reconocimiento de estos actos de diálogo con los datos disponibles.

Diálogo	Duración		Turnos	Enunciados
	Segundos	Minutos		
d01	495.6	8.3	63	116
d03	412.9	6.9	114	168
d12	438.8	7.3	69	117
d13	837.5	14.0	105	191
d14	524.9	8.7	97	137
d15	346.3	5.8	54	90
d17	655.2	10.9	136	237
d19	608.7	10.1	62	105
d21	262.8	4.4	49	69
d22	811.1	13.5	113	181
d23	352.7	5.9	49	81
d26	1,201.4	20.0	127	210
Total	6,948	116	1,038	1,702

Tabla 5.1 Duraciones, número de turnos y de enunciados de los 12 diálogos seleccionados para los experimentos.

Los datos de etiquetación de INTSINT presentan una distribución estadística como se describe a continuación: el número promedio tonos por enunciado es 7.1, el máximo es 45, el mínimo 2 y la desviación estándar 5.5. Considerando que las dos últimas etiquetas INTSINT de la secuencia son particularmente relevantes para el reconocimiento de la modalidad del enunciado, se realizó un análisis de sus combinaciones y éste muestra que la mayoría (80.1%) están constituidas por 10 combinaciones: *BS*, *DB*, *UT*, *TS*, *TB*, *BT*, *MB*, *US*, *BH* y *TL*.

La distribución estadística de las duraciones de enunciado es: promedio, 2103.96 milisegundos; máximo, 31157.94; mínimo, 96.25 y desviación estándar, 2,697.41.

Las correlaciones se analizan computando las frecuencias absolutas y relativas de los siguientes pares de atributos: acto de diálogo *versus* modalidad del enunciado, acto de diálogo *versus* rol del hablante, modalidad del enunciado *versus* rol del hablante, y acto de diálogo del plano de obligaciones *versus* el correspondiente del common ground.

Aunque la modalidad declarativa es la más frecuente en la mayoría de los actos de diálogo, la interrogativa es la más común para algunos de ellos en particular, p. ej. para la solicitud de información y algunas de las combinaciones de ésta (*sol-inf_resp*, *sol-inf_oferta*, etc.), en 75% o mayor porcentaje; para la *espera* y sus combinaciones (*reafirma_espera*, *afirm_espera*, etc.), en 84% o más; y para la *oferta* y sus combinaciones

(*oferta_acepta*, *oferta_conv-abre*, etc.), en el 100% de sus casos. Los actos de diálogo que no pertenecen al plano del acuerdo, ocurren como interrogativas (49.9%) o declarativas (39.1%).

Etiqueta	%	% Acum.
acepta	24.2	24.2
sol-inf	15.6	39.8
resp	14.5	54.3
dir-acción	6.8	61.0
afirm	6.8	67.8
acción-gráf	6.4	74.2
compromiso	6.1	80.3
Otras etiquetas	19.7	100.0

Tabla 5.2 Análisis de Pareto de las etiquetaciones de actos de diálogo.

El análisis de acto de diálogo *versus* rol del hablante muestra que algunos actos de diálogo son expresados más frecuentemente por uno de los dos roles y también que algunos actos nunca lo son por alguno de ellos. En el plano de las obligaciones, la solicitud de información es más frecuente en el *sistema* (72.2%), la respuesta en el *usuario* (73.3%), la directiva de acción siempre (100%) es expresada por el *usuario* y el compromiso por el *sistema* (100%). En el plano del acuerdo, el acto de aceptación muestra una distribución estadística similar entre *sistema* (53.1%) y *usuario* (46.9%); en cambio, la afirmación es más frecuente en el *usuario* (71.8%); mientras que la espera, la opción abierta y la oferta son más comunes en el *sistema* (85.5, 100 y 100%, respectivamente).

Analizando la correlación entre modalidad del enunciado y rol del hablante, las declarativas se usan en una proporción similar por ambos roles: 43.9% por el *sistema* y 56.1% por el *usuario*. Las otras modalidades presentan patrones distintos: la interrogativa se usa en el 74.2% de los casos por el *sistema*; la imperativa, 92.9% por el *usuario* y la categoría *otra* en 92.9% por el *usuario*. Esto sugiere que el rol del hablante influye en la modalidad de los enunciados que expresa.

Ciertos actos de diálogo en cada uno de los planos DIME-DAMSL muestran una clara correlación estadística con el correspondiente acto en el plano complementario. En el plano de obligaciones, por ejemplo, 65.6% de las solicitudes de información se relacionan con el valor de *sin-etiqueta* en el plano del acuerdo; 69.2% de las directivas de acción presentan el mismo comportamiento; 97.3% de los compromisos y 100% de las ofertas están relacionados con una aceptación. Algunas combinaciones de etiquetas de obligaciones también muestran correlaciones con el plano de acuerdo: 81.3% los casos de *dir-acción_respuesta* se relacionan con *acepta-parte*; 62.5% de *sol-inf_respuesta* se relacionan con *sin-etiqueta* en el acuerdo. Desde el punto de vista del plano del acuerdo, 91.3% de las esperas se relacionan con solicitud de información en las obligaciones; 68.1% de las opciones abiertas, con respuesta; 65% de las aceptaciones parciales, con la combinación de etiquetas *dir-acción_respuesta* y, 64.7% de las reafirmaciones, con respuesta. La correlación con el valor de

sin-etiqueta refleja el hecho de que los actos de diálogo expresados por algunos enunciados tienden a contribuir predominantemente en sólo uno de los planos.

Los resultados estadísticos muestran patrones en el *dataset* que pueden usarse para guiar la implementación de modelos de reconocimiento y también describen por anticipado cómo podrían interactuar los atributos en el modelo resultante. Los resultados de las correlaciones muestran que el rol del hablante y el tipo de acto de diálogo del plano complementario pueden ser útiles para reconocer el acto de diálogo en un plano determinado.

El principal atributo *target* es el acto de diálogo y los otros atributos se usan y se evalúan como predictores de éste. Por otra parte, la modalidad del enunciado también necesita manejarse como *target* porque es una entrada requerida para el reconocimiento del acto de diálogo y no puede obtenerse de otra forma. Por lo tanto, se implementan dos grupos de modelos de reconocimiento: uno para acto de diálogo y otro para modalidad del enunciado; la salida del segundo es una de las entradas (predictores) para el primero. Los modelos para acto de diálogo son dos: para obligaciones y para common ground; ésto se debe a que un enunciado puede comunicar actos de diálogo en ambos planos simultáneamente. El modelo para common ground maneja solamente el subplano de acuerdo y no el de entendimiento porque de éste último no se tiene un número suficientemente grande de instancias en el *dataset*. Los principales atributos a evaluar como predictores para la modalidad del enunciado son los obtenidos de la etiquetación de entonación y de la duración del enunciado, entre otros que se detallan en la sección de resultados. En ambos modelos, la selección de atributos para ser evaluados como predictores se basa en aquellos sugeridos tanto por diversas teorías en el área como por análisis estadísticos de los datos empíricos disponibles. Dado que el acto de diálogo de un enunciado está disponible como una combinación de tres atributos separados, es decir, valores nominales en las obligaciones, el acuerdo y el entendimiento, se implementan y evalúan varios modelos para acto de diálogo usando combinaciones de uno o dos de estos atributos como predictores y el restante como *target*. Los experimentos se desarrollan usando la herramienta de software denominada WEKA, Waikato Environment for Knowledge Analysis (Witten y Frank, 2005).

5.7 Configuración de los experimentos

El principal atributo *target* es el tipo del acto de diálogo; los otros atributos se evalúan como predictores. Por otra parte, la modalidad del enunciado también necesita establecerse como un *target* especial porque es un dato de entrada requerido para el reconocimiento del acto de diálogo y en una aplicación real no podría obtenerse en otra forma. Por lo tanto, también se implementa un modelo de reconocimiento para modalidad del enunciado y su salida se usa como una de las entradas para el modelo de acto de diálogo.

La tabla 5.3 presenta el diccionario de datos de los atributos implementados a partir de los datos del corpus para los experimentos. La primera columna presenta los nombres de los atributos en los modelos; la segunda, es una descripción del atributo; la tercera explica por qué se evalúa el atributo. Finalmente, la columna de la extrema derecha especifica si el atributo se usa como *predictor (P)*, *target (T)* o ambos (*T/P*); el valor *T/P* especifica que el atributo se usa como *target* en un modelo en particular y como un *predictor* en otro. No se usan atributos de información léxica en los experimentos.

Los predictores se eligieron con base en diversas teorías lingüísticas y en los análisis estadísticos de los datos empíricos. El algoritmo de aprendizaje automático evalúa la contribución de cada predictor a la tarea de reconocimiento y eventualmente podría descartar cualquiera de ellos si no es útil para el reconocimiento.

Se evalúan tres configuraciones experimentales. En cada configuración se crean dos modelos para acto de diálogo: uno para obligaciones y uno para common ground. El motivo es que un enunciado puede transmitir simultáneamente actos de diálogo en ambos planos; también, porque se supone que ciertos actos de diálogo en un plano determinado podrían determinar los actos que ocurren en el otro. Dado que el plano del common ground está estructurado en dos subplanos (acuerdo y entendimiento), su modelación tiene que tomar esto en cuenta. El análisis estadístico muestra que los enunciados con etiquetas en entendimiento sin una etiqueta simultánea en el subplano de acuerdo son muy poco frecuentes, así que para propósitos experimentales la etiqueta de entendimiento se concatena con la del acuerdo para crear el atributo *commgr* en ciertos modelos. A continuación se describen las tres configuraciones experimentales.

La **Configuración 1** se enfoca en los atributos entonativos del enunciado. Las teorías sobre la entonación del Español establecen que ésta se relaciona estrechamente con la modalidad del enunciado y que también está asociada con ciertos tipos de actos de diálogo, así que en algunos modelos la entonación real del enunciado (manualmente anotada) se incluye como uno de los predictores en lugar de las etiquetaciones INTSINT. En una aplicación del mundo real, la modalidad del enunciado necesita ser determinada a partir de la información del habla, así que se genera un modelo para producir una modalidad etiquetada automáticamente y después su salida se usa en lugar de la modalidad real como una de las entradas para los modelos de acto de diálogo.

La **Configuración 2** asume que ciertos atributos *no* entonativos pueden contribuir tanto al reconocimiento de la modalidad del enunciado como al del tipo de acto de diálogo, así que el modelo para *optimal_pred_mood* evalúa la contribución de *speaker_role*; esto se justifica por una correlación estadística entre *speaker_role* y *utt_mood*. Otro modelo evalúa la contribución de *speaker_role* y de la etiqueta del acto de diálogo complementario para reconocer el tipo de acto de diálogo. El acto de diálogo complementario es *commgr* para *obligations* y *obligations* para *commgr*. La configuración asume que el tipo de acto de diálogo de uno de los dos planos ya se conoce mediante el uso de algún otro método, tal como otro algoritmo de aprendizaje automático, modelos de lenguaje, etc. Este análisis permite además estudiar el comportamiento del esquema de DIME-DAMSL.

Atributo	Descripción	Por qué se evalúa	P ó T
<i>first_1</i>	La primera etiqueta INTSINT del enunciado	La región inicial del contorno entonativo contribuye al reconocimiento de la modalidad del enunciado; cada uno de los tres atributos es evaluado para determinar cuál es útil	P
<i>first_2</i>	Las primeras dos etiquetas INTSINT del enunciado		P
<i>first_3</i>	Las primeras tres etiquetas INTSINT del enunciado		P
<i>last_2</i>	Las últimas 2 etiquetas INTSINT del enunciad	Experimentos preliminares mostraron que es altamente contribuyente al reconocimiento de la modalidad del enunciado porque contiene el tonema	P
<i>utt_mood</i>	La modalidad del enunciado etiquetada manualmente	Se relaciona con el contorno entonativo y quizá con el acto de diálogo	P
<i>predicted_mood</i>	Se obtiene mediante un modelo complementario de reconocimiento antes del reconocimiento del acto de diálogo	Se usa una modalidad reconocida automáticamente en lugar de la anotada manualmente porque esta implementación es más similar a una aplicación de mundo real	T/P
<i>optimal_pred_mood</i>	Se obtiene mediante un modelo complementario de reconocimiento antes del reconocimiento del acto de diálogo. Funciona mejor que <i>predicted_mood</i> al usar otros predictores	La tasa de reconocimiento de tipos de actos de diálogo es mejor que al usar <i>predicted_mood</i>	T/P
<i>utt_duration</i>	Duración del enunciado en milisegundos (no está normalizada)	Experimentos preliminares sugieren que podría contribuir al reconocimiento del acto de diálogo	P
<i>number_of_tones</i>	Número de etiquetas INTSINT de la etiquetación entonativa completa del enunciado	Es una abstracción de la duración del enunciado y no necesita un proceso de normalización	P
<i>speaker_role</i>	Rol del hablante en el diálogo, ya sea <i>Sistema</i> o <i>usuario</i>	Las estadísticas sugieren que el rol del hablante está correlacionado con el tipo de acto de diálogo; p. ej. <i>sistema</i> con <i>compromiso</i> y <i>usuario</i> con <i>directiva de acción</i>	P
<i>obligations</i>	Etiqueta anotada manualmente del tipo de acto de diálogo en el plano de obligaciones	Es el <i>target</i> en el modelo de obligaciones y uno de los predictores en el modelo de common ground	T/P
<i>obligations_minus1</i>	Etiqueta de acto de diálogo (manualmente anotada) de obligaciones en el enunciado <i>n-1</i> , donde <i>n</i> es el enunciado cuyo tipo de acto de diálogo es el <i>target</i>	Se desea evaluar la magnitud de su contribución al reconocimiento del tipo de acto de diálogo	P
<i>commgr</i>	Etiqueta anotada manualmente del tipo de acto de diálogo en el plano del common ground; las etiquetas de acuerdo y de entendimiento se concatenan como un sólo atributo	Es el <i>target</i> en el modelo del common ground y uno de los predictores en el modelo de obligaciones	T/P
<i>commgr_minus1</i>	Etiqueta del acto de diálogo (manualmente anotada) del common ground en el enunciado <i>n-1</i> , donde <i>n</i> es el enunciado cuyo acto de diálogo es el <i>target</i>	Se desea evaluar la magnitud de su contribución al reconocimiento del tipo de acto de diálogo	P

Tabla 5.3 Atributos implementados en el *dataset* usando datos del corpus.

La **Configuración 3** analiza la contribución del acto de diálogo previo como uno de los predictores; es decir, el tipo del acto de diálogo del enunciado previo en el diálogo para predecir el tipo correspondiente a un enunciado actual. El acto de diálogo previo se representa como dos atributos: *obligations_minus1* y *commgr_minus1*.

Todos los modelos de aprendizaje automático en este trabajo se crean usando validación cruzada de 10 subconjuntos (*10-fold cross validation*) como método de entrenamiento-prueba; es decir, el conjunto de entrada es dividido automáticamente en 10 subconjuntos por la herramienta de software de aprendizaje automático para realizar 10 pruebas, como se describe a continuación: en una primera iteración, se usa una combinación de 9 subconjuntos para entrenar al modelo y el subconjunto restante se usa para probarlo; en una segunda iteración, una segunda combinación de 9 subconjuntos (donde uno es distinto de los primeros nueve) se usa para entrenar y otro subconjunto (distinto del usado primeramente) para prueba; después, se implementa una tercera combinación de 9 subconjuntos, etc. hasta que se completan 10 iteraciones y los 10 subconjuntos han sido usados alternadamente para entrenamiento y para prueba. Este método es altamente confiable y frecuentemente usado para crear modelos de reconocimiento o clasificación.

Los modelos se evalúan calculando automáticamente su precisión (*accuracy*) y su estadístico *Kappa* (Carletta, 1996) y se selecciona el modelo óptimo considerando los valores más altos correspondientes a estos dos parámetros. En aprendizaje automático los criterios establecen que la *accuracy* es aceptable si es mayor o igual que 0.75; y *Kappa* es confiable si es mayor o igual que 0.7. Además, los *recalls*, *precisions* y *medidas F* de los tipos (clases) de actos de diálogo se calculan con base en sus correspondientes matrices de confusión. La *accuracy* se asume como el porcentaje de instancias clasificadas correctamente por el modelo durante la prueba. *Kappa* es una medida de cuán consistente es el modelo sobre el nivel del azar. El *recall* es la proporción de instancias que fueron clasificadas en una clase entre todas las instancias que realmente pertenecen a la clase; la *precision* es la proporción de instancias que realmente pertenecen a una clase entre todas aquellas que fueron clasificadas en esa clase. La *medida F* es otro estadístico comúnmente usado en aprendizaje automático para evaluar la tasa de reconocimiento para cada clase del dato *target*; se calcula con la fórmula: $F = 2 \times (Precision \times Recall) / (Precision + Recall)$.

Un árbol de clasificación contiene reglas *if-then*; por ejemplo, una regla extraída de un árbol de obligaciones podría ser:

IF common_ground=acepta
AND speaker_role=sistema
AND optimal_predicted_mood=declarativa,
THEN compromiso,

la cual expresa que si la etiqueta del common ground es *acepta* y el rol del hablante es *sistema* y la modalidad predicha óptima es *declarativa*, entonces la etiqueta en el plano de obligaciones es *compromiso*. La utilidad de cada regla de un árbol se evalúa mediante dos parámetros, denominados *soporte* y *confianza*. El soporte (*support*) es una razón calculada como el número de casos donde la premisa de la regla (el conjunto de condiciones booleanas antes del *THEN*) ocurre, dividido entre el número total de instancias del *dataset*; es decir, el soporte representa qué tan frecuentemente ocurre en el *dataset* la combinación de valores de los atributos predictores. La confianza (*confidence*) es otra proporción: el número de casos donde ocurre la regla, dividido entre el número de casos donde ocurre la premisa. Por ejemplo, si la premisa de la regla del ejemplo se cumpliera en 182 enunciados de un total de 1,043, su soporte sería $182/1,043=17.4\%$ y, si la regla se cumpliera en 103 de los 182, entonces su confianza sería de $103/182=56.6\%$.

5.8 Resultados de aprendizaje automático

Una vez que el *dataset* fue preparado y formateado para ser compatible con WEKA, se produjo una serie de árboles de clasificación para acto de diálogo y para modalidad del enunciado. Los árboles completos se presentan en la sección de Apéndices. A continuación se presentan y se analizan los resultados de cada una de las tres configuraciones.

La tabla 5.4 muestra los atributos predictores que resultaron mejores en cada modelo, así como los respectivos valores de *accuracy* y Kappa. Los predictores fueron identificados por el algoritmo J48 como contribuyentes para reconocer los *targets* correspondientes; es decir, J48 descarta automáticamente los atributos que no contribuyen al reconocimiento del *target*.

5.8.1 Configuración 1: Atributos entonativos como predictores

Los resultados sugieren que al usar sólo atributos entonativos y de duración (modelo 1.1), se produce una baja tasa de reconocimiento del tipo de acto de diálogo. El mejor reconocimiento se obtiene en el modelo 1.2, el cual usa la modalidad real (etiquetada manualmente) del enunciado, la cual no estaría disponible en una aplicación real. Un modelo más realista es el 1.4, donde se usa una modalidad etiquetada automáticamente en lugar de la real. El modelo 1.2 sugiere que la modalidad del enunciado contribuye a esta tarea. Dado que las *accuracies* y *Kappas* de los modelos 1.1, 1.2 y 1.4 son bajas, no se presentan más análisis de sus correspondientes árboles de clasificación.

Se realizó un análisis particular para el reconocimiento de la modalidad del enunciado, así que se crearon y evaluaron dos modelos: en el primero, el conjunto de etiquetas (*tagset*) tiene tres etiquetas para la modalidad interrogativa: pregunta sí/no (*preg-s/n*), *pregunta pronominal* (*preg-pron*) y pregunta genérica (*preg*), siendo la última aplicada para las interrogativas que no sean etiquetables con las primeras dos. En el segundo

modelo, se usa una sola etiqueta (*int*) para cualquier tipo de interrogativa. Los resultados muestran que al usar una sola etiqueta interrogativa se obtiene un mejor desempeño que al usar tres y el modelo 1.3 surge del primer caso.

Configuración	Modelo	Target	Predictores	Accuracy (%)	Kappa
1	1.1.A	<i>obligations</i>	<i>first_3, last_2, utt_duration</i>	37.4031	0.1571
	1.1.B	<i>commgr</i>	<i>first_3, last_2, utt_duration</i>	46.3178	0.1817
	1.2.A	<i>obligations</i>	<i>utt_mood, utt_duration</i>	47.9386	0.3264
	1.2.B	<i>commgr</i>	<i>utt_mood, utt_duration</i>	56.3758	0.3758
	1.3	<i>predicted_mood</i>	<i>first_2, last_2, utt_duration</i>	74.3049	0.4164
	1.4.A	<i>obligations</i>	<i>predicted_mood, utt_duration</i>	39.0221	0.2179
	1.4.B	<i>commgr</i>	<i>predicted_mood, utt_duration</i>	50.6232	0.2832
2	2.1	<i>optimal_pred_mood</i>	<i>last_2, first_2, speaker_role, utt_duration</i>	77.3730	0.4531
	2.2.A	<i>obligations</i>	<i>commgr, speaker_role, optimal_pred_mood, utt_duration</i>	66.2512	0.5812
	2.2.B	<i>commgr</i>	<i>obligations, speaker_role, optimal_pred_mood, utt_duration</i>	66.9223	0.5482
3	3.1.A	<i>obligations</i>	<i>speaker_role, commgr_minus1, obligations_minus1, commgr, optimal_pred_mood, first_3, number_of_tones</i>	73.4420	0.6783
	3.1.B	<i>commgr</i>	<i>commgr_minus1, first_3, obligations, speaker_role, last_2, obligations_minus1, utt_duration, first_2, number_of_tones</i>	69.0316	0.6054

Tabla 5.4. Resultados de los experimentos.

El modelo 1.3 ofrece una *accuracy* prometedora de 74.3049% y *Kappa* igual a 0.4164 para el reconocimiento de la modalidad del enunciado. Los predictores contribuyentes en esta tarea son *first_2*, *last_2* y *utt_duration*. Esta modalidad predicha (*predicted_mood*) se usa como una de las entradas para el modelo 1.4. La región final del contorno entonativo, es decir, las últimas 2 etiquetas INTSINT de la representación entonativa, se confirma como altamente contribuyente al reconocimiento de la modalidad, como se ha establecido en teorías de la entonación del Español. Además, el reconocimiento de la modalidad mejora al adicionar información de la región inicial (*first_3*), lo cual también es consistente con aseveraciones teóricas previas. El modelo de árbol completo contiene 41 reglas. La tabla 5.5 presenta las cinco reglas con mayor soporte; éstas son consistentes con las teorías de la entonación del Español, tales como la de Navarro Tomás (1948), respecto al contorno ascendente (etiquetaciones *UT*, *TS*) de algunas modalidades interrogativas y el contorno descendente o plano (*BS*, *DB*, *TB*) de las declarativas.

Núm. regla	Regla	Casos donde Ocurre la Premisa	Casos No Satisfactorios	Confianza %	Soporte %
1	IF <i>last_2</i> =BS, THEN <i>dec</i>	136	16	88.2	13.0
2	IF <i>last_2</i> =DB, THEN <i>dec</i>	107	17	84.1	10.4
3	IF <i>last_2</i> =UT, THEN <i>int</i>	98	34	65.3	9.4
4	IF <i>last_2</i> =TS, THEN <i>int</i>	94	37	60.6	9.0
5	IF <i>last_2</i> =TB, THEN <i>dec</i>	90	17	81.1	8.6

Tabla 5.5. Las 5 reglas con mayor soporte para modalidad del enunciado (modelo 1.3).

La tabla 5.6 presenta las medidas *F*, *recalls* y *precisions* de las clases *target* del modelo 1.3. No se presentan los valores para *imperativa* ni para otras clases; sus valores son menores que 0.5.

Modalidad	<i>F</i>	Recall	Precision
<i>dec</i>	0.826	0.868	0.787
<i>int</i>	0.615	0.598	0.633

Tabla 5.6. Medidas *F*, *recalls* y *precisions* de modalidad predicha optimizada (modelo 1.3).

Los resultados para modalidad del enunciado se toman en cuenta para la implementación de modelos en las configuraciones 2 y 3, tanto para modalidad como para acto de diálogo. En ambas se usa una sola clase de interrogativa para el reconocimiento de la modalidad del enunciado.

5.8.2 Configuración 2: Evaluando atributos no entonativos

El modelo 2.1 para reconocimiento de modalidad del enunciado contiene 55 reglas. Su dato *target* es la denominada modalidad predicha optimizada (*optimal_pred_mood*) y usa *last_2*, *first_2*, *speaker_role* y *utt_duration* como predictores. Los demás atributos fueron descartados automáticamente por el algoritmo J48. La *accuracy* es de 77.373% y *Kappa* es 0.4531; la tabla 5.7 presenta las medidas *F*, *recalls* y *precisions*. No se presentan los valores para *imperativa* ni para otras clases, ya que son menores que 0.5. Se obtienen estos resultados al manejar una sola etiqueta para la modalidad interrogativa. Por otra parte, si se manejan las tres etiquetas para modalidad interrogativa, la *accuracy* y *Kappa* son menores: 74.6% y 0.38825, respectivamente. Se usa el término *optimizada* para distinguir el modelo 2.1 de una modalidad predicha más simple, donde la *accuracy* y *Kappa* son más bajas.

Modalidad	<i>F</i>	Recall	Precision
<i>dec</i>	0.849	0.938	0.776
<i>int</i>	0.640	0.543	0.778

Tabla 5.7. Medidas *F*, *recalls* y *precisions* de modalidad predicha optimizada (modelo 2.1).

En la tabla 5.8 se presentan las reglas con soporte más alto para modalidad predicha optimizada, donde se observa la contribución de *speaker_role* como uno de los predictores.

Núm. regla	Regla	Casos donde Ocurre la Premisa	Casos No Satisfactorios	Confianza %	Soporte %
1	IF <i>last_2</i> =BS, THEN <i>dec</i>	136	16	88.2	11.5
2	IF <i>last_2</i> =DB, THEN <i>dec</i>	107	17	84.1	8.6
3	IF <i>last_2</i> =MB, THEN <i>dec</i>	87	2	97.7	8.1
4	IF <i>last_2</i> =TB, THEN <i>dec</i>	90	17	81.1	7.0
5	IF <i>last_2</i> =TS AND <i>speaker_role</i> =System, THEN <i>int</i>	58	9	84.5	4.7

Tabla 5.8. Las 5 reglas con mayor soporte para modalidad predicha optimizada (modelo 2.1).

En estas reglas, *last_2* es el atributo más relevante para la tarea del reconocimiento porque está presente en todas las reglas del modelo completo. *Speaker_role* es el segundo atributo más contribuyente; se encuentra en el 34.5% de los casos. El 21.8% de las reglas usan *first_2* como predictor y 14.5% usan *utt_duration*.

En el modelo 2.2 para actos de diálogo la mayor *accuracy* para reconocimiento de obligaciones se obtiene al usar estos predictores: *optimal_pred_mood*, *utt_duration*, *commgr* (usando *acuerdo* y *entendimiento* como una concatenación en un solo atributo) y *speaker_role*.

El modelo más preciso para *common ground* usa estos predictores: *optimal_pred_mood*, *utt_duration*, *obligations* y *speaker_role*.

La tabla 5.9 presenta las medidas *F*, *recalls* y *precisions* del modelo para los actos de diálogo de obligaciones y la 5.10 los valores para *common ground*. Los valores correspondientes a las etiquetas que no se presentan en la tabla son menores que 0.4.

Los análisis estadísticos muestran que ciertas modalidades de enunciado y ciertos actos de diálogo son más frecuentes en uno de los dos roles de hablante; por lo tanto, el uso de *speaker_role* como uno de los predictores incrementa los valores de *accuracy* y *Kappa* de los árboles de clasificación para modalidad del enunciado así como para acto de diálogo. Sin embargo, su influencia es un fenómeno contingente porque está determinada por las condiciones y protocolos para la creación del corpus. A pesar de esto, se espera que los diálogos prácticos de otros dominios muestren una relación similar entre rol del hablante, acto de diálogo y modalidad del enunciado.

Los árboles de clasificación muestran que la etiqueta particular de un acto de diálogo en cualquier plano de DIME-DAMSL (obligaciones o *common ground*) podría determinar o restringir las posibles etiquetas del componente del acto de diálogo del plano complementario; es por ello que la etiqueta de un plano usada como

uno de los predictores mejora las *accuracies* y las *Kappas* de reconocimiento del componente del acto de diálogo del otro plano. Esta configuración asume la suposición de que ya se ha determinado la etiqueta de alguno de los dos planos.

Obligaciones	<i>F</i>	Recall	Precision
<i>dir-acción_respuesta</i>	0.813	0.813	0.813
<i>sol-info</i>	0.789	0.725	0.865
<i>respuesta</i>	0.715	0.751	0.681
<i>compromiso</i>	0.683	0.893	0.552
<i>dir-acción</i>	0.654	0.833	0.538
<i>sin-etiqueta</i>	0.486	0.405	0.607

Tabla 5.9 Medidas *F*, *recalls* y *precisions* del modelo para actos de diálogo de obligaciones (modelo 2.2.A).

Common Ground	<i>F</i>	Recall	Precision
<i>acepta</i>	0.799	0.811	0.788
<i>sin-etiqueta</i>	0.739	0.830	0.666
<i>acepta-parte</i>	0.667	0.650	0.684
<i>opción-abierta</i>	0.661	0.872	0.532
<i>espera_repetir-refr</i>	0.577	0.593	0.561
<i>SNE</i>	0.500	0.455	0.556
<i>afirma</i>	0.411	0.403	0.419

Tabla 5.10 Medidas *F*, *recalls* y *precisions* del modelo para actos de diálogo de common ground (modelo 2.2.B).

Hay 64 reglas en el árbol de clasificación para actos de diálogo de obligaciones y 47 en el árbol para common ground.

En las reglas para obligaciones, 49.2% de ellas corresponden a *sin-etiqueta*, 18.6% a *respuesta* y 16.9% a *sol-info*; estas tres clases constituyen el 84.7% de las reglas. Las reglas muestran que el atributo más relevante para la tarea de reconocimiento es *commgr*, presente en el 100% de las reglas, seguido por *utt_duration* (en el 37.3%), *speaker_role* (en 25.4%) y *optimal_pred_mood* (8.5%). La tabla 5.11 presenta las 5 reglas con mayor soporte.

Similarmente a la configuración 1, la 2 sugiere que la información entonativa por sí misma no contribuye en forma significativa al reconocimiento del acto de diálogo en el plano de obligaciones; esto se demuestra al comparar los resultados con un modelo alternativo donde se incluyen las etiquetaciones INTSINT como predictores, obteniendo *accuracy* y *Kappa* menores: 62.7037% y 0.537, respectivamente.

En el conjunto completo de reglas para common ground, 34.1% de éstas corresponden a *sin-etiqueta*, 22.7% a *acepta*, 13.6% a *afirma* y 9.1% a *espera_repet-refr*; estas cuatro clases constituyen el 79.5% de las reglas. Cinco de las reglas con mayor soporte se presentan en la tabla 5.12.

Núm. regla	Regla	Casos donde Ocurre la Premisa	Casos No Satisfactorios	Confianza %	Soporte %
1	IF <i>commgr=acepta</i> AND <i>speaker_role=sistema</i> AND <i>optimal_pred_mood=dec</i> , THEN <i>commit</i>	182	79	56.6	17.4
2	IF <i>commgr=sin-etiqueta</i> AND <i>speaker_role=User</i> AND <i>utt_duration>1328.6875</i> , THEN <i>dir-acción</i>	155	77	50.3	14.9
3	IF <i>commgr=acepta</i> AND <i>speaker_role=User</i> AND <i>utt_duration<=2218.4375</i> , THEN <i>answer</i>	146	33	77.4	14.0
4	IF <i>commgr=sin-etiqueta</i> AND <i>speaker_role=sistema</i> AND <i>utt_duration>509.75</i> , THEN <i>sol-info</i>	127	13	89.8	12.2
5	IF <i>commgr=espera_repet-refr</i> , THEN <i>sol-info</i>	54	1	98.1	5.2

Tabla 5.11. Las 5 reglas con mayor soporte para actos de diálogo de obligaciones (modelo 2.2.A).

En el reconocimiento de actos de diálogo de common ground, *obligations* es el atributo más relevante; está presente en el 100% de las reglas del modelo 2.2.B. Éstas también incluyen: *utt_duration* (84.1% de las reglas), *speaker_role* (84.1%) y *optimal_pred_mood* (15.9%). Para comparar la contribución de los atributos exclusivamente entonativos en esta configuración, se creó un modelo alternativo usando etiquetaciones INTSINT entre los predictores, obteniendo una *accuracy* de 62.9914% y *Kappa* de 0.4717, las cuales son menores que las del modelo 2.2.B.

Como la información entonativa, especialmente *last_2*, es el atributo más contributivo para *optimal_pred_mood*, puede establecerse que este *target* es una representación indirecta de la entonación, así que la entonación puede ser ubicada por debajo del cuarto predictor más contributivo al reconocimiento del acto de diálogo en la configuración 2.

5.8.3 Configuración 3: Evaluando el acto de diálogo previo como predictor

La configuración 3 usa la salida del modelo 2.1 para *optimal_pred_mood*. El acto de diálogo previo se representa usando los atributos de obligaciones y de common ground del enunciado *n-1*, donde *n* es el enunciado actual. Los atributos *obligations_minus1* y *commgr_minus1* se evalúan como predictores tanto para el target *obligations* como para *commgr*. Los atributos de etiquetación INTSINT se incluyen como predictores

para acto de diálogo en la configuración 3 pero no en la 2 porque el algoritmo J48 encontró que mejoran la tasa de reconocimiento en 3 pero no en 2. Esto fue determinado al comparar con otros modelos donde estos cuatro atributos (*first_1*, *first_2*, *first_3* y *last_2*) no se usan como predictores. Los valores de base para evaluar los resultados de la configuración 3 son las *accuracies* y *Kappas* obtenidas en la configuración 2.

Núm. regla	Regla	Casos donde Ocurre la Premisa	Casos No Satisfactorios	Confianza %	Soporte %
1	IF <i>obligations=respuesta</i> AND <i>utt_duration</i> ≤1304.3125, THEN <i>acepta</i>	135	20	85.2	12.9
2	IF <i>obligations=dir-accion</i> , THEN <i>sin-etiqueta</i>	120	37	69.2	11.5
3	IF <i>obligations=sol-info</i> AND <i>speaker_role=sistema</i> AND <i>utt_duration</i> >516.375, THEN <i>sin-etiqueta</i>	113	22	80.5	10.8
4	IF <i>obligations=compromiso</i> , THEN <i>acepta</i>	112	3	97.3	10.7
5	IF <i>obligations=sin-etiqueta</i> AND <i>utt_duration</i> ≤921.875 AND <i>speaker_role=sistema</i> , THEN <i>acepta</i>	101	30	70.3	9.7

Tabla 5.12. Las 5 reglas con mayor soporte para actos de diálogo de *common ground* (modelo 2.2.B)

El modelo 3.1.A, un árbol para obligaciones, contiene 148 reglas y el modelo 3.1.B, un árbol para *common ground*, contiene 153. Los resultados muestran que las *accuracies* y *Kappas* son mayores que sus líneas base cuando se usa el acto de diálogo previo como uno de los predictores. Las tablas 5.13 y 5.14 presentan las cinco reglas con mayor soporte de cada modelo.

En el plano de las obligaciones, la etiqueta del acto de diálogo complementario, es decir, *commgr*, es el atributo más contributivo para la predicción. Cuatro de las cinco reglas en la tabla 5.13 muestran un atributo de un acto de diálogo previo, ya sea *obligations_minus1* o *commgr_minus1*, como uno de los predictores. Una de las 5 reglas usa *number_of_tones*, el cual proveyó una mayor capacidad clasificatoria que *utt_duration*; esto podría ser causado por un efecto comparable con la normalización de las duraciones de los enunciados, ya que *number_of_tones* es una representación indirecta de la *utt_duration*, pero descartando la influencia de las diferencias en velocidad del habla de los diversos hablantes.

En el modelo del *common ground*, también la etiqueta del acto de diálogo complementario, es decir, *obligations*, es el atributo más contributivo y el acto de diálogo previo está presente en dos de las cinco reglas de la tabla 5.14. En este modelo, tanto *number_of_tones* como *utt_duration* resultaron útiles en la tarea de clasificación.

Núm. de regla	Regla	Casos donde ocurre la premisa	Casos no Satisfactorios	Confianza %	Soporte %
1	IF <i>commgr</i> = <i>sin-etiqueta</i> AND <i>commgr_minus1</i> = <i>acepta</i> AND <i>number_of_tones</i> <= 15, THEN <i>sol-inf</i>	91	52	42.9%	8.7%
2	IF <i>commgr</i> = <i>accion-graf</i> AND <i>obligations_minus1</i> = <i>compromiso</i> , THEN <i>sol-inf_accion-graf</i>	72	1	98.6%	6.9%
3	IF <i>commgr</i> = <i>acepta</i> AND <i>speaker_role</i> = <i>sistema</i> AND <i>obligations_minus1</i> = <i>dir-accion</i> , THEN <i>compromiso</i>	71	19	73.2%	6.8%
4	IF <i>commgr</i> = <i>espera_repet-refr</i> , THEN <i>sol-inf</i>	54	1	98.1%	5.2%
5	IF <i>commgr</i> = <i>acepta</i> AND <i>speaker_role</i> = <i>usuario</i> AND <i>commgr_minus1</i> = <i>accion-graf</i> , THEN <i>resp</i>	51	0	100.0%	4.9%

Tabla 5.13 Las cinco reglas con mayor soporte para actos de diálogo de *obligaciones* (modelo 3.1.A).

Núm. de regla	Regla	Casos donde ocurre la premisa	Casos no Satisfactorios	Confianza %	Soporte %
1	IF <i>obligations</i> = <i>compromiso</i> , THEN <i>acepta</i>	112	3	97.3%	10.7%
2	IF <i>obligations</i> = <i>sol-inf</i> AND <i>speaker_role</i> = <i>sistema</i> , THEN <i>espera_repet-refr</i>	99	47	52.5%	9.5%
3	IF <i>obligations</i> = <i>sol-inf_accion-graf</i> , THEN <i>accion-graf</i>	98	2	98.0%	9.4%
4	IF <i>obligations</i> = <i>resp</i> AND <i>commgr_minus1</i> = <i>accion-graf</i> , THEN <i>acepta</i>	56	5	91.1%	5.4%
5	IF <i>obligations</i> = <i>resp</i> AND <i>commgr_minus1</i> = <i>espera_repet-refr</i> , THEN <i>acepta</i>	48	7	85.4%	4.6%

Tabla 5.14 Las cinco reglas con mayor soporte para actos de diálogo de *common ground* (modelo 3.1.B).

La tabla 5.15 presenta los porcentajes de presencia de los atributos en cada uno de los modelos de la configuración 3. Los atributos con mayores porcentajes están asociados a una mayor contribución en la tarea de clasificación porque tienen una mayor capacidad discriminativa.

La tabla 5.16 muestra las medidas *F*, *recalls* y *precisions* de las clases del modelo para *obligaciones*, ordenadas en forma descendente respecto a *F*; la tabla 5.17 muestra las correspondientes a *common ground*. En ambos casos se omitió presentar las clases cuyas medidas *F* tienen valor 0 (cero).

Los cinco actos de diálogo de *obligaciones* con mayor medida *F* son: *sol-inf_accion-graf*, *sol-inf_accion-graf_resp*, *resp*, *compromiso* y *oferta*. Esto involucra que son los que tienen mayor probabilidad de ser

correctamente reconocidos por el modelo. Los cinco actos de diálogo de common ground que tienen las mayores medidas F son: *afirm_quiza*, *accion-graf*, *oferta_conv-abre*, *acepta* y *reafirm*.

Model 3.1.A			Model 3.1.B		
Atributo	Núm. de Reglas	%	Atributo	Núm. de Reglas	%
<i>commgr</i>	148	100.0	<i>obligations</i>	153	100.0
<i>commgr_minus1</i>	79	53.4	<i>commgr_minus1</i>	140	91.5
<i>obligations_minus1</i>	40	27.0	<i>first_3</i>	42	27.5
<i>speaker_role</i>	33	22.3	<i>speaker_role</i>	32	20.9
<i>first_3</i>	28	18.9	<i>obligations_minus1</i>	18	11.8
<i>number_of_tones</i>	10	6.8	<i>number_of_tones</i>	12	7.8
<i>optimal_pred_mood</i>	4	2.7	<i>first_2</i>	12	7.8
<i>first_1</i>	0	0.0	<i>utt_duration</i>	9	5.9
<i>first_2</i>	0	0.0	<i>last_2</i>	3	2.0
<i>last_2</i>	0	0.0	<i>first_1</i>	0	0.0
			<i>optimal_pred_mood</i>	0	0.0

Tabla 5.15 Porcentajes de presencia de los atributos en las reglas de los árboles del modelo 3.1: A) *Obligaciones* y B) *Common ground*.

Obligaciones	F	Recall	Precision
<i>sol-inf_accion-graf</i>	0.969	0.969	0.969
<i>sol-inf_accion-graf_resp</i>	0.875	0.875	0.875
<i>resp</i>	0.857	0.853	0.861
<i>compromiso</i>	0.812	0.866	0.764
<i>oferta</i>	0.800	0.800	0.800
<i>dir-accion_resp</i>	0.765	0.813	0.722
<i>sin-etiqueta</i>	0.669	0.637	0.704
<i>sol-inf</i>	0.654	0.680	0.630
<i>dir-accion</i>	0.525	0.533	0.516
<i>sol-inf_resp</i>	0.200	0.125	0.500

Tabla 5.16 Medidas F , recalls y precisions de actos de diálogo de *obligaciones* (modelo 3.1.A).

5.9 Discusión de las reglas

Al analizar los resultados de las tres configuraciones, se observa que en la primera, la información entonativa por sí misma, aquí representada por los etiquetajes INTSINT y la duración del enunciado, produce tasas de reconocimiento de 38 y 47% para *obligaciones* y *common ground*, respectivamente, y Kappas menores a 0.19 en ambos casos. El uso de la modalidad del enunciado como atributo predictor mejora la tasa de reconocimiento del acto de diálogo; al usar la etiqueta manualmente asignada de la modalidad se obtiene 47% para *obligaciones* y 56% para *common ground*; es decir, aproximadamente 9 puntos porcentuales de mejora. En una aplicación de mundo real, la modalidad del enunciado tendría que ser reconocida a partir de la señal del habla, por lo cual se implementa el atributo denominado *modalidad predicha* (*predicted_mood*),

una etiqueta obtenida automáticamente por un árbol de decisión a partir de los etiquetajes INTSINT (*first_2* y *last_2*) y la duración del enunciado. Al usar esta representación de la modalidad, se obtiene 39 y 50% de exactitud en el reconocimiento para los respectivos planos; ésta es menor que al usar la modalidad del enunciado etiquetada manualmente, pero mayor que si no se usa modalidad alguna. Las reglas obtenidas para la modalidad predicha son consistentes con las teorías de la entonación del español, que establecen que la región final del contorno entonativo del enunciado determina significativamente el tipo de modalidad, existiendo también una contribución marginal de la región inicial.

Common Ground	F	Recall	Precision
<i>afirm_quiza</i>	1.000	1.000	1.000
<i>accion-graf</i>	0.986	0.990	0.981
<i>oferta_conv-abre</i>	0.833	1.000	0.714
<i>acepta</i>	0.790	0.828	0.755
<i>reafirm</i>	0.692	0.563	0.900
<i>opcion-abierta_despliega</i>	0.689	0.756	0.633
<i>acepta-parte</i>	0.683	0.700	0.667
<i>sin-etiqueta</i>	0.671	0.710	0.637
<i>espera_repet-refr</i>	0.671	0.926	0.526
<i>opcion-abierta</i>	0.600	0.500	0.750
<i>afirm</i>	0.446	0.425	0.470
<i>conv-cierra</i>	0.250	0.200	0.333
<i>espera</i>	0.235	0.154	0.500
<i>SNE</i>	0.154	0.083	1.000
<i>afirm_acepta</i>	0.154	0.111	0.250

Tabla 5.17 Medidas *F*, *recalls* y *precisions* de actos de diálogo de *common ground* (modelo 3.1.B).

La contribución de la modalidad del enunciado como predictor es tomada en cuenta y la necesidad de una etiqueta óptima producida automáticamente conduce a la introducción de la modalidad predicha optimizada (*optimal_pred_mood*), cuyos predictores son: el rol del hablante (*speaker_role*), etiquetajes INTSINT y la duración del enunciado. La inclusión del rol del hablante proporciona una mayor exactitud en el reconocimiento de la modalidad (77% contra 74). Como consecuencia de ello, mejora el reconocimiento del tipo de acto de diálogo en la configuración 2, cuyos resultados se discuten a continuación.

La configuración 2 ofrece mayor exactitud que la 1, logrando 66.2 y 66.9%, para obligaciones y *common ground*, respectivamente. Se usan además de atributos entonativos, otros *no* entonativos, tales como el rol del hablante y el tipo de acto de diálogo del plano complementario (*common ground* para reconocer *obligaciones* y *obligaciones* para *common ground*). El hecho de que la tasa de reconocimiento mejore al usar el acto del plano complementario sugiere que existe una interacción entre ambos, y que los interlocutores pueden apoyarse en el conocimiento de uno para reconocer el otro.

Las mejores tasas de reconocimiento del acto de diálogo, 73 y 69% para obligaciones y *common ground*, respectivamente, se obtienen en la tercera configuración. Los predictores útiles en las obligaciones, en orden de mayor a menor contribución, son: *commgr*, *commgr_minus1*, *obligations_minus1*, *speaker_role*, *first_3*, *number_of_tones* y *optimal_pred_mood*. Para los actos del *common ground* los predictores útiles, también ordenados descendientemente por su contribución, son: *obligations*, *commgr_minus1*, *first_3*, *speaker_role*, *obligations_minus1*, *number_of_tones*, *first_2*, *utt_duration*, *last_2*, *first_1* y *optimal_pred_mood*. En ambos casos, el tipo de acto del plano complementario es el que presenta mayor capacidad discriminativa; es decir, éste aparece en una mayor cantidad de reglas que los otros atributos. De las dos componentes del acto de diálogo previo, el atributo del *common ground* (*commgr_minus1*) es el que más contribuye al reconocimiento, tanto para obligaciones como para *common ground*; siendo mayor para el segundo que para el primero.

Para el reconocimiento de obligaciones el rol del hablante contribuye más que la entonación (etiquetaje INTSINT); sin embargo, para el de *common ground* ocurre lo contrario. La duración del enunciado parece contribuir en la misma medida sobre ambos planos, ya que en ambos casos su grado de contribución ocupa el sexto lugar. La modalidad predicha optimizada (*optimal_pred_mood*) es el atributo menos contribuyente para las obligaciones y no contribuye en el reconocimiento del *common ground*.

En el capítulo destinado a las conclusiones se discuten los resultados del presente trabajo a la luz de algunas teorías de la entonación y del acto de diálogo; asimismo, estos resultados se contrastan con los de trabajos previos en el área y se exploran algunas aplicaciones prácticas.

Capítulo 6. Conclusiones

En esta investigación se analizó y modeló la relación entre el acto de diálogo y la entonación desde el punto de vista de la lingüística computacional. En el estudio se consideran también fuentes de información adicionales a la entonación, que el interlocutor aprovecha en la determinación del tipo de acto de diálogo expresado mediante el enunciado, tales como el rol del hablante y la modalidad del enunciado, entre otros.

6.1 Postulación final

La postulación es que la información entonativa contribuye al reconocimiento del tipo de acto de diálogo; pero su contribución depende de las interacciones que se dan con diversos elementos del contexto del diálogo, tales como el rol del hablante y el acto de diálogo correspondiente al enunciado previo. Por ello, la información entonativa en forma aislada no garantiza un reconocimiento exitoso del tipo de acto de diálogo, pero éste mejora al utilizarla. Las interacciones de la entonación involucran la presencia de los actos de habla indirectos, altamente frecuentes en el habla espontánea y que requieren un proceso de interpretación basado en múltiples fuentes informativas. Adicionalmente, se ha confirmado empíricamente la relación entre la forma del contorno entonativo y el tipo de modalidad del enunciado; en particular, la fuerte influencia que ejerce la región final, asociada al *tonema*, y la contribución marginal de la región inicial.

6.2 Acto de habla

El acto de habla se aborda bajo una noción de acto de diálogo que ha sido sintetizada a partir de las aproximaciones de diversos autores, manteniendo un gran apego a la perspectiva computacional. Más que considerar al acto en forma aislada se le concibe con base en su función comunicativa dentro de diálogos prácticos; así que es entendido como la intención del hablante expresada dentro del contexto de una transacción y teniendo componentes sobre tres planos comunicativos: el de las *obligaciones*, el del *acuerdo* y el del *entendimiento*. Los actos de diálogo contenidos dentro de una transacción se encuentran en un estado de equilibrio; ya que los tipos de actos se organizan en pares, donde un cierto tipo genera un *cargo* y otro tipo el *abono* que lo equilibra. Con base en DIME-DAMSL se han identificado y etiquetado los pares de tipos que producen estados equilibrados.

El *grounding* o aterrizamiento se establece mediante un subconjunto de tipos actos de diálogo que tienen la función principal de establecer, mantener o reconstruir el conocimiento común entre los interlocutores; este concepto se deriva de la noción proporcionada por el esquema DIME-DAMSL. El concepto difiere

sensiblemente de la noción de Clark y Schaefer, una de las más conocidas, según la cual el *grounding* es el proceso colaborativo desarrollado por los interlocutores que tiende a incrementar el conocimiento, las creencias o las presuposiciones compartidos por ambos.

6.3 Niveles de representación lingüística

El acto de diálogo puede expresarse mediante diversos niveles de representación lingüística: fonético, fonológico, léxico, morfológico, sintáctico, semántico y pragmático. El recurso empírico utilizado, el corpus DIME, contiene la transcripción ortográfica de diálogos, segmentada por turno de hablante y por enunciado; contiene también etiquetajes de alófonos, fonemas, sílabas fonéticas, palabras, partes del habla (*parts of speech*), índices de ruptura (*break indices*) según el esquema Sp-ToBI, reparaciones del habla (*speech repairs*), marcadores de discurso, transcripción de entonación con el esquema INTSINT y modalidad del enunciado. Adicionalmente, para cada enunciado se etiquetó también el tipo de rol del hablante (*usuario* o *sistema*). El acto de diálogo se representa mediante el esquema de etiquetaje DIME-DAMSL.

A partir de estos niveles, se han determinado los tipos de información lingüística adecuados para los objetivos particulares, enumerados a continuación: el contorno de la frecuencia fundamental, la duración del enunciado (representada tanto en unidades de tiempo como en número de tonos), la modalidad del enunciado y los etiquetajes de acto de diálogo; finalmente, se ha incluido también el rol del hablante que realiza el enunciado. Aunque este último atributo no proviene estrictamente de los niveles de representación lingüística, se ha incorporado al conjunto porque enriquece al nivel pragmático.

Aunque los etiquetajes del recurso empírico utilizado ofrecen, entre otros, datos provenientes del nivel léxico, en este trabajo la atención se ha concentrado en fuentes de información *no* léxica, principalmente la entonación, siendo uno de los objetivos principales la evaluación de su contribución al reconocimiento del tipo de acto de diálogo. Es de esperar que cada tipo de información contribuya en diversa magnitud a la tarea del reconocimiento y también puede esperarse que existan interacciones entre los diversos tipos; por ello se requiere de análisis estadísticos que permitan entender estos comportamientos. Una vez identificados los niveles de representación o los tipos de información con mayor contribución para la tarea, esto puede aprovecharse en la implementación de aplicaciones de administración de diálogo.

6.4 Comparación con el estado del arte

Al contrastar los resultados con los del estado del arte, se observa que los de proyectos como VERBMOBIL (Wahlster, 1993) no son directamente comparables con los aquí obtenidos, ya que en VERBMOBIL la tipología de actos de diálogo incluye algunos que fueron definidos para los dominios específicos del

agendamiento de reuniones de trabajo y la planeación de viajes (VERBMOBIL-2). Sin embargo, al analizar su esquema general de etiquetaje, descrito por Alexandersson *et al.* (1998), se encuentran algunos aspectos comparables, tales como la identificación de etapas en el desarrollo del diálogo, de modo similar a las fases de la transacción en DIME-DAMSL. VERBMOBIL aborda los denominados *diálogos de negociación*, considerando cinco fases: 1) *Hello*, donde los interlocutores se saludan y se presentan mutuamente; 2) *Opening*, donde se introduce el tópico que será tratado en el diálogo; 3) *Negotiation*, constituida por el proceso de negociación en sí; 4) *Closing*, donde la negociación concluye, los interlocutores expresan su acuerdo y eventualmente se recapitula el tópico y 5) *Good_bye*, la despedida entre los participantes. Por su parte, DIME-DAMSL, al analizar diálogos prácticos, plantea las fases de *especificación* y *satisfacción* de la intención. VERBMOBIL establece que un diálogo puede evolucionar a lo largo de las fases sin que exista un orden estricto; es decir, un hablante puede avanzar o (retroceder) entre cualquier par de fases antes de llegar a la última, e incluso podría retroceder para después regresar a ella. En DIME-DAMSL pueden existir transacciones anidadas, de modo que tanto la especificación como la satisfacción de la intención pueden ocurrir en modo recursivo, haciendo que el proceso se complete gradualmente.

En VERBMOBIL, los actos de diálogo se clasifican en tres grupos: *control de diálogo*, *administración de la tarea* y *promoción de la tarea*. Esta clasificación es similar a las categorías del *nivel informativo* de DIME-DAMSL (heredado de DAMSL), mediante el cual se determina si el enunciado trata acerca de la *administración de la tarea*, la *administración de la comunicación* o la *realización de la tarea*. En VERBMOBIL es posible asignar cualquier cantidad de etiquetas de acto de diálogo a un enunciado, mientras que en DIME-DAMSL se prefiere limitar la cantidad de etiquetas posibles apoyándose en las nociones de *planos de comunicación*, y *cargos* y *abonos*.

Comparando con los resultados de Fernández y Picard (2002), en este último se analizan y modelan actos de diálogo de conversaciones telefónicas sobre temas de interés general, lo cual tampoco permite una comparación consistente con la presente investigación; sin embargo, es destacable que los valores de exactitud del reconocimiento al usar información entonativa exclusivamente, son menores que 48%, similares a los obtenidos en algunas de las configuraciones experimentales del presente trabajo. Fernández y Picard usaron el esquema de etiquetaje de actos de diálogo de Levin *et al.* (1998); éste es comparable con DIME-DAMSL en el sentido de que ambos toman a DAMSL como punto de partida; también, porque en ambos se consideran estructuras informativas de mayor alcance que el enunciado. Levin *et al.*, además de usar la noción de acto de habla, usan las denominadas *juego de diálogo* y *actividad*. El *juego de diálogo* es un conjunto de enunciados de los interlocutores en el que se dan interacciones de *iniciativa* e *intención*; los límites de un *juego* están determinados por el cambio de la persona que tiene la iniciativa en la conversación y por el cambio en la intención del hablante (*e.g.* al cambiar de una intención de afirmar a una de preguntar). Los turnos dentro de un juego se etiquetan con: *Initiative*, *Response* y *Feedback*. Los juegos pueden traslaparse, ya sea como anidamientos o como juegos intercalados. Existen ocho tipos principales de juegos: *seeking information*, *giving information*, *giving directive*, *action commit*, *giving opinion*, *expressive*, *seeking*

confirmation y *communication filler*. La *actividad* tiene un nivel jerárquico mayor que el *juego de diálogo* y es un segmento de discurso que se concentra en un propósito o meta específicos de los hablantes. Dado que los diálogos etiquetados no son orientados a tareas, los segmentos no pueden dividirse en subtareas y se han dividido con base en un criterio de tópico, considerando diez etiquetas de actividad: *inform*, *inquire*, *plan*, *convince*, *negotiate*, *gossip*, *argue*, *conversation management*, *greet-close* e *instruct*. El juego de diálogo es similar a la estructura de *cargos* y *abonos* de DIME-DAMSL y la *actividad* es comparable a la *transacción* de este último.

Quizá el trabajo previo del área con el cual podría haber mayores posibilidades de comparación es el de Shriberg *et al.* (1998); aunque sus análisis y modelos se hicieron para el inglés americano, existen similitudes entre los esquemas de etiquetaje de actos de diálogo y entre los algoritmos de Aprendizaje Automático utilizados. Shriberg *et al.* usan información entonativa exclusivamente, representándola con un enfoque de procesamiento de señales; sus resultados sugieren que los actos de diálogo están marcados redundantemente en los enunciados de la conversación natural y que una variedad de atributos prosódicos extraíbles automáticamente podrían apoyar al procesamiento del diálogo en aplicaciones de habla.

6.5 Esquemas para etiquetaje de actos de diálogo

Se han explorado diversos esquemas de etiquetaje de actos de diálogo con el propósito de elegir el más adecuado para los objetivos particulares. Se utilizó el esquema DAMSL (Allen y Core, 1997) como punto de partida para implementar esquemas de etiquetaje de actos de diálogo para propósitos especializados. La exploración del estado del arte así lo ha mostrado, ya que diversos trabajos previos en el área se han apoyado en él; p. ej. en las etiquetaciones de actos de diálogo usadas por Shriberg *et al.* (1998) y por Fernández y Picard (2002), entre otros. Una ventaja aplicativa de DAMSL es su gran flexibilidad, pues permite asignar múltiples etiquetas; sin embargo, una debilidad es la ausencia de estructuras de mayor alcance que el enunciado.

DAMSL es altamente influyente por ser uno de los primeros y más versátiles. Sin embargo, su alta flexibilidad también tiende a producir bajos niveles de consistencia entre los etiquetajes realizados por distintos anotadores. Por otra parte, para que el esquema de etiquetaje de VERBMOBIL pudiera aprovecharse en la presente investigación su conjunto de etiquetas tendría que modificarse, ya que originalmente se limita a los actos que se expresan oralmente en diálogos realizados por teléfono y no contiene las etiquetas necesarias para actos que ocurren en un entorno multimodal, tales como las *directivas de acción* o las acciones de naturaleza gráfica. Finalmente, en el esquema de Levin *et al.* (1998), utilizado por Fernández y Picard (2002), también habría que extender su tipología de actos, ya que originalmente sólo es aplicable a la conversación informal sobre temas generales y no a los diálogos prácticos.

6.5.1 Esquema DIME-DAMSL

El esquema DIME-DAMSL, apoyado fuertemente en DAMSL, analiza y etiqueta al acto de diálogo considerando un alcance mayor que el enunciado; es decir, incorpora estructuras tales como la *transacción*, *tres planos* de expresión y pares equilibrados de *cargos* y *abonos*. Como consecuencia de ello, produce etiquetajes que tienden a presentar un alto grado de consistencia.

DIME-DAMSL es la principal fuente para sintetizar un concepto de acto de diálogo en esta investigación. Se ha observado que su marco teórico y sus lineamientos prácticos favorecen niveles de consistencia satisfactorios en los etiquetajes de acto de diálogo, lo cual, a su vez, favorece la exactitud y la confiabilidad de los modelos producidos mediante algoritmos de aprendizaje automático.

Aunque en los experimentos no se analizan atributos descriptivos de la *transacción* o de la fase de la misma en la cual ocurre una instancia de un acto de diálogo, sería de esperar que el uso de esa información incremente la exactitud del reconocimiento del tipo de acto de diálogo.

La estructura de planos de comunicación de DIME-DAMSL ha permitido una modelación del acto de diálogo en la cual se percibe que existen fenómenos de interacción entre los dos planos básicos, de modo que el tipo de acto que ocurre sobre uno de ellos aporta información para reconocer el tipo de acto que ocurre en su plano complementario; es decir, el reconocimiento del tipo que ocurre en las *obligaciones* permite reconocer al que ocurre en el del *common ground* y viceversa. En una primera aproximación, el fenómeno puede analizarse en los siguientes términos: en algunas instancias de actos de diálogo el propósito preponderante es establecer, reforzar o reconstruir las suposiciones y conocimientos compartidos entre los interlocutores, mientras que en otras instancias es generar la responsabilidad, en el hablante o en el oyente, de realizar una acción, ya sea verbal o de otra índole. El propósito puede ser también una mezcla, en diversas proporciones, de los dos descritos; en este caso existen ciertos pares de tipos de actos que tienden a ser preferidos por los hablantes ante ciertos contextos. Algunos ejemplos son: *compromiso* en *obligaciones* con *aceptación* en *common ground*, *directiva de acción* en *obligaciones* con ninguna componente en *common ground* o simultáneamente *respuesta* y *directiva de acción* en *obligaciones* con *aceptación parcial* en *common ground*. Sin embargo, la existencia de una correlación estadística no es suficiente para confirmar la existencia de una relación causal, por lo cual este fenómeno tiene que investigarse en mayor profundidad en trabajos futuros.

6.5.2 DIME-DAMSL: Aplicación

En la experiencia del presente trabajo, el esquema DIME-DAMSL (Pineda *et al.*, 2006 y 2006a) ha resultado confiable para la etiquetación de actos de diálogo y puede recomendarse su uso para etiquetar diálogos prácticos en dominios diversos. Con base en DIME-DAMSL, se ha evaluado un conjunto de atributos para reconocer tipos de actos de diálogo mediante métodos de aprendizaje automático usando información entonativa como uno de los predictores.

6.6 Uso del estadístico *Kappa*

El estadístico *Kappa* ha resultado de gran utilidad para evaluar la consistencia en los etiquetajes de actos de diálogo desarrollados con el esquema DIME-DAMSL. También, para la evaluación de los modelos basados en aprendizaje automático de la modalidad del enunciado y del acto de diálogo.

6.7 Representación de la entonación

Al considerar los requerimientos inicialmente establecidos para representar la entonación, el enfoque del esquema INTSINT resulta útil; la adaptación que se hace descartando la información de duraciones de tonos, permite un mayor grado de generalización y evita la necesidad de normalización. Además, al representar el contorno entonativo como una secuencia alfabética se podrían desarrollar diversos tipos de análisis que no serían factibles al usar representaciones numéricas.

La representación de la entonación con un enfoque de procesamiento de señales habría sido factible, aunque requeriría una mayor complejidad en la preparación de los datos, tanto por la necesidad de un proceso de normalización de las *f0* y de las duraciones, como por la necesidad de algún esquema para abstraer la forma del contorno entonativo en una representación de tipo discreto.

El enfoque del esquema ToBI o aún el de Sp-ToBI, habrían sido inadecuados para nuestras necesidades porque, al ser ToBI en un esquema fonológico, no considera la representación de información fonética, misma que constituye una de las componentes fundamentales dentro de los análisis y modelos desarrollados.

6.8 Experimentos y resultados

Los experimentos para modelar y analizar el acto de diálogo han estado basados en algoritmos de aprendizaje automático en particular, para creación de árboles de decisión; antes de diseñar los experimentos, se ha

realizado una serie de análisis estadísticos para conocer la distribución e interacciones de los datos y para definir diversas configuraciones experimentales. Un proceso necesario antes de los análisis estadísticos y de los experimentos ha sido la determinación y creación de atributos informativos (*features*) a partir de los etiquetajes disponibles; por ello, se han creado atributos para representar al acto de diálogo, la modalidad del enunciado, la entonación, etc.

Aunque el esquema DIME-DAMSL dispone de un subconjunto de etiquetas que ofrece granularidad fina para etiquetar actos de diálogo, denominado *DIME-DAMSL detallado*, para los objetivos del presente trabajo ha resultado suficiente el subconjunto de *DIME-DAMSL preliminar*, que ofrece granularidad gruesa. El primero establece subtipos de ciertos tipos de actos; p. ej. considera tres variedades de la *solicitud de información*: *pregunta sí/no*, *pregunta pronominal* y *pregunta imperativa*. También, distingue entre tipos específicos de acciones gráficas, tales como *agregar*, *mover* o *borrar un objeto*, desplegar un catálogo, rotar el punto de vista del espacio virtual, etc. Dado que en esta investigación la atención se concentra en actos de diálogo expresados mediante enunciados, el etiquetaje de las acciones de naturaleza gráfica no requiere demasiada granularidad y se limita a identificarlas como acción gráfica en general. Los atributos básicos producidos con los etiquetajes de DIME-DAMSL han sido dos: la etiqueta del plano de *obligaciones* y la del *common ground*, estando esta última formada por una concatenación de las correspondientes al *acuerdo* y al *entendimiento*. Aunque el *common ground* está estructurado en dos subplanos, se ha optado por concatenar sus etiquetajes en un sólo atributo, ya que los análisis estadísticos han mostrado que los actos del *entendimiento* presentan un escaso número de instancias, lo cual habría imposibilitado la creación de modelos representativos del fenómeno. Adicionalmente, se han establecido otros dos atributos para *obligaciones* y *common ground*, que representan a las dos componentes del acto de diálogo del enunciado previo.

El atributo representativo de la modalidad del enunciado y su respectivo conjunto de etiquetas (*tagset*) se han formalizado considerando dos *tagsets*. En ambos se consideran cuatro tipos de modalidad: *declarativa*, *interrogativa*, *imperativa* y *otra*, pero en uno de ellos se han considerado tres subtipos de *interrogativa* (*pregunta pronominal*, *pregunta sí/no* y *pregunta genérica*, aplicando esta última cuando no son aplicables las primeras dos); mientras que en el otro se considera un sólo tipo interrogativo. El propósito ha sido evaluar cuál de los dos permite un mejor reconocimiento automático de la modalidad.

Los atributos de la entonación son diez subcadenas obtenidas de la secuencia de etiquetaje INTSINT; esta última está constituida por las etiquetas INTSINT sin incluir información de la duración de los tonos. De las diez subcadenas, cinco representan a la región final y cinco a la inicial del contorno entonativo. Concretamente, contienen a los últimos 5, 4, 3, 2 y 1 y a los primeros 1, 2, 3, 4 y 5 tonos de la secuencia INTSINT.

Para representar la *duración del enunciado* los atributos implementados ha sido dos: la duración en milisegundos y el número de tonos INTSINT del etiquetaje entonativo. El primero expresa un valor absoluto

de la duración y el segundo un valor normalizado que elimina el efecto de dispersión introducido por las diferencias en las velocidades del habla de los sujetos.

El *rol del hablante* se ha representado mediante un solo atributo cuyo valor puede ser *sistema* o *usuario*.

6.8.1 Análisis estadístico

El análisis estadístico de los diversos etiquetajes ha permitido saber, por ejemplo, que los dos roles de hablante, *usuario* y *sistema*, tienen una participación cuantitativamente similar en los diálogos. La duración promedio del enunciado es 2.1 seg. Respecto a los tipos de actos de diálogo más frecuentes, la mayoría de los casos está constituida por: aceptaciones, solicitudes de información, respuestas, directivas de acción, afirmaciones, acciones gráficas y compromisos, en ese orden de ocurrencia; los actos del subplano del entendimiento tienen una escasa presencia. La mayoría de los enunciados presentan modalidad *declarativa* (67.2%) o *interrogativa* (25.2%). Los etiquetajes de entonación tienen en promedio 7.1 tonos (etiquetas INTSINT) por enunciado, con mínimo de 2 y máximo de 45. Dado que la región final del contorno entonativo presenta especial interés por estar asociada con el *tonema* y con la modalidad, se analizaron las combinaciones de las dos últimas etiquetas de las secuencias, hallándose que las más comunes son: *BS, DB, UT, TS, TB, BT, MB, US, BH* y *TL*.

De especial utilidad han sido los análisis que exploran las correlaciones entre los siguientes pares de atributos: acto de diálogo con modalidad del enunciado, acto de diálogo con rol del hablante, modalidad con rol de hablante y tipo de acto en el plano de *obligaciones* con tipo de acto en el *common ground*.

Aunque la modalidad declarativa es la más frecuente, se ha observado que la interrogativa ocurre en el 75% de las solicitudes de información, en el 84% de las esperas y en el 100% de las ofertas; esto puede ser indicio de la existencia de ciertas modalidades preferidas por los hablantes para expresar ciertos tipos de actos.

Algunos actos de diálogo son expresados más frecuentemente por uno de los dos tipos de rol de hablante e incluso, algunos actos simplemente no se observan en alguno de ellos. En el plano de las *obligaciones*, por ejemplo, la solicitud de información es más frecuente en el rol del *sistema* (72.2%), la respuesta en el *usuario* (73.3%), la directiva de acción parece ser exclusiva del *usuario* (100%) y el compromiso, exclusivo del *sistema* (100%). En el plano del *acuerdo* el acto de aceptación presenta una distribución estadística similar entre ambos roles (53.1% por el *sistema* y 46.9% por el *usuario*); la afirmación es más frecuente en el *usuario* (71.8%), mientras que la espera, la opción abierta y la oferta son típicas del *sistema* (85.5, 100 y 100%, respectivamente).

El análisis de las modalidades usadas por cada tipo de rol de hablante ha mostrado que las declarativas se usan en proporciones similares por ambos roles: 56.1% por el *usuario* y 43.9% por el *sistema*. En las otras

modalidades se observan claros sesgos: las interrogativas son más comunes en el *sistema* (74.2%); las imperativas, en el *usuario* (92.9%), igual proporción que se observa para el mismo tipo de rol respecto a la modalidad denominada *otra*.

Existe una correlación entre la componente del acto de diálogo que ocurre sobre un plano y la correspondiente al plano complementario; es decir, ciertos pares de etiquetas de *obligaciones/common ground* tienen una mayor frecuencia que otros. En el plano de las *obligaciones*, por ejemplo, la mayoría (69.2%) de las directivas de acción están asociadas con una etiqueta nula (sin etiqueta) en el plano del *acuerdo*; lo mismo ocurre en el 65.6% de las solicitudes de información. El 97.3% de los compromisos y el 100% de las ofertas están vinculados a una aceptación. Algunas de las etiquetas compuestas también muestran correlaciones significativas; por ejemplo, 81.3% de los casos de directiva de acción y respuesta coinciden con aceptación parcial, 62.5% de los de solicitud de información y respuesta se asocian con etiqueta nula en el *acuerdo*. Desde el punto de vista del plano del *acuerdo*, 91.3% de las esperas se vinculan con solicitudes de información en las *obligaciones*; 68.1% de las opciones abiertas, con respuestas; 65% de las aceptaciones parciales con la combinación de directiva de acción con respuesta y 64.7% de las reafirmaciones, con respuestas.

6.8.2 Configuración de los experimentos

Se eligió el algoritmo J48 (Witten y Frank, 2005), basado a su vez en CART (Breiman *et al.*, 1983), para generar árboles de decisión. El planteamiento general de los experimentos ha consistido en crear una serie de modelos para el reconocimiento (o la predicción) de tres *targets*: los dos principales son el acto de diálogo en el plano de las *obligaciones* y el acto en el plano del *common ground*; otro *target* es la modalidad del enunciado, que se requiere para producir uno de los datos de entrada para los primeros dos modelos. Para cada *target* se evalúa la utilidad que ofrecen diferentes subconjuntos de los atributos disponibles en el *dataset* al ser implementados como predictores, obteniéndose en cada caso conjuntos de modelos con diferentes grados de exactitud para un mismo *target*. Para los modelos de actos de diálogo se han evaluado tres configuraciones: 1) sólo atributos de naturaleza entonativa, 2) atributos entonativos con otros no entonativos y 3) acto de diálogo previo.

El algoritmo J48 se encarga de descubrir cuáles son los atributos que tienen capacidad clasificatoria (o predictiva), distinguiéndolos de aquéllos que no. Los primeros aparecen en los árboles de decisión y los últimos, no. Por ello, en algunos árboles producidos dentro de las tres configuraciones experimentales no están presentes uno o más de los atributos usados como predictores.

Configuración 1: Sólo atributos entonativos

Los atributos entonativos evaluados para el reconocimiento o predicción del tipo de acto de diálogo han sido algunos de los etiquetajes INTSINT, en particular los primeros dos y tres y los últimos dos tonos de la secuencia; también, la duración del enunciado, tanto en *mseg* como en número de tonos y la modalidad del enunciado. De la modalidad se ha evaluado la contribución de la etiquetada manualmente y la de otra, denominada *modalidad predicha*, que se genera automáticamente a partir de los últimos dos y primeros dos tonos y de la duración del enunciado en *mseg*.

Configuración 2: Atributos entonativos con otros no entonativos

En la Configuración 2, los atributos evaluados han sido los primeros dos y últimos dos tonos, siendo el rol del hablante el dato no entonativo incorporado; se ha incluido una forma de modalidad predicha, denominada *optimizada*, que explota la contribución del rol del hablante y ofrece una mayor exactitud que la modalidad predicha básica usada en la Configuración 1. También se usa la duración del enunciado. Aprovechando las posibilidades del esquema DIME-DAMSL, se explora la contribución de cada uno de los dos planos comunicativos (*obligaciones* y *common ground*) para hacer el reconocimiento del tipo de acto de su respectivo plano complementario; es decir, se usa el tipo de acto de las *obligaciones* como uno de los predictores para reconocer el tipo de acto del *common ground* y viceversa.

Configuración 3: Utilización del acto de diálogo previo

La Configuración 3 evalúa, además de los atributos explorados en la Configuración 2, las etiquetas de acto de diálogo de ambos planos, correspondientes al enunciado previo, partiendo de la hipótesis de que el tipo de acto de diálogo de un enunciado puede estar influenciado o determinado por del enunciado inmediato anterior. Se analizan también los primeros tres tonos y la duración del enunciado, tanto en *mseg* como en número de tonos.

6.8.3 Resultados de los experimentos

Como resultado de los experimentos se han obtenido tres series de modelos: una para actos de diálogo de *obligaciones*, otra para actos de *common ground* y otra para modalidad del enunciado. El último se usa para producir uno de los atributos predictores que se alimenta a los otros dos. En cada serie se ha evaluado y comparado la exactitud del reconocimiento (o predicción) de los diversos modelos, identificando los óptimos con sus respectivos conjuntos de atributos predictores.

De las tres configuraciones experimentales, la tercera es la que ofrece mayor exactitud, la primera es la menos exacta y la segunda tiene una capacidad intermedia. Puede interpretarse que la exactitud mejora al incorporar fuentes de información que enriquecen a la estrictamente entonativa. Las *exactitudes* y *Kappas* óptimas obtenidas son, 73.442% y 0.6783 para obligaciones, 69.0316% y 0.6054 para *common ground* y 77.373% y 0.4531 para modalidad del enunciado. Las *exactitudes* y *Kappas* más bajas son: para *obligaciones*, 37.4031% y 0.1571; para *common ground*, 46.3178 y 0.1817 y para modalidad, 74.3049 y 0.4164. Los valores de las óptimas pueden considerarse prometedores porque el proceso de reconocimiento no usa ninguna fuente de información léxica; sin embargo, no son suficientemente altos para ser satisfactorios y en la implementación de una aplicación confiable se tendrían que explotar también las fuentes léxica y sintáctica.

Los resultados son consistentes con los fundamentos teóricos de la entonación del español, según los cuales la región final del contorno entonativo determina significativamente la modalidad del enunciado y la región inicial proporciona una contribución marginal. Los atributos predictores más útiles para la modalidad del enunciado son: las 2 últimas y las primeras 2 etiquetas INTSINT, el rol del hablante y la duración del enunciado en *mseg*. Los resultados del reconocimiento son mejores si en la etiquetación de la modalidad del enunciado se usa un sólo valor de *interrogativa* para la clase del mismo nombre en lugar de tres valores (*pregunta pronominal, pregunta sí/no y otra*).

Se observa que ciertos tipos de acto de diálogo de un plano están presentes en una alta proporción de casos cuando ocurren otros tipos específicos en el plano complementario. Además de despertar interés desde el punto de vista teórico, este fenómeno resulta interesante para aplicaciones prácticas, pues al conocer el tipo de acto que está ocurriendo sobre un plano se tendría un indicio para conocer el que está ocurriendo sobre el otro. En el caso general, no se conocería ninguno de los dos, así que tendría que usarse un modelo alternativo que explotara fuentes de información distintas a las aquí descritas; la fuente que podría resultar más útil es el contenido léxico, al cual se le aplicaría un proceso de parseo. Dado que no está plenamente identificada la causa de la correlación entre los tipos de actos de los dos planos, no está claro cómo podría aprovecharse este hecho de manera práctica; sin embargo, podría resultar útil para complementar las capacidades de un sistema reconocedor de actos de diálogo basado en parseo y en modelos de lenguaje de los tipos de actos.

6.8.4 Discusión de las reglas

De los árboles de decisión óptimos de cada configuración se han extraído los conjuntos correspondientes de reglas *if-then*. La cantidad de reglas en los modelos varía entre aproximadamente 40 y poco más de 150, siendo mayor en las configuraciones que evalúan más atributos, ya que en éstas últimas existe un mayor número de interacciones entre ellos. Para cada regla se han computado los valores de *soporte* y *confianza* que han permitido identificar a las más representativas del fenómeno estudiado. También, se han calculado los

porcentajes de las reglas en las cuales se tiene presencia de cada uno de los atributos; esto permite ponderar el grado de contribución de cada atributo al proceso de reconocimiento o predicción.

Sobre cada uno de los dos planos del acto de diálogo, el atributo de su plano complementario ha sido el más contribuyente para el reconocimiento; es decir, el tipo de acto de *obligaciones* es el atributo con mayor capacidad discriminadora para reconocer el tipo de acto del *common ground*, y este último lo es para reconocer al primero. Este comportamiento está presente tanto cuando se usa el tipo de acto previo como cuando no se usa. Esto es consecuencia de la naturaleza de la teoría de DIME-DAMSL y muestra que ciertos pares de actos de diálogo *obligaciones/common ground* son más frecuentes que otros y que ciertos actos de diálogo ocurren sin un acto de diálogo en su plano complementario.

En los modelos para acto de diálogo que han resultado óptimos, es decir, en los que usan atributos del acto previo, la componente del *common ground* del tipo de acto previo ha sido el segundo atributo con mayor contribución para la predicción, tanto para *obligaciones* como para *common ground*; es destacable que sea el mismo para ambos planos. La contribución de ese atributo ha sido mayor para el *common ground* que para las *obligaciones*. Las causas de estos hechos no son claras y merecen un análisis en mayor amplitud que podría desarrollarse en trabajo futuro.

Para el reconocimiento de *obligaciones* el rol del hablante contribuye más que la entonación (etiquetaje INTSINT); sin embargo, para el de *common ground* la entonación contribuye más que el rol del hablante.

La duración del enunciado parece contribuir en la misma medida sobre ambos planos, ya que en ambos casos su grado de contribución ocupa el sexto lugar.

Para el reconocimiento de la modalidad del enunciado, se han desarrollado dos modelos: uno básico que usa sólo atributos entonativos y duración del enunciado, y otro que, además de éstos, explota la información del rol del hablante. El segundo, denominado *modalidad predicha optimizada*, tiene mayor exactitud que el primero (77.373 contra 74.3049%). A pesar de que el reconocimiento de la modalidad predicha optimizada tiene una exactitud relativamente alta, al alimentarla como uno de los predictores en los modelos para acto de diálogo es el atributo menos contribuyente para las *obligaciones* y parece no contribuir en el reconocimiento del *common ground*. Aunque su baja contribución pareciera contradecir a lo observado en los análisis de correlación entre modalidad y acto de diálogo, donde se percibe que algunos tipos de modalidad están frecuentemente asociados a tipos particulares de actos, no existe tal contradicción. Este fenómeno se puede explicar intuitivamente en los siguientes términos: el hecho de que se conozca la modalidad del enunciado no garantiza que se pueda conocer el tipo del acto de diálogo. Para demostrarlo empíricamente se creó un modelo que usa la modalidad predicha optimizada como único predictor para el tipo de acto, obteniéndose una exactitud de 31.9271 % para *obligaciones* y 42.3778 % para *common ground*, con *Kappas* de 0.1141 y 0.1541, respectivamente, lo cual resulta insatisfactorio en cualquiera de los dos casos para propósitos de

reconocimiento automático; por otra parte, estos valores son interesantes si interpretamos que la modalidad por sí sola permite reconocer 3 ó 4 de cada 10 casos. El hecho de que la exactitud y el valor de *Kappa* para *common ground* sean mayores que para *obligaciones* podría ser indicio de que la modalidad es más útil en la clasificación de los diversos tipos de actos del *common ground* y menos útil en los de *obligaciones*. Podrían existir al menos dos causas que expliquen el bajo nivel de contribución de la modalidad. Una es que la modalidad predicha optimizada, al provenir de un modelo de aprendizaje automático, contiene ruido informativo (incertidumbre, margen de error) que se suma al introducido por los demás atributos predictores. Otra causa es que un oyente usa además información del contexto del diálogo; por ejemplo, el o los tipos de actos de diálogo previos, el rol que su interlocutor está desarrollando en el diálogo, entre otros. Estas y otras fuentes de información complementarias le permiten al oyente interpretar actos de habla indirectos, altamente frecuentes en el habla espontánea.

La contribución del rol del hablante al reconocimiento tanto de la modalidad del enunciado como del tipo de acto de diálogo es un fenómeno contingente porque está determinado por la configuración particular del corpus analizado. Sin embargo, resulta útil para el objetivo práctico de implementar sistemas de reconocimiento automático y sería esperable observar este mismo comportamiento en diálogos prácticos de otros dominios. Con base en los análisis estadísticos, puede afirmarse que ciertas modalidades de enunciado y ciertos actos de diálogo son más frecuentes y, por lo tanto, más probables, en alguno de los dos roles en particular.

En conclusión, los resultados sugieren que al usar solamente información entonativa, es decir, las etiquetaciones INTSINT del inicio y del final del contorno, la duración del enunciado y el número de etiquetas INTSINT como predictores, sin usar otros atributos, se logra un reconocimiento pobre del acto de diálogo; la *accuracy* (exactitud) es menor que 50% y *Kappa* es menor que 0.2 (en promedio). Agregando un atributo de modalidad del enunciado reconocida automáticamente, como uno de los predictores, no se producen mejoras significativas; además, las medidas *F* de todos los actos de diálogo es menor que 0.5, así que la entonación, incluyendo en esta acepción a la modalidad, es un predictor débil para el acto de diálogo. La exactitud del reconocimiento de acto de diálogo mejora si se usan adicionalmente otros atributos no entonativos como predictores. Los evaluados son el rol del hablante y el tipo del acto de diálogo del plano complementario, es decir, la etiqueta del acuerdo como uno de los predictores para obligaciones y la etiqueta de obligaciones como uno de los predictores para el acuerdo. Los resultados óptimos se logran al incluir atributos del acto previo como predictores.

El enfoque del presente estudio es diferente a los enfoques anteriores ya que al usar el esquema DIME-DAMSL se logra una alta consistencia en los etiquetajes del acto de diálogo y al usar las secuencias INTSINT se obtiene una representación de la entonación con mayor nivel de abstracción.

6.9 Aplicaciones potenciales

Algunas aplicaciones potenciales de los resultados presentados podrían ser la implementación de: *a)* herramientas para etiquetación automática de la modalidad del enunciado y *b)* soporte a la etiquetación automática de actos de diálogo basada en parseo y modelos de lenguaje. Dichas implementaciones pueden aprovechar tanto los modelos basados en aprendizaje automático como los análisis estadísticos; es decir, se podrían implementar las reglas *if-then* de los modelos de árbol y también usar los análisis estadísticos de los etiquetajes para establecer parámetros de configuración. En los primeros tres casos sería necesario automatizar el etiquetaje de entonación basado en MOMEL e INTSINT; para ello, tendría que realizarse la extracción de la f_0 (frecuencia fundamental) de la señal del habla mediante alguno de los algoritmos más conocidos (*AMDF*, autocorrelación, función *comp*, etc.); después, la estilización del contorno y luego la asignación automática de etiquetas INTSINT. A continuación se describen algunas posibles implementaciones de las cuatro aplicaciones propuestas.

6.9.1 Etiquetación automática de la modalidad del enunciado

Se podría implementar una herramienta para la etiquetación automática de la modalidad del enunciado usando las reglas de los árboles de decisión creados para tal fin; habría dos posibles implementaciones: 1) cuando *no* se conoce el rol del hablante y 2) cuando sí se conoce. En el primer caso, el árbol se alimentaría con la duración del enunciado (en *mseg*) y las primeras dos y las últimas dos etiquetas INTSINT; en la segunda implementación, además de los tres atributos mencionados, se incluiría el rol del hablante. La segunda corresponde a la denominada *modalidad predicha optimizada*.

6.9.2 Soporte a la etiquetación automática de actos de diálogo

La etiquetación automática de actos de diálogo basada en parseo y modelos de lenguaje de los tipos de actos de diálogo podría beneficiarse del uso de los modelos propuestos en el presente trabajo. Al igual que en la presente investigación, los modelos se implementarían por separado: uno para el plano de las *obligaciones* y otro para el del *common ground*. Para cada uno de ellos habría cuatro configuraciones posibles, dependiendo de los atributos predictores disponibles:

- a) Las primeras 3 y las últimas 2 etiquetas INTSINT y la duración del enunciado
- b) Duración del enunciado y modalidad, reconociendo automáticamente esta última a partir de las primeras 2 y últimas 2 etiquetas INTSINT y de la duración del enunciado
- c) Duración del enunciado, modalidad predicha *optimizada*, rol del hablante y tipo de acto de diálogo del plano complementario (*obligaciones* para reconocer del *common ground* y

viceversa); la modalidad predicha *optimizada* se reconocería a partir de la duración, el rol del hablante y las primeras 2 y últimas 2 etiquetas INTSINT. La etiqueta del denominado *plano complementario* sería la proveniente del proceso de parseo y de los modelos de lenguaje de los tipos de actos.

- d) Rol del hablante, tipo de acto de diálogo del plano complementario, tipo de acto de diálogo de las obligaciones y del *common ground* del enunciado previo, modalidad predicha *optimizada* (ésta es requerida para predecir en el plano de *obligaciones*), primeras 3 etiquetas INTSINT, número total de tonos INTSINT, y últimas 2 y primeras 2 etiquetas INTSINT (éstos dos últimos atributos, para predecir actos del *common ground*). Igual que en el inciso *c*, la etiqueta del acto del plano complementario se obtendría del parseo y de los modelos de lenguaje de tipos de actos.

Sería pertinente evaluar empíricamente el desempeño de las implementaciones descritas en los incisos *c* y *d*, con el propósito de cuantificar la mejora obtenida al usar en combinación el parseo y los modelos de lenguaje de tipos de actos de diálogo con los árboles de decisión planteados en el presente trabajo.

6.9.3 Trabajo futuro

Uno de los temas de trabajo futuro que podrían explorarse sería la construcción de modelos de lenguaje para los diversos tipos de acto de diálogo a partir de los etiquetajes de partes del habla (P.O.S.); la exactitud de estos modelos podría compararse con la de los modelos aquí expuestos.

También sería interesante evaluar la contribución de otros atributos para el reconocimiento (o predicción) del tipo de acto de diálogo; entre éstos, podrían considerarse aquellos que indican la fase (o subfase) de la transacción dentro de la cual ocurre el acto; es decir, fase de *especificación* de la intención o fase de *satisfacción*. Un atributo para representar si el enunciado constituye un *cargo* o un *abono* podría enriquecer los modelos y sería posible que su exactitud se incrementara.

Sería valioso abordar un análisis más profundo y extenso de la correlación entre actos de diálogo del plano de *obligaciones* y los del *common ground* para determinar la posible existencia de una relación causal. Por ejemplo, habría que responder a las siguientes preguntas: ¿por qué el acto de *common ground* del enunciado previo es el segundo mayor contribuyente para ambos planos? ¿Por qué su contribución para predecir el acto de *common ground* es mayor que para el de las *obligaciones*? ¿Por qué para el reconocimiento de actos de *obligaciones* el rol del hablante contribuye más que la entonación y para el reconocimiento del *common ground* ocurre lo contrario?

Finalmente, se podrían desarrollar las implementaciones de las aplicaciones potenciales sugeridas.

APÉNDICES

Apéndice I. El Corpus DIME

En este trabajo el enfoque hacia el análisis tanto de la entonación como de los actos de diálogo es empírico, así que se requieren datos de diálogos reales. Estos diálogos necesitan cumplir algunas restricciones para ser útiles a nuestros objetivos: (1) conversación orientada a tarea y (2) representatividad del habla espontánea. La primera restricción surge porque esta clase de diálogo es uno de los más atractivos para la implementación de sistemas de interacción humano-computadora y también porque el número y la complejidad de los tipos de actos de diálogo en este tipo particular de diálogo es menor que los presentes en otros tipos de conversación, lo cual facilita los análisis bajo condiciones controladas. Segundo, los datos empíricos; es decir, los enunciados en los diálogos, deben reflejar la conducta de un usuario hablando tan espontáneamente como sea posible mientras participa en un diálogo orientado a tarea. Estos dos aspectos determinan el requerimiento de usar diálogos generados mediante un protocolo similar al del *Mago de Oz* (Dahlback *et al.*, 1993), donde los interlocutores desempeñan roles de *usuario* y de *sistema*.

Dentro del Proyecto DIME (Pineda *et al.*, 2002) está disponible un corpus que satisface los requerimientos descritos anteriormente, el Corpus DIME (Pineda *et al.*, 2007), una colección de archivos de video, audio y etiquetaciones que contiene 26 diálogos donde los participantes desarrollan alguno de dos roles: *sistema* o (*Mago*), que actúa como si fuera la computadora, y *usuario*, que actúa como usuario del sistema. La meta en los diálogos es colocar y organizar muebles en una cocina virtual usando una herramienta de software de diseño asistido por computadora según especificaciones dadas a cada usuario en dibujos sobre papel. El sistema participa en cada diálogo desarrollando acciones gráficas en el ambiente virtual; p. ej. agregando un mueble, desplazándolo, borrándolo, etc. según las especificaciones dadas por el usuario mediante lenguaje hablado; el sistema puede hablar con el usuario para ofrecer o para pedir información.

Se definieron niveles de etiquetación tomando en cuenta los datos de índole fonética o fonológica que las investigaciones previas en el área han mostrado o sugerido como correlacionados con el acto de diálogo o la modalidad del enunciado. Por lo tanto, las etiquetaciones se definieron como se describe a continuación: primeramente, la transcripción ortográfica (transliteración) y la segmentación en enunciados; después, en el nivel fonético segmental, alófonos; en el nivel fonético suprasegmental, sílabas fonéticas y contornos entonativos; en el nivel fonológico, palabras, partes del habla (*Parts of Speech, POS*), índices de ruptura (*break indices*) del esquema *Sp-ToBI* (Beckman *et al.*, 2002) y modalidad del enunciado. Finalmente, se implementó un nivel para actos de diálogo, etiquetado con DIME-DAMSL (Pineda *et al.*, 2006 y 2006a).

La etiquetación de actos de diálogo con DIME-DAMSL es uno de los elementos de información más importantes para el presente trabajo porque permite analizar el fenómeno en cuestión y representar el dato *target*, es decir, el dato a predecir, en los modelos de reconocimiento automático. Esta etiquetación garantiza una consistencia suficientemente alta para análisis estadísticos y para modelos de aprendizaje de máquina.

Una vez determinados los niveles de etiquetación, se eligieron esquemas de etiquetación adecuados. La transcripción ortográfica se hizo con los caracteres del alfabeto español. Los alófonos y las sílabas fonéticas y las palabras fueron etiquetados con el alfabeto fonético MexBet (Cuétara, 2004). Las etiquetas para el nivel de P.O.S. fueron establecidas y refinadas por Moreno (2006). Los índices de ruptura fueron etiquetados con una adaptación especial del esquema Sp-ToBI, desarrollado para los propósitos particulares del proyecto DIME. La entonación fue etiquetada semi-automáticamente con el sistema Internacional de Transcripción para Entonación (*International Transcription System for Intonation*), INTSINT (Hirst *et al.*, 2000). La modalidad del enunciado fue etiquetada manualmente basándose tanto en la percepción acústica de la entonación como en la estructura sintáctica del enunciado.

Además de la etiquetación de actos de diálogo, la de entonación es otra importante fuente de información para analizar el fenómeno bajo estudio. Su relevancia proviene de la posibilidad de abstraer el contorno entonativo de un enunciado en forma de dato categórico simple, es decir, una cadena de etiquetas de tonos, o una estructura de datos definida por tonos y marcas de tiempo, para usarse como uno de los predictores del acto de diálogo en modelos de aprendizaje automático.

I.1 Motivación y objetivos

El proyecto DIME, Diálogos Inteligentes Multimodales en Español (Pineda *et al.*, 2002) tiene entre sus objetivos principales la creación de sistemas de administración de diálogo para la interacción humano-computadora. Dentro de los recursos desarrollados en el contexto del proyecto, se creó el denominado Corpus DIME (Pineda *et al.*, 2007), una serie de grabaciones de audio y video que contienen diálogos humano-humano; el propósito del corpus es proporcionar un recurso empírico para el estudio de los fenómenos fonéticos, fonológicos y de estructura y de administración del diálogo.

Los diálogos del corpus pertenecen a la categoría de los denominados *diálogos prácticos*, según la acepción de Allen *et al.* (2001); es decir, son conversaciones orientadas a tareas dentro de dominios específicos. En ellas se asumen dos hipótesis: primeramente, la denominada *del diálogo práctico*, involucrando que en éste la competencia requerida por los interlocutores es más simple que en la conversación de tipo general y, segundo, la hipótesis de la *independencia del dominio*, que establece que la administración del diálogo y la interpretación del lenguaje son independientes de la tarea desarrollada.

Mediante el estudio de los diálogos prácticos se pueden producir modelos descriptivos, reconocedores o predictores que podrían ser útiles para implementar aplicaciones.

Se decidió que los diálogos trataran sobre un dominio que no demandara conocimientos especializados de los hablantes, que permitiera el uso del lenguaje hablado espontáneo y que facilitara la grabación. Se identificó que el dominio del diseño arquitectónico de interiores, apoyado en software, era adecuado para el propósito. En particular, se decidió que los participantes desarrollaran la tarea de colocar muebles en un espacio arquitectónico virtual que representa a una cocina.

I.2 Condiciones de la grabación

El Corpus DIME fue grabado con la participación de aproximadamente 15 hablantes del Español de la ciudad de México. Los hablantes tienen nivel educativo de estudios profesionales y algunos de ellos también de postgrado; la mayoría están relacionados con el área de la computación.

Cada uno de los hablantes, denominados *usuarios*, interactúa en una sesión de trabajo con un hablante en particular, denominado *sistema* o *Mago*. Cada sesión constituye un diálogo práctico, durante el cual un usuario y el sistema colaboran mediante una interfaz de software de C.A.D. para diseñar una cocina. El usuario da órdenes o plantea preguntas al sistema para seleccionar y colocar muebles en el espacio virtual de diseño. La tarea de diseño es guiada de modo general por un dibujo en papel en manos del usuario que muestra el estado final deseado de la cocina; no se usa un guión escrito, de modo que el lenguaje utilizado por los interlocutores es espontáneo. Se definieron algunas restricciones de diseño, que previamente se informaron a los participantes. Dichas restricciones involucran algunas reglas sobre la disposición espacial de los muebles en la cocina virtual y permiten adicionar un cierto grado de complejidad en la tarea, haciendo que los fenómenos observados en las conversaciones presenten una mayor variedad y riqueza para el análisis.

La fig. I.1 muestra la ubicación del hardware y de los interlocutores al desarrollar los diálogos. El usuario y el sistema se ubican en habitaciones distintas, separadas por una pared que impide el contacto visual y auditivo directo. En su locación, el usuario tiene a cargo una computadora personal (identificada en la figura como *CPU3*) que, además del monitor, el teclado y el *mouse*, tiene un micrófono y un altoparlante. En la locación del sistema hay dos computadoras (identificadas en la figura como *CPUI* y *CPU2*); *CPUI* tiene, además del monitor, el teclado, el *mouse*, un micrófono y un altoparlante, un *multiplexer* y un *switch*, cuyo uso se explicará más adelante. En *CPUI* está instalado el software de C.A.D., controlado por el sistema; a través del *multiplexer* existe una conexión hacia *CPU3* que permite al usuario ver lo mismo que el sistema. Los dos interlocutores pueden conversar mediante sus respectivos micrófonos y altoparlantes. El usuario no tiene permitido el manejo de los muebles (agregar, mover, borrar); sólo se le permite señalar objetos o zonas sobre el espacio virtual, lo cual realiza mediante el *mouse*. El manejo de los muebles es responsabilidad exclusiva del sistema. Guiándose por el dibujo en papel, el usuario dirige al sistema mediante lenguaje hablado para seleccionar y colocar los muebles requeridos en las ubicaciones especificadas. El *switch* le permite al sistema activar o desactivar el *mouse* del usuario, de modo que éste último no puede usarlo cuando el sistema está

desarrollando alguna actividad en el ambiente virtual. La computadora *CPU2* se usa para grabar el audio de la conversación y el video de los eventos visuales ocurridos en el ambiente de la herramienta C.A.D. El audio y el video se graban en archivos que después se someten a un proceso de edición y segmentación para producir los audios correspondientes a los enunciados expresados por cada interlocutor.

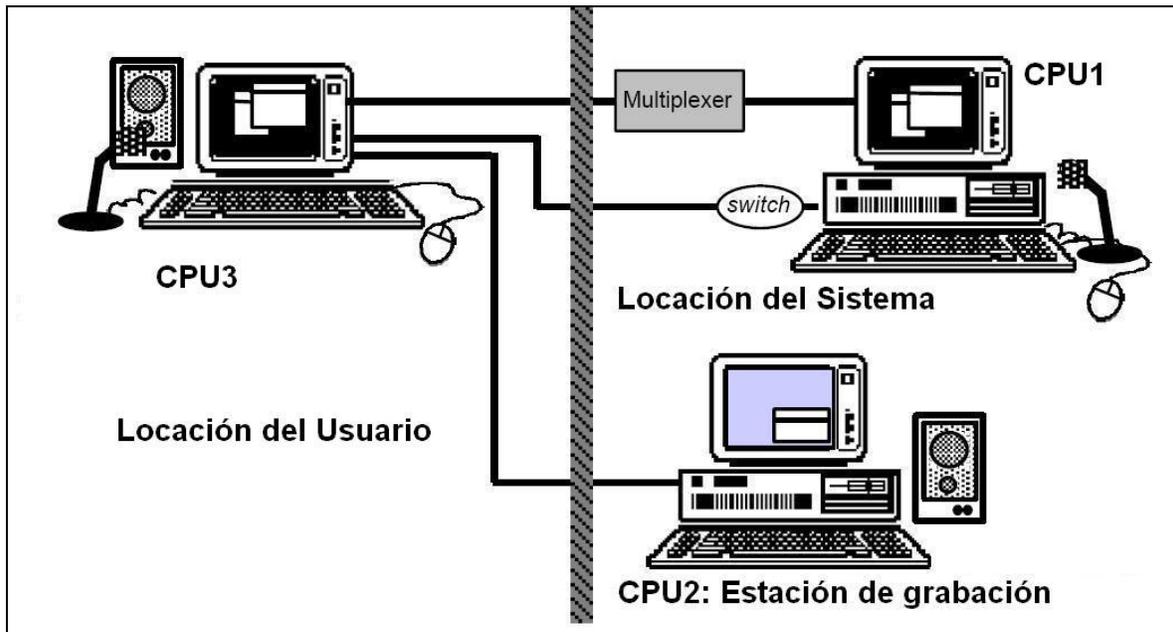


Figura I.1 Configuración de la instalación de grabación del corpus DIME.

El corpus está constituido por los audios y videos de 26 diálogos. La fig. I.2 muestra una vista de la interfaz del software de C.A.D. al ser utilizado en una sesión de trabajo por los participantes de uno de los diálogos. La interfaz muestra el espacio virtual de la cocina y está dividida en tres secciones que se describen a continuación. En la esquina superior izquierda se presenta la vista de planta de la cocina, en la cual los muebles se representan en forma transparente y vistos desde arriba; en la esquina superior derecha se muestra la vista de frente, con los muebles en color. La vista de frente permite la observación de la cocina desde diversos puntos de vista, siendo posible hacer rotaciones al espacio virtual alrededor del eje vertical. Finalmente, la sección inferior de la interfaz se usa para ver catálogos de muebles que la herramienta tiene disponibles para colocar en la cocina. El sistema muestra al usuario los catálogos de muebles, entre los cuales éste va seleccionando aquellos que le resultan satisfactorios según las especificaciones dadas en el dibujo en papel y según las restricciones de diseño.

Una vez generadas las grabaciones de audio (en formato *.wav*) y video (en formato *.avi*) de cada diálogo, se produjo manualmente su correspondiente transcripción ortográfica (transliteración), anotando los enunciados

expresados por cada hablante. Para ello, los anotadores observaron y escucharon las grabaciones usando software de uso típico, como Windows Media Player, Real Player y similares. La transcripción se hizo usando editores de texto sencillos, produciendo archivos en formato *.txt*. La fig. I.3 presenta una muestra de la transcripción ortográfica de un diálogo.

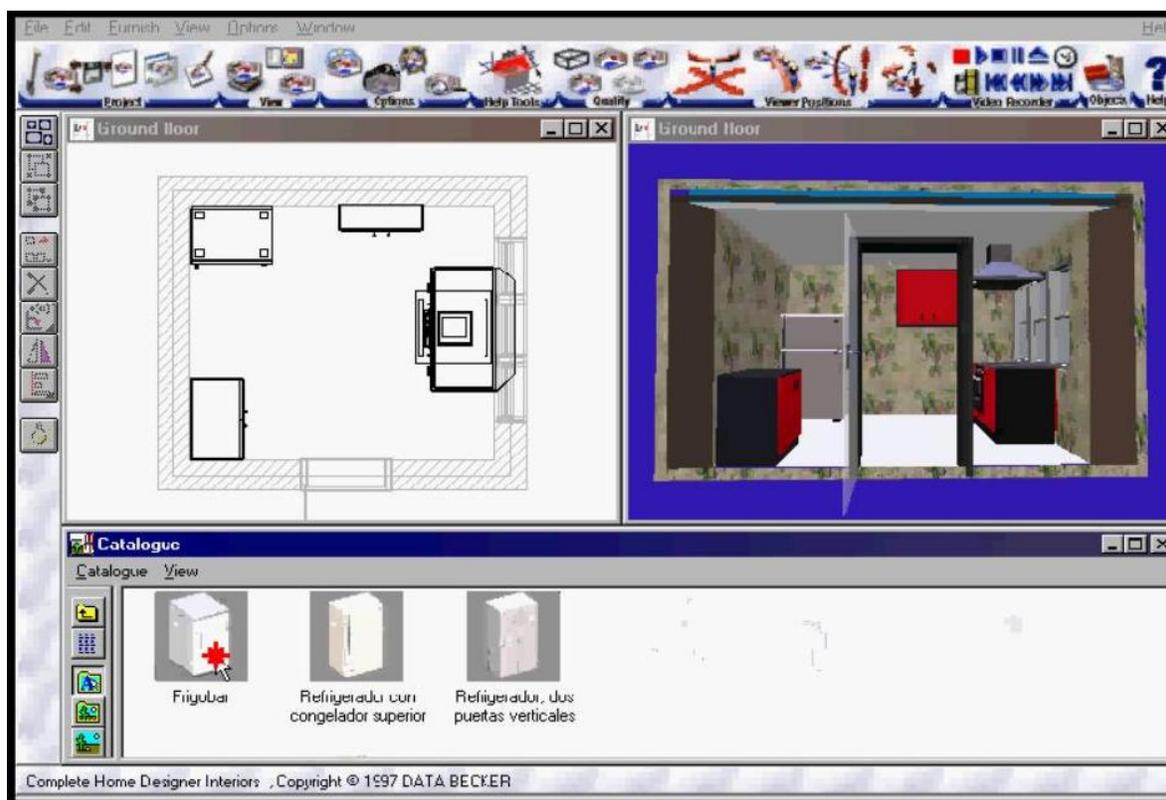


Figura I.2 Vista del software de C.A.D. usado en los diálogos del Corpus DIME.

En la transcripción, la secuencia *utt* seguida por un número es el identificador del enunciado; *s* indica un enunciado emitido por el sistema y *u* uno por el usuario. Si la transcripción de un enunciado no especifica *s* ni *u*, se asume que el hablante fue el mismo que en el enunciado inmediato anterior. Las palabras emitidas por el hablante son el contenido principal de la transcripción; sin embargo, también se transcriben algunos eventos del lenguaje hablado que podrían tener algún interés para el estudio del diálogo o que podrían requerir un manejo especial, como las pausas, etiquetadas como *<sil>*; los sonidos no vocales, como los suspiros y los chasquidos de labios, marcados como *<no-vocal>* y el ruido de fondo, etiquetado como *<ruido>*. Por otra parte, también se etiquetó la omisión de alófonos o sílabas en una palabra por parte del hablante; para ello, se marcan entre corchetes [y] los fonemas o sílabas de la palabra que no fueron pronunciados por el hablante; p. ej. en *utt19*: *u*: *<no-vocal> ne[ce]sito un fregadero*, la sílaba *ce* no fue pronunciada.

La segmentación de enunciados en la transcripción ortográfica se basa primordialmente en los turnos de cada hablante; en segundo término, cada turno puede estar formado por uno o más enunciados, asumiendo éste

último como una unidad de comunicación que expresa una intención distinguible de otras en un turno. Ejemplo de un turno formado por un solo enunciado es *utt10: s: ¿Cuál es la pared del fondo?* Un turno formado por varios enunciados es *utt2: u: Sí*, seguido por *utt3: Necesito una estufa*.

```
utt1 : s: quieres que traiga algún mueble a la cocina ?
utt2 : u: sí
utt3 :  necesito una estufa
utt4 : s: un segundo <ruido>
utt5 :  éstos son los modelos de estufas que contamos
utt6 :  tu estufas sencillas y estufas con alacenas laterales
utt7 : u: mm <sil> voy a seleccionar esa estufa
utt8 : s: <no-vocal> okey
utt9 : u: eh por favor lo necesito <sil> en <sil> en la pared del fondo
utt10: s: cuál es la pared del fondo ?
utt11: u: <ruido> a ver aquí
utt12: s: ahí ?
utt13: u: sí
utt14: s: <ruido> un segundo <ruido>
utt15:  ahí está bien ?
utt16: u: sí
utt17:  de momento sí
utt18:  eh
utt19:  <no-vocal> ne[ce]sito un fregadero
utt20: s: okey
utt21:  éstos son los cuatro tipos de fregaderos que tenemos
utt22: u: a ver éste
utt23:  selecciono éste fregadero con máquina lavatrastes
utt24: s: okey
utt25:  dónde quieres que lo ponga ?
utt26: u: a ver en este <sil> en este lugar
utt27: s: <ruido> quieres que ponga <sil> este fregadero <sil> a un lado de la estufa ?
utt28: u: no
utt29:  eh <sil> en la pared de las ventanas
utt30: s: okey
utt31:  <no-vocal> ahí está bien ?
utt32: u: por ahora sí
```

Figura I.3 Muestra de la transcripción ortográfica de un diálogo del corpus DIME.

En algunos casos la distinción entre dos enunciados de un turno es reconocible por una pausa del hablante; o bien, en ausencia de pausa, por una diferencia clara entre las intenciones de los diversos enunciados. Sin embargo, en otros casos no es fácil identificar si un turno contiene un solo enunciado o más; p. ej. *utt22: u: a ver éste*, seguido por *utt23: selecciono este fregadero con máquina lavatrastes*.

I.3 Determinación de los niveles de etiquetación

La transcripción ortográfica es la básica para poder desarrollar etiquetaciones en otros niveles de representación. Esa transcripción se desarrolló en la etapa posterior a la grabación de los diálogos. Las etiquetaciones del corpus se organizan en grupos denominados *niveles*; cada nivel atiende a un cierto tipo de fenómeno del habla y los fenómenos de interés se determinaron en función de los objetivos generales del

corpus y de los particulares del presente trabajo. A partir de ello, se identifica la necesidad de los siguientes niveles de etiquetación, adicionales a la transliteración:

- a) alófonos
- b) sílabas fonéticas
- c) palabras
- d) índices de ruptura (*break indices*) basados en el esquema Sp-ToBI (Beckman *et al.*, 2002)
- e) partes del habla (POS, *Parts Of Speech*)
- f) marcadores del discurso
- g) reparaciones del habla
- h) modalidad del enunciado
- i) contornos entonativos usando INTSINT (Hirst *et al.*, 2000)
- j) actos de diálogo

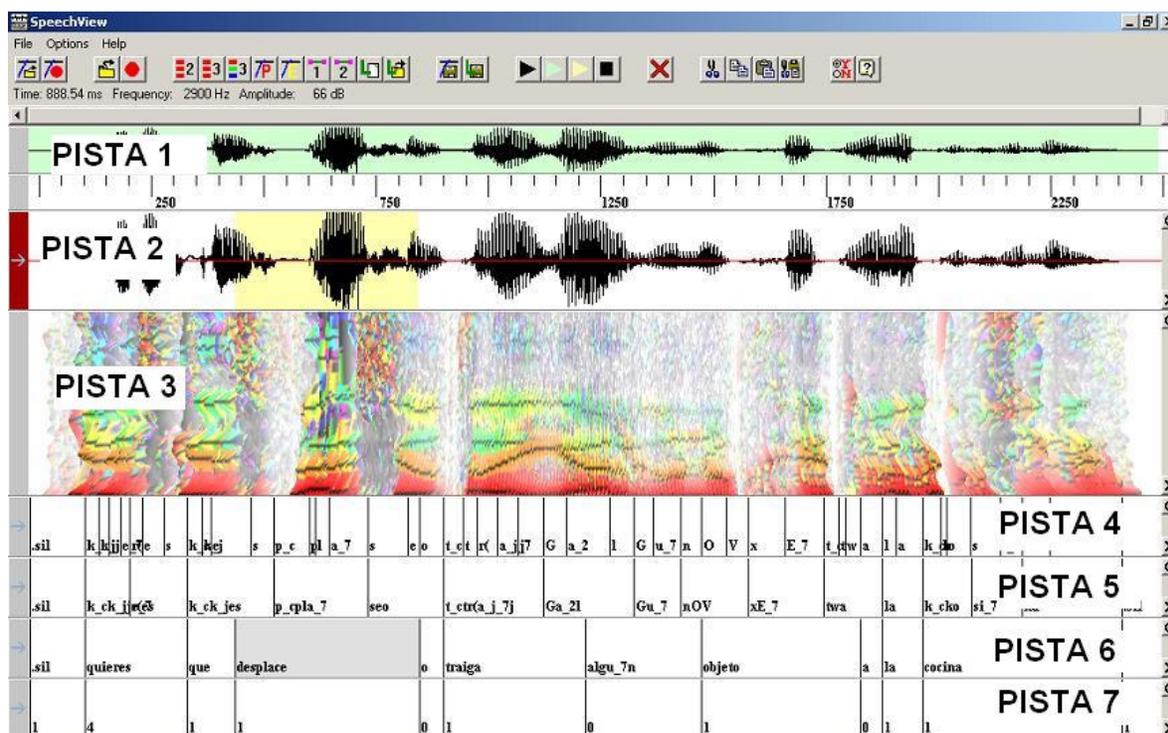


Figura I.4 Interfaz del software SpeechView con etiquetaciones alineadas sobre el eje de tiempo de un enunciado del corpus DIME.

La fig. I.4 presenta una vista de la interfaz de SpeechView (CSLU, 2004), el software utilizado para etiquetar la mayoría de los niveles fonéticos y fonológicos del corpus. El software permite ver el oscilograma de la señal de audio en las pistas (*tracks*) identificadas como 1 y 2 en la figura; sobre éstas, se pueden seleccionar fragmentos de la señal para escucharlos y verlos con una granularidad más detallada e identificar más fácilmente los límites entre segmentos o suprasegmentos, con lo que se favorece la alineación de las etiquetas

con el tiempo en el cual ocurre el fenómeno etiquetado. Entre las pistas 1 y 2 se muestra una escala para medición del tiempo, dada en milisegundos.

En la pista numerada en la figura con 3 se observa el espectrograma de la señal, pudiendo éste visualizarse en colores o en blanco y negro. El espectrograma muestra los formantes de la frecuencia fundamental de la señal, contribuyendo así al reconocimiento visual de fonemas y alófonos; ésto puede resultar útil cuando el habla tiene una alta velocidad, o bien, cuando el habla con velocidad típica no es inteligible claramente.

La herramienta ofrece la funcionalidad de mostrar la frecuencia fundamental (f_0) (no mostrada en esta figura). La f_0 ofrece una vista del contorno entonativo del enunciado y de la ocurrencia de períodos de sonoridad o de sordera, éstos últimos asociados a la presencia de fonemas o alófonos sonoros, cierres u oclusiones fonéticas, pausas, etc. Aunque SpeechView ofrece algunas funcionalidades para analizar el contorno entonativo, no dispone de las suficientes para producir los etiquetajes requeridos en la presente investigación; por ello, se decidió usar otra herramienta de software para este propósito: *Motif Environment for Speech* (Espesser, 1999), descrita en la sección de entonación de este apéndice.

La pista 4 sirve para la etiquetación de alófonos, en la cual se usa el alfabeto MexBet (Cuétara, 2004), descrito en la siguiente sección.

Otras herramientas de etiquetación usadas en el Corpus DIME son el transcriptor semi-automático Transcribemex (Cuétara y Villaseñor, 2004) y Microsoft Excel.

I.3.1 Alfabeto MexBet y herramienta Transcribemex

Un aspecto crítico al etiquetar un corpus de habla es la selección del alfabeto fonético para hacer las etiquetaciones fonéticas y fonológicas segmentales. La selección es relevante porque de ella dependen la riqueza y la versatilidad de uso de las etiquetaciones, lo cual tiene consecuencias en procesos posteriores de análisis de los fenómenos del habla o en la creación de modelos acústicos para sistemas de reconocimiento automático. MexBet, basado en diversos alfabetos fonéticos precedentes, es el elegido para etiquetar el corpus DIME, ya que fue creado especialmente para representar los fonemas y alófonos del Español de la ciudad de México. MexBet se creó a partir del análisis de datos empíricos de habla obtenidos del corpus DIME e intenta facilitar los análisis y las implementaciones computacionales. Contiene una serie de símbolos fonéticos y alofónicos basados en los caracteres del código ASCII e incluye reglas que representan a los contextos en los cuales ocurren los diversos alófonos en particular; es decir, una regla de contexto está expresada como una combinación de alófonos que ocurre con una frecuencia suficientemente alta en el habla para ser relevante.

En MexBet, la necesidad de analizar los fenómenos fonéticos segmentales en diversos niveles de granularidad derivó en el establecimiento de tres conjuntos de etiquetas, denominados T22, T44 y T54, correspondiendo a 22, 44 y 54 etiquetas, respectivamente. A mayor número de etiquetas disponibles corresponde una mayor variedad de fenómenos etiquetables, de modo que T22 permite la granularidad más gruesa (*transcripción ancha*), T54 la más fina (*transcripción estrecha*) y T44 una intermedia. El conjunto usado en la etiquetación de alófonos del corpus DIME es T54, en el cual se representarán 37 alófonos (incluidos también en el transcriptor Transcribemex), mas los 8 cierres de oclusivas y africadas ($[p_c, t_c, k_c, b_c, d_c, g_c, tS_c, dZ_c]$) y las 9 vocales que pueden recibir acento ($[i_7, e_7, E_7, a_j_7, a_7, a_2_7, O_7, o_7$ y $u_7]$). Las tablas I.1, I.2 y I.3 reproducen el conjunto de etiquetas T54 de MexBet.

Consonantes	Labiales	Labiodental	Dentales	Alveolares	Palatales	Velares
Oclusivas sordas	p / p_c		t / t_c		k_j	k / k_c
Oclusivas sonoras	b / b_c		d / d_c			g / g_c
Africada sorda					tS / tS_c	
Africada sonora					dZ / dZ_c	
Fricativas sordas		f	s_l	s		x
Fricativas sonoras	V		D	z	Z	G
Nasales	m		n_l	n	$n\sim$	N
Vibrantes				r / r		
Lateral				l		

Tabla I.1 Alfabeto MexBet, conjunto de etiquetas T54: consonantes.

Vocales átonas	Anteriores	Central	Posteriores
Paravocales	j		w
Cerradas	i		u
Medias	e		o
Medias abiertas	E		O
Abiertas	a_j	a	a_2

Tabla I.2 Alfabeto MexBet, conjunto de etiquetas T54: vocales átonas

Transcribemex es un software creado a partir del alfabeto MexBet y de sus reglas de contexto con el objetivo de producir automáticamente etiquetaciones fonéticas (fonemas y alófonos) usando como entrada la transcripción ortográfica (grafemas) de enunciados. Las transcripciones así generadas, aunque no están estrictamente alineadas sobre el eje de tiempo, resultan de gran utilidad como base para producir la transcripción alineada manualmente. Esta herramienta se basa en reglas de contexto para hacer la conversión de grafema a fonema o a alófono, halladas empíricamente.

Vocales tónicas	Anteriores	Central	Posteriores
Cerradas	i_7		u_7
Medias	e_7		o_7
Medias abiertas	E_7		O_7
Abiertas	a_j_7	a_7	a_2_7

Tabla I.3 Alfabeto MexBet, conjunto de etiquetas T54: vocales tónicas.

A continuación se describen los diversos niveles de etiquetación considerados en el Corpus DIME.

I.3.2 Alófonos

La etiquetación de alófonos permite representar fenómenos fonéticos que ocurren en el nivel lingüístico de granularidad más fina. Esta tarea se desarrolló usando el conjunto de etiquetas T54 de Mexbet. Las etiquetaciones de alófonos se encuentran disponibles en formato *.phn*, de la herramienta SpeechView. Este formato, en archivo de texto plano, permite una fácil transformación a formatos de archivo de otras herramientas mediante scripts programados en lenguajes como Perl, C, Java, etc.

I.3.3 Sílabas fonéticas

Una vez etiquetado el nivel de alófonos del enunciado, se desarrolla el etiquetaje de sílabas fonéticas; al hacerlo en este orden se aprovechan los archivos de etiquetaje alofónico, en los cuales ya se tiene una alineación que coincide con los límites de sílaba. Sólo se eliminan los límites de alófonos dentro de la sílaba y así se obtienen las sílabas fonéticas. Este nivel también se etiqueta con Mexbet y los archivos producidos se encuentran en formato de SpeechView.

I.3.4 Palabras

Después de etiquetar las sílabas fonéticas, estos etiquetajes se usan como base para etiquetar las palabras. Éstas se etiquetan usando los caracteres de las letras minúsculas del alfabeto del español; usando símbolos del alfabeto Mexbet para las vocales acentuadas y para la letra eñe (*ñ*).

I.3.5 Índices de disyunción (*break indices*)

En el esquema Sp-ToBI (Beckman *et al.*, 2002) los índices de disyunción (*break indices*) representan el grado de cohesión fonética que existe entre dos palabras consecutivas dentro de un enunciado. El conjunto de etiquetas se presenta en la tabla I.4.

Etiqueta	Fenómeno
0	Reducción silábica por contacto de vocales entre palabras (sinalefa)
1	Cualquier otra juntura ordinaria entre palabras
2	(Uso optativo para frase intermedia, grupo tónico o grupo clítico)
3	(Uso optativo para frase intermedia, grupo tónico o grupo clítico)
4	Límite de frase entonativa (grupo melódico)

Tabla I.4 Conjunto de etiquetas de *break indices* en Sp-ToBI (Beckman *et al.*, 2002).

I.3.6 Entonación

La entonación fue etiquetada con el esquema INTSINT (*International Transcription System for Intonation*) de Hirst (Hirst *et al.*, 2000) en un proceso semi-automático donde se usó el software MES (Espesser, 1999). Las etiquetas representan a los tonos relativos que ocurren a lo largo del contorno entonativo. El conjunto de etiquetas se presenta en la tabla I.5.

Etiqueta	Tono relativo
<i>M</i>	Medio (<i>medium</i>)
<i>T</i>	Máximo absoluto (<i>top</i>)
<i>B</i>	Mínimo absoluto (<i>bottom</i>)
<i>H</i>	Máximo local (<i>higher</i>)
<i>L</i>	Mínimo local (<i>lower</i>)
<i>U</i>	Ascendente (<i>upstep</i>)
<i>D</i>	Descendente (<i>downstep</i>)
<i>S</i>	Mismo tono que un tono predecesor

Tabla I.5 Conjunto de etiquetas de INTSINT (Hirst *et al.*, 2000).

Para producir los etiquetajes del contorno, el anotador se apoya tanto en una representación gráfica estilizada del mismo, generada con el algoritmo MOMEL (Hirst *et al.*, 2000), como en la percepción auditiva del audio en el archivo correspondiente. La fig. I.5 presenta un ejemplo de un contorno entonativo estilizado al cual se le han incorporado etiquetajes de tonos INTSINT.

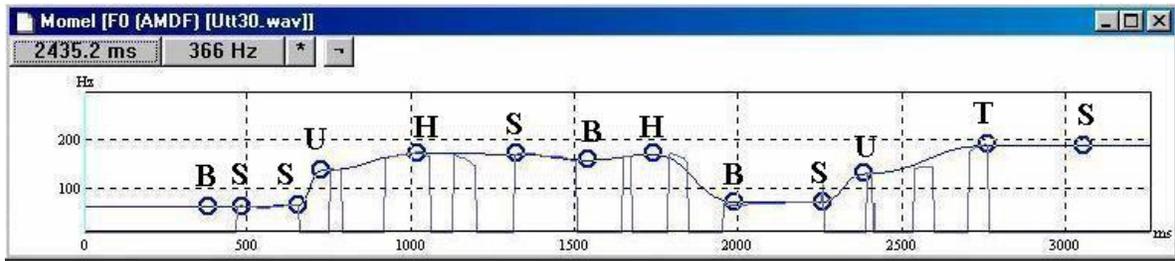


Figura I.5 Ejemplo de un contorno entonativo con etiquetaje INTSINT.

Dado que el algoritmo MOMEL es sensible a la presencia de ruido en los archivos de audio, no fue posible producir los etiquetajes INTSINT de todos los enunciados del corpus.

I.3.7 Partes del habla

El etiquetaje de partes del habla (*POS, Parts of Speech*), entendido también como etiquetaje de categorías léxicas, se realizó usando el conjunto de etiquetas definido en (Moreno y Pineda, 2006), reproducido en la tabla I.6.

Etiqueta	Categoría
<i>N</i>	Sustantivo
<i>V</i>	Verbo
<i>VAM</i>	Verbo auxiliar – modal
<i>VC</i>	Verbo con pronombre clítico
<i>A</i>	Adjetivo
<i>AD</i>	Adjetivo demostrativo
<i>TD</i>	Artículo determinado
<i>TI</i>	Artículo indefinido
<i>R</i>	Adverbio
<i>RI</i>	Adverbio interrogativo
<i>RR</i>	Adverbio relativo
<i>RN</i>	Adverbio de negación
<i>RA</i>	Adverbio de afirmación
<i>P</i>	Pronombre
<i>PD</i>	Pronombre demostrativo
<i>PR</i>	Pronombre relativo
<i>PI</i>	Pronombre interrogativo
<i>PC</i>	Pronombre clítico
<i>S</i>	Preposición
<i>C</i>	Conjunción

Tabla I.6 Conjunto de etiquetas para partes del habla, definido en (Moreno y Pineda, 2006).

I.3.8 Marcadores del discurso

Un marcador del discurso es una palabra o frase que se usa con un sentido distinto al que tiene sintácticamente, teniendo la función de guiar las inferencias que se realizan en la comunicación y que dan al

oyente información acerca de la estructura del discurso. El etiquetaje de marcadores del discurso se hizo usando el conjunto de etiquetas definido en (Moreno y Pineda, 2006), reproducido en la tabla I.7.

Etiqueta	Categoría
<i>MDN</i>	Sustantivo
<i>MDV</i>	Verbo
<i>MDA</i>	Adjetivo
<i>MDR</i>	Adverbio
<i>MDC</i>	Conjunción
<i>MDI</i>	Interjección
<i>MDK</i>	<i>Acknowledgment</i>
<i>MDeste</i>	Palabra <i>este</i>

Tabla I.7 Etiquetas para marcadores del discurso definidas en (Moreno y Pineda, 2006).

I.3.9 Reparaciones del habla (*speech repairs*)

Una reparación del habla (*speech repair*) es una disfluencia donde el hablante corrige, total o parcialmente, el contenido léxico de un enunciado. En la conversación esto ocurre generalmente porque el sujeto inicia la realización del enunciado antes de tener completamente definida la idea o intención que desea expresar. En el estudio de las reparaciones presentado por Moreno y Pineda (2006), utilizado en el Corpus DIME, se describen los siguientes fenómenos típicos: interrupción del enunciado para corregir o repetir algo de lo dicho; presencia de una pausa, acompañada por alguna palabra o frase mientras se piensa lo que se va a decir a continuación, o abandono intencional del enunciado para expresar una nueva idea.

La reparación del habla presenta generalmente una estructura constituida por: *reparandum*, *término de edición* y *alteración*. El *reparandum* es la parte del enunciado que el sujeto desea corregir. El *término de edición*, no siempre presente, establece un vínculo entre el *reparandum* y la *alteración*; puede ocurrir en forma de pausa, o bien, como una palabra o frase que permite al hablante pensar lo que va a decir. La *alteración* es el contenido léxico que sustituye al *reparandum* y constituye lo que el hablante realmente quiere decir. A partir de esta estructura, Moreno y Pineda (*op. cit.*) presentan una tipología de las reparaciones con un esquema de etiquetaje adaptado a las características del Corpus DIME.

I.3.10 Modalidad del enunciado

La modalidad de un enunciado especifica si su forma es interrogativa, declarativa, imperativa, etc. La etiquetación de la modalidad es especialmente importante en este trabajo, ya que ésta es una caracterización del enunciado basada en su forma sintáctica y en su contorno entonativo.

Para etiquetar la modalidad en el corpus DIME se establecieron el procedimiento y las convenciones que se describen a continuación.

El anotador escucha el archivo de audio correspondiente al enunciado, además de observar su transcripción ortográfica. Debe tomarse en cuenta la entonación que se percibe auditivamente, especialmente de la parte final. La etiquetación de modalidad se hace utilizando como herramienta una hoja de trabajo de Microsoft Excel, en la cual se cuenta con la transcripción ortográfica de cada enunciado.

Las etiquetas son: *dec* (declarativa), *imp* (imperativa), *int_sn* (interrogativa sí/no), *int_pron* (interrogativa pronominal), *int* (interrogativa genérica) y *otra*; ésta última incluye a cualquier otra modalidad que no corresponda a las categorías enumeradas anteriormente. A continuación se detalla el modo de uso de cada una.

Se etiqueta como *dec* el enunciado que no tenga entonación interrogativa ni exclamativa, y que además no contenga un verbo conjugado en modo imperativo; p. ej. *Así está bien.*

Se etiqueta como *imp* el enunciado que no tenga entonación interrogativa y que tenga un verbo conjugado en modo imperativo como verbo principal; p. ej. *Mejor ponme la estufa donde está la ventana.*

La etiquetación de la modalidad interrogativa se realizó en forma especial, produciendo dos datasets: en el primero, además de las etiquetas *dec*, *imp* y *otra*, se usan las tres variedades de la modalidad interrogativa especificadas anteriormente; es decir, interrogativa sí/no, interrogativa pronominal e interrogativa genérica. Se etiqueta como *int_sn* el enunciado que tenga entonación interrogativa y cuya respuesta esperada sea de tipo “sí/no”; p. ej. *¿Quieres saber qué mueble es éste?* Se etiqueta como *int_pron* el enunciado que, además de tener entonación interrogativa, inicia con un pronombre interrogativo, tal como *qué*, *dónde*, *cuánto*, etc.; p. ej. *¿Cuál es la diferencia entre el tercero y el cuarto?* La etiqueta *int* (interrogativa genérica) se usa para cualquier otro enunciado que tenga entonación interrogativa pero que no pertenezca a alguna de las dos primeras clases; p. ej. *¿Perdón?* En el segundo dataset, en lugar de usar tres etiquetas para las interrogativas, se usa sólo una (*int*) para cualquier tipo de éstas sin hacer distinción.

La modalidad se etiqueta como *otra* si:

- a) al final del audio hay ruido excesivo
- b) al final del audio hay un silencio demasiado prolongado, cuya duración supere a la de una pausa
- c) se trata de un sonido no verbal del hablante, como suspiro, respiración, risa, chasquido, etc.
- d) el audio del enunciado está cortado abruptamente haciendo que a la última palabra le falte algún fonema

Un caso que se maneja en forma especial es cuando el archivo de audio parece contener dos o más enunciados en lugar de uno sólo; ésto se observa en el corpus debido a la ambigüedad sintáctica o entonativa de algunos enunciados. En este caso la etiqueta de modalidad es la que corresponde al último enunciado que aparece en el audio, pudiendo ocurrir que ésta sea diferente de las de los enunciados que le anteceden en el mismo audio. Por ejemplo, en *Ahora necesito desplazar lo que es el extractor o qué qué es este objeto (d01, utt48)* parece haber dos enunciados: uno con entonación (y modalidad) declarativa (*Ahora necesito desplazar lo que es el extractor*) y uno con interrogativa (*o qué qué es este objeto*); en este caso la etiqueta asignada es la correspondiente a la interrogativa. Se asumió esta convención porque, según las principales teorías de la entonación del Español, la región final del contorno entonativo es la que determina en mayor grado la modalidad.

	A	B
1	ELOCUCIÓN	
	modalidad	
2	utt1 : s: quieres que desplace o traiga algún objeto a la cocina ?	int_sn
3	utt2 : u: <ruido> no	dec
4	utt3 : puedes mover la estufa hacia la izquierda ?	int_sn
5	utt4 : s: <ruido> hacia dónde ?	int_pron
6	utt5 : u: <ruido> hacia <sil> hacia la derecha	dec
7	utt6 : s: <no-vocal> hacia la derecha	dec
8	utt7 : <no-vocal> okey	dec
9	utt8 : que tanto quieres que la desplace ?	int_pron
10	utt9 : u: a la mitad del espacio que hay entre la ventana y la pared	dec
11	utt10 : s: okey	dec
12	utt11 : quieres que mueva este objeto <sil> hacia acá ?	int_sn
13	utt12 : u: no	dec
14	utt13 : hacia el otro lado	dec
15	utt14 : s: okey	dec
16	utt15 : este objeto hacia acá ?	int_sn
17	utt16 : u: sí	dec

Figura I.6 Muestra de etiquetaciones de modalidad del enunciado de un diálogo del corpus DIME

La modalidad fue etiquetada por un sólo anotador, quien además fue el mismo que formalizó las convenciones de etiquetación; por ello, aunque no se hizo una evaluación estadística de la consistencia de las etiquetaciones, es de esperar que ésta sea satisfactoria.

La fig. I.6 presenta una muestra de las etiquetaciones de modalidad de enunciado de uno de los diálogos del corpus DIME.

Las etiquetaciones de modalidad del enunciado se encuentran disponibles en archivos de formato *.xls* (Microsoft Excel) y también en *.csv* (valores separados por comas).

I.3.11 Actos de diálogo: DIME-DAMSL

La etiquetación de actos de diálogo, basada en DIME-DAMSL (Pineda *et al.*, 2006 y 2006a), está estructurada en tres niveles: el de las *transacciones* es el de mayor jerarquía, seguido por el de *cargos* y *abonos* y finalmente el de *actos de diálogo* en sí. La *transacción* consiste en un conjunto de enunciados dentro de los cuales se especifica y satisface una intención del *usuario*. Dentro de ella ocurren pares equilibrados de *cargos* y *abonos*, donde cada par está constituido por dos instancias de actos de diálogo. De la transacción se etiqueta su inicio y su fin; cada cargo y cada abono se etiquetan con el número del enunciado correspondiente; los actos de diálogo se etiquetan con el conjunto de etiquetas de DIME-DAMSL, considerando dos niveles granularidad del tipo de acto: DIME-DAMSL Preliminar, que agrupa a las etiquetas en *obligaciones* y *common ground* (dividiendo a éstas en *acuerdo* y *entendimiento*) y DIME-DAMSL Detallado, donde se agrupan en las dimensiones básicas de DAMSL (Allen y Core, 1997), adicionando la dimensión *graf* para acciones gráficas. Las tablas I.8, I.9, I.10 y I.11 presentan el conjunto de etiquetas de DIME-DAMSL.

Etiqueta	Acto
<i>dir-accion</i>	Directiva de acción
<i>sol-info</i>	Solicitud de información
<i>oferta</i>	Oferta
<i>compromiso</i>	Compromiso

Tabla I.8 Conjunto de etiquetas del plano de las *obligaciones*.

Los etiquetajes de actos de diálogo con DIME-DAMSL Preliminar y Detallado están disponibles en archivos con formato de Microsoft Excel (.xls).

Etiqueta	Acto
<i>opcion-abierta</i>	Opción abierta
<i>oferta</i>	Oferta
<i>afirm</i>	Afirmación
<i>reafirm</i>	Reafirmación
<i>acepta</i>	Aceptación
<i>rechaza</i>	Rechazo
<i>acepta-parte</i>	Aceptación parcial
<i>rechaza-parte</i>	Rechazo parcial
<i>espera</i>	Espera
<i>quizá</i>	Quizá
<i>conv-abre</i>	Convencional-abre
<i>conv-cierra</i>	Convencional-cierra

Tabla I.9 Conjunto de etiquetas del plano del *acuerdo*, subplano del *acuerdo*.

Etiqueta	Acto
<i>back-channel</i>	Back-channel
<i>ack</i>	Acknowledgment
<i>SNE</i>	Señal de no entendimiento
<i>repet-refr</i>	Repetición o parafraseo
<i>complement</i>	Complementación
<i>correccion</i>	Corrección

Tabla I.10 Conjunto de etiquetas del plano del *acuerdo*, subplano del *entendimiento*.

Etiqueta	Acto
<i>apt-obj</i>	Apuntar objeto
<i>apt-zona</i>	Apuntar zona
<i>apt-trayect</i>	Apuntar trayectoria
<i>apt-objs-coord</i>	Apuntar objetos coordinados
<i>despliega</i>	Desplegar catálogo
<i>rotar-vista</i>	Rotar la vista del ambiente virtual del software de C.A.D.
<i>agrega-obj</i>	Agregar objeto
<i>mueve-obj</i>	Mover objeto
<i>quita-obj</i>	Quitar objeto
<i>plan-graf</i>	Secuencia de acciones gráficas consecutivas
<i>visual</i>	Interpretación visual del resultado de acciones gráficas

Tabla I.11 Conjunto de etiquetas para acciones gráficas.

I.4 Control del proceso y de la calidad de la etiquetación

El orden en el cual se etiquetan los diversos niveles de un corpus puede simplificar o dificultar la tarea, ya que las etiquetaciones de algunos niveles pueden servir como base para producir las de otros. Por ello, al etiquetar el Corpus DIME se identificaron las ventajas de empezar con el nivel de alófonos, después el de sílabas fonéticas, después los de palabras, *break indices*, partes del habla, marcadores de discurso y reparaciones del habla. Esto garantiza que la correcta alineación en tiempo de los fenómenos etiquetados sobre los diversos niveles.

La calidad de las etiquetaciones se evaluó y controló como se describe a continuación:

Los etiquetajes de alófonos, sílabas fonéticas, palabras, *break indices*, POS, marcadores del discurso y reparaciones del habla fueron desarrollados por un equipo de anotadores y posteriormente se revisaron manualmente por los anotadores con más experiencia en el grupo. La modalidad del enunciado fue etiquetada por un único anotador que siguió convenciones formalizadas, lo cual ofrece altas probabilidades de consistencia. El etiquetaje de contornos entonativos estuvo sujeto a la verificación perceptiva de un único anotador. Finalmente, los etiquetajes de actos de diálogo se desarrollaron por dos anotadores expertos y la consistencia se evaluó con el método del estadístico *Kappa*.

Los 26 diálogos del Corpus DIME contienen aproximadamente 5,369 enunciados con una duración de 372 minutos, equivalente a 6.2 horas. La tabla I.12 presenta la información detallada de cada diálogo.

Diálogo	Duración		Turnos	Enunciados
	Segundos	Minutos		
d01	496	8.3	63	116
d02	1,080	18.0	111	196
d03	413	6.9	114	168
d04	1,256	20.9	261	384
d05	476	7.9	66	118
d06	1,656	27.6	230	371
d07	448	7.5	47	98
d08	754	12.6	85	162
d09	2732	45.5	357	692
d10	453	7.6	46	96
d11	1,211	20.2	139	285
d12	439	7.3	69	117
d13	837	14.0	105	191
d14	525	8.7	97	137
d15	346	5.8	54	90
d16	1,488	24.8	246	454
d17	655	10.9	136	237
d18	1,232	20.5	121	216
d19	609	10.1	62	105
d20	929	15.5	99	179
d21	263	4.4	49	69
d22	811	13.5	113	181
d23	353	5.9	49	81
d24	1,181	19.7	195	300
d25	488	8.1	73	116
d26	1,201	20.0	127	210
Total	22,333	372	3,114	5,369

Tabla I.12 Duraciones, número de turnos y número de enunciados en el Corpus DIME.

Apéndice II. DIME-DAMSL detallado y uso del estadístico Kappa

II.1 DIME-DAMSL Detallado

En el etiquetaje DIME-DAMSL Detallado, se usa el conjunto de etiquetas para actos de diálogo presentado en la tabla II.1, que es más específico que el usado en DIME-DAMSL Preliminar, en el cual se apoya. La mayoría de las etiquetas fueron tomadas de DAMSL, excepto las denominadas *segmento*, *inicio/fin de transacción* y las de la dimensión *graf*. La *sol-inf* se divide en tres categorías: 1) *pregunta sí/no* (*preg-s/n*), donde la respuesta esperada es un “sí” o un “no”; 2) *pregunta pronominal* (*preg-pro*), donde se usa un adverbio interrogativo, p. ej. *dónde*, *cuánto*, *cómo*, etc. y 3) *pregunta imperativa* (*preg-imp*), donde la solicitud de información tiene forma de orden o comando, p. ej. *Muéstrame los muebles que tengo*. En la *respuesta* (*resp*) se usa un número (o una serie) que indica el enunciado (o enunciados) previo(s) a los cuales se está respondiendo.

<i>Segmento</i> (núm. del enunciado inicial)	<i>Sol-inf: preg-s/n, preg-pro, preg-imp</i>	<i>Exclama</i>
<i>(i) inicio ó (f) fin de transacción</i>	<i>Influ-acc-fut-esc: dir-accion, opcion-abierta</i>	<i>Otra (cualquier otra función hacia delante)</i>
<i>Edo-com: uninterp, mono, abandonada</i>	<i>Comp-acc-fut-hab: oferta, compromiso</i>	<i>Acuerdo: ref-ant, acepta, acepta-parte, quiza, espera, rechaza, rechaza-parte</i>
<i>Niv-inf: tarea, admin-tarea, admin-com, otro</i>	<i>Conv: abre, cierra</i>	<i>Ent: ref-ant, SNE, ack, back-channel, repet-refr, complem, correccion</i>
<i>Dec: afirm, reafirm, otra</i>	<i>Resp: ref-ant o lista-ref-ant</i>	<i>Graf: apt-obj, apt-zona, apt-trayect, despliega, apt-objs-coord, mueve-obj, agrega-obj, quita-obj, plan-graf, visual, rotar-vista</i>

Tabla II.1 Dimensiones y etiquetas usadas en DIME-DAMSL detallado.

Las etiquetas de la dimensión *graf* fueron introducidas en el esquema para los actos de diálogo con modalidad gráfica que se presentan en el ambiente del software de C.A.D. usado por los diálogos del corpus DIME. La fig. II.1 presenta como muestra un *frame* del video de un diálogo del corpus; en la figura (región inferior izquierda) puede verse el apuntador señalando un objeto (frigoriferador).

La tabla II.2 describe las etiquetas de actos de diálogo con modalidad gráfica. Estos actos pueden clasificarse en las siguientes categorías: 1) *apuntar* (objeto, zona, trayectoria y objetos coordinados), 2) *actos informativos* (desplegar catálogo y rotar vista), 3) *manipulación de objeto* (agregar, mover, quitar, plan gráfico) y 4) *acto de observación* (etiqueta *visual*).

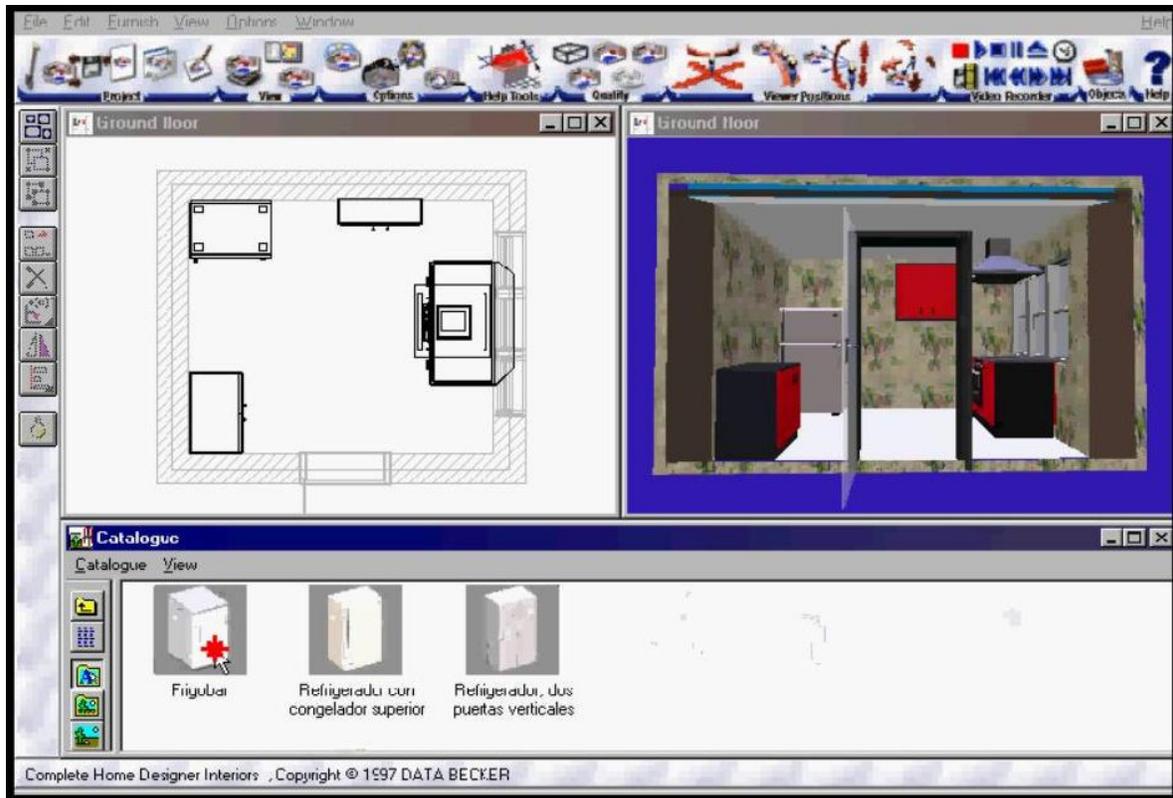


Figura II.1 Un *frame* de video de un diálogo del corpus DIME.

Etiqueta	Descripción
<i>apt-obj</i>	Se señala con el apuntador un objeto (mueble de cocina)
<i>apt-zona</i>	Se señala una zona del espacio virtual con el apuntador
<i>apt-trayect</i>	Se describe con el apuntador una trayectoria dentro del espacio virtual
<i>apt-objs-coord</i>	Se señalan con el apuntador varios objetos como una colección
<i>despliega</i>	Se despliega un catálogo de muebles
<i>rotar-vista</i>	La interfaz gráfica completa del ambiente virtual es girada para facilitar su observación
<i>agrega-obj</i>	Se agrega un nuevo objeto al ambiente
<i>mueve-obj</i>	Se mueve un objeto dentro del ambiente
<i>quita-obj</i>	Se borra un objeto del ambiente
<i>plan-graf</i>	Plan gráfico: se realiza una serie de actos con modalidad gráfica en una secuencia unitaria (p. ej. agregar un objeto e inmediatamente reubicarlo)
<i>visual</i>	Se observa con especial atención el resultado de una acto de diálogo que tiene modalidad gráfica

Tabla II.2 Etiquetas para actos de diálogo con modalidad gráfica (*graf*) de DIME-DAMSL.

El acto de girar la vista del ambiente virtual (*rotar-vista*), se considera un acto informativo porque tiene el propósito de dar al observador información gráfica que sólo es accesible en esa forma; por ello, su naturaleza es similar a la de *despliega*.

Los actos de diálogo con modalidad gráfica, excepto el etiquetado como *visual*, generan cargos y abonos en forma similar a la *afirmación*; la etiqueta *visual* es una excepción porque no introduce información al cuerpo de conocimiento compartido de los interlocutores y es un acto más bien pasivo.

El esquema DIME-DAMSL fue desarrollado en forma evolutiva; es decir, el conjunto de etiquetas y las convenciones de etiquetaje fueron refinadas a lo largo de una serie de etapas de etiquetación de dos diálogos del corpus DIME, realizada por equipos de anotadores del grupo DIME. Al final de cada etapa se evaluó con el estadístico Kappa (Carletta, 1996) la consistencia de los etiquetajes producidos para inicio/fin de transacción, segmento, cargos y abonos y actos de diálogo (DIME-DAMSL Preliminar) y también la consistencia de DIME-DAMSL Detallado. Cada discrepancia en etiquetaje al final de cada etapa fue discutida entre los etiquetadores para determinar su causa, que generalmente se redujo a alguna de tres posibilidades: error tipográfico, desconocimiento (u olvido) de la convención aplicable o falta de una convención para guiar el etiquetaje del tipo de situación. En este último caso se definían nuevas convenciones que se usaban por los anotadores en la etapa siguiente. Después de cuatro iteraciones de la tarea de etiquetación, los valores de Kappa finalmente obtenidos, presentados por Pineda et al. (2006 y 2006a), son mayores o iguales a 0.8, lo cual muestran que DIME-DAMSL es un esquema consistente y suficientemente confiable.

II.2 DIME-DAMSL: Aplicación

Para hacer etiquetaciones con DIME-DAMSL, las herramientas de software principales son formatos predefinidos en hojas de cálculo de Excel: se usa uno para *DIME-DAMSL Preliminar* (fig. II.2) y otro para *DIME-DAMSL Detallado* (figs. II.3, II.4 y II.5). Esto facilita la transformación de los etiquetajes a diversos formatos de archivo, así como el cálculo de valores Kappa y la producción de *datasets* para diversos análisis y modelaciones. Los formatos contienen la transcripción ortográfica de los diálogos del corpus DIME, segmentados a nivel de enunciado, así como columnas predefinidas para que el anotador teclee etiquetas o valores relacionados con éstas.

Para lograr una correcta interpretación de los actos de diálogo al etiquetar, es necesario observar y escuchar la grabación de video (fig. II.1) y audio del diálogo para analizar el contexto en el cual se da cada enunciado y cada evento con modalidad gráfica. Se analiza también la entonación del hablante, los fenómenos de traslape del habla de los interlocutores y la relación de temporalidad entre enunciados y eventos gráficos.

1	A	B	C	D	E	F	G	H	I	J	K										
												INICIO (i) O FIN (f) DE TRAN	SEGM (0 / 1)	OBLIGACIONES		COMMON GROUND				DIME-DAMSL	
														CARGOS	ABONOS	ACUERDO		ENTENDIMIENTO		OBLS	COMM. GR.
2	ELOCUCIÓN																				
3	utt1 : s: quieres que desplace o traiga algún objeto a la cocina ?	i				1					oferta, abre										
4	utt2 : u: <ruido> no						1				rechaza										
5	utt3 : s: puedes mover la estufa hacia la izquierda ?			3		3				dir-accion	dir-accion										
6	utt4 : s: <ruido> hacia dónde ?			4				4		sol-inf	sne										
7	utt5 : u: <ruido> hacia <sil> hacia la derecha				4	5			4	resp	afirm										
8	utt6 : s: <no-vocal> hacia la derecha										repeticion										
9	utt7 : <no-vocal> okey						5.3				acepta										
10	utt8 : que tanto quieres que la desplace ?			8						sol-inf											
11	utt9 : u: a la mitad del espacio que hay entre la ventana y la pared				8	9				resp	afirm										
12	utt10 : s: nkey										ack										
13	utt11 : quieres que mueva este objeto <sil> hacia acá ?			11				11		sol-inf	hold, repeticion, apt-obj, apt-zona										
14	utt12 : u: no				11				11	resp	rechaza, visual										
15	utt13 : hacia el otro lado					13					afirm										
16	utt14 : s: okey										ack										
17	utt15 : este objeto hacia acá ?			15				15		sol-inf	hold, repeticion, apt-										
18	utt16 : u: sí				15				15	resp	acepta										
19	utt17 : s: okey			17			13.9			compromiso	acepta										
20	utt18 : <ruido>				17.3	18				accion-graf	accion-graf										
21	utt19 : ahí está bien ? <no-vocal>			19		19				sol-inf	sol-inf										
22	utt20 : u: un poco menos por favor			20	19	20	19.18			resp, dir-accion	acepta-parte, dir-										
23	utt21 : s: okey			21			20			compromiso	acepta										
24	utt22 : <ruido>				21,20	22				accion-graf	accion-graf										
25	utt23 : ahí está bien ?			23		23				sol-inf	sol-inf										
26	utt24 : u: ahí está bien	f			23		23.22			resp	acepta										
27	utt25 : después <sil> me puedes poner <sil> [e] el extractor de aire encima de la <sil> de la estufa	i		25		25				dir-accion	dir-accion										
28	utt26 : s: okey			26			25			compromiso	acepta										
29	utt27 : <ruido> <no-vocal> así está bien ?									accion-graf, sol-	accion-graf										

Figura II.2 Formato en hoja de cálculo de Excel para etiquetaje de DIME-DAMSL Preliminar.

II.3 Uso del estadístico Kappa

Al definir un esquema de etiquetaje es necesario evaluar su consistencia; es decir, su capacidad para producir etiquetajes suficientemente similares por parte de anotadores independientes que etiquetan un mismo conjunto de instancias. El método más sencillo para evaluar la consistencia es computando el porcentaje simple de coincidencias entre los etiquetajes producidos por los diversos anotadores; sin embargo, este método tiene una deficiencia: no toma en cuenta la contribución del azar, ya que existe una probabilidad de que se obtenga la misma etiqueta de manera aleatoria, por lo cual los etiquetajes presentarían un porcentaje de acuerdo alto, cuando realmente no lo tendrían. Por lo tanto, el método del porcentaje simple es poco confiable y comúnmente se opta por el denominado *estadístico Kappa* (Carletta, 1996).

El estadístico Kappa es un método para medir la consistencia en tareas de etiquetación, en las cuales se aplican generalmente criterios clasificatorios predefinidos que pueden incluir un cierto grado de subjetividad. Kappa fue presentada por Siegel y Castellan (1988) para usarse en las ciencias del comportamiento; Carletta (1996) propuso usarla para análisis del discurso y para investigación en ciencias cognitivas.

Microsoft Excel - d12dime.damsl

Type a question for help

BC7

	A	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	
1	ELOCUCIÓN	SEGMENTO	INICIO / FIN DE TRANSACC. (i/f)	edo-com uninterp	edo-com mono	edo-com abandonada	niv-inf tarea	niv-inf admin-tarea	niv-inf admin-com	niv-inf otro	dec afirm	dec reafirm	dec otra	sol-inf preg-s/n	sol-inf preg-pro	sol-inf preg-imp	influ-acc-fut-esc dir-accion	influ-acc-fut-esc opcion-abierta	comp-ac oferta
2	utt1 : s: quieres que desplace o traiga algún objeto a la cocina ?						1												
3	utt2 : u: <ruido> no						1												
4	utt3 : puedes mover la estufa hacia la izquierda ?						1											1	
5	utt4 : s: <ruido> hacia dónde ?								1						1				
6	utt5 : u: <ruido> hacia <sil> hacia la derecha								1		1								
7	utt6 : s: <no-vocal> hacia la derecha								1										
8	utt7 : <no-vocal> okey						1												
9	utt8 : que tanto quieres que la desplace ?						1								1				
10	utt9 : u: a la mitad del espacio que hay entre la ventana y la pared						1					1							
11	utt10 : s: okey								1										
12	utt11 : quieres que mueva este objeto <sil> hacia acá ?								1						1				
13	utt12 : u: no								1										
14	utt13 : hacia el otro lado								1			1							
15	utt14 : s: okey									1									
16	utt15 : este objeto hacia acá ?								1						1				
17	utt16 : u: sí								1										
18	utt17 : s: okey						1												
19	utt18 : <ruido>						1												
20	utt19 : ahí está bien ? <no-vocal>						1	1						1					
21	utt20 : u: un poco menos por favor						1	1											1
22	utt21 : s: okey						1												
23	utt22 : <ruido>						1												
24	utt23 : ahí está bien ?						1	1						1					
25	utt24 : u: ahí está bien			f			1	1											
26	utt25 : después <sil> me puedes poner <sil> [e] el extractor de aire encima de la <sil> de la estufa			i					1										1
27	utt26 : s: okey								1										
28	utt27 : <ruido> que voy a hacer está								1										

Ready

start | etiquetasd12 | formatos_ algunos | Recycle Bin | Windows Media ... | cap_teoría dime... | Microsoft Excel - ... | EN | 11:24 PM

Figura II.3 Formato en hoja de cálculo Excel para etiquetaje DIME-DAMSL Detallado (parte 1 de 3).

Microsoft Excel - d12dime.damsl

Type a question for help

BC7

	A	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL
1	ELOCUCIÓN																		
2		comp-acc-fut-hab	comp-acc-fut-hab	conv	conv	performativo	exclama	otra	acuerdo	acuerdo	acuerdo	acuerdo	acuerdo	acuerdo	acuerdo	ent	ent	ent	ent
3		oferta	compromiso	abre	cierra				ref-ant	acepta	acepta-parte	quiza	espera	rechaza	rechaza-parte	ref-ant	SNE	ack	back-channel
4	utt1 : s: quieres que desplace o traiga algún objeto a la cocina ?	1		1															
5	utt2 : u: <ruido> no								1					1					
6	utt3 : puedes mover la estufa hacia la izquierda ?																		
7	utt4 : s: <ruido> hacia dónde ?															3	1		
8	utt5 : u: <ruido> hacia <sil> hacia la derecha																		
9	utt6 : s: <no-vocal> hacia la derecha															6			
10	utt7 : <no-vocal> okey								5.3	1									
11	utt8 : que tanto quieres que la desplace ?																		
12	utt9 : u: a la mitad del espacio que hay entre la ventana y la pared																		
13	utt10 : s: okey															9		1	
14	utt11 : quieres que mueva este objeto <sil> hacia acá ?								9				1			9			
15	utt12 : u: no								11					1					
16	utt13 : hacia el otro lado																		
17	utt14 : s: okey															13		1	
18	utt15 : este objeto hacia acá ?									13.9			1			13			
19	utt16 : u: sí								15	1									
20	utt17 : s: okey			1					13.9	1									
21	utt18 : <ruido>																		
22	utt19 : ahí está bien ? <no-vocal>																		
23	utt20 : u: un poco menos por favor																		
24	utt21 : s: okey			1						20	1			1					
25	utt22 : <ruido>																		
26	utt23 : ahí está bien ?																		
27	utt24 : u: ahí está bien									23.22	1								
28	utt25 : después <sil> me puedes poner <sil> [e] el extractor de aire encima de la <sil> de la estufa																		
29	utt26 : s: okey			1						25	1								
30	utt27 : <ruido> me puedes decir qué																		

Ready

start | etiquetad12 | formatos_al... | Recycle Bin | Windows Me... | cap_teoría... | Microsoft Ex... | untitled - Paint | EN | 11:25 PM

Figura II.4 Formato en hoja de cálculo Excel para etiquetaje DIME-DAMSL Detallado (parte 2 de 3).

Microsoft Excel - d12dime.damsl

File Edit View Insert Format Tools Data Window Help

Type a question for help

BC7

	A	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU	AV	AW	AX	AY	AZ	BA	E
1	ELOCUCIÓN																	
2	ent	ent	ent	ent	resp	graf	graf	graf	graf	graf	graf	graf	graf	graf	graf	graf	graf	
3	back-channel	repet-refr	complem	correccion	ref-ant ó lista-ref-ant	apt-obj	apt-zona	apt-tragect	despliega	apt-objs-coord	mueve-obj	agrega-obj	quita-obj	plan-graf	visual	rotar-vista		
4	utt1 : s: quieres que desplace o traiga algún objeto a la cocina ?																	
5	utt2 : u: <ruido> no																	
6	utt3 : puedes mover la estufa hacia la izquierda ?																	
7	utt4 : s: <ruido> hacia dónde ?																	
8	utt5 : u: <ruido> hacia <sil> hacia la derecha																	
9	utt6 : s: <no-vocal> hacia la derecha																	
10	utt7 : <no-vocal> okey																	
11	utt8 : que tanto quieres que la desplace ?																	
12	utt9 : u: a la mitad del espacio que hay entre la ventana y la pared																	
13	utt10 : s: okey																	
14	utt11 : quieres que mueva este objeto <sil> hacia acá ?																	
15	utt12 : u: no																	
16	utt13 : hacia el otro lado																	
17	utt14 : s: okey																	
18	utt15 : este objeto hacia acá ?																	
19	utt16 : u: sí																	
20	utt17 : s: okey																	
21	utt18 : <ruido>																	
22	utt19 : ahí está bien ? <no-vocal>																	
23	utt20 : u: un poco menos por favor																	
24	utt21 : s: okey																	
25	utt22 : <ruido>																	
26	utt23 : ahí está bien ?																	
27	utt24 : u: ahí está bien																	
28	utt25 : después <sil> me puedes poner <sil> [e] el extractor de aire encima de la <sil> de la estufa																	
29	utt26 : s: okey																	
30	utt27 : <ruido> me puedes decir qué																	

Ready

start | etiquetasd12 | formatos_al... | Recycle Bin | Windows Me... | cap_teoría... | Microsoft Ex... | untitled - Paint | EN | 11:26 PM

Figura II.5 Formato en hoja de cálculo Excel para etiquetaje *DIME-DAMSL Detallado* (parte 3 de 3).

Se asume que los etiquetadores clasifican instancias en una serie de categorías nominales predefinidas. Cada etiquetador asigna cada una de las instancias en una categoría que él considera adecuada basándose en los criterios predefinidos. En general, Kappa mide la consistencia entre los etiquetajes de los diversos anotadores e indirectamente refleja el acuerdo en la aplicación e interpretación de esos criterios al analizar casos particulares. Un etiquetador puede ser asumido como un individuo, como un equipo de individuos o como una herramienta de software que genere etiquetaciones automáticamente.

En el contexto de la etiquetación de actos de diálogo, Kappa mide qué tanto coinciden los etiquetadores anotando la misma etiqueta (o serie de etiquetas) a un mismo enunciado al desarrollar la tarea independientemente. Este estadístico responde a la siguiente pregunta: ¿Qué tanto se parecen los etiquetajes realizados por los anotadores independientes? E indirectamente se responde también a otras: ¿Están los etiquetadores interpretando en la misma forma los criterios para anotar un mismo enunciado? ¿Cuánta consistencia tiene el modelo expresado mediante los criterios de etiquetaje?

El número de coincidencias entre los etiquetajes de anotadores independientes es inversamente proporcional al producto del número de etiquetas posibles por el número de instancias a etiquetar; es decir, cuanto menos etiquetas o instancias a etiquetar se tengan, más probables son las coincidencias. Por ello, Kappa descuenta el acuerdo aleatorio al acuerdo *bruto* usando la siguiente expresión:

$$K = \frac{P(A) - P(E)}{1 - P(E)}$$

donde K (mayúscula) es Kappa, $P(A)$ es la proporción de veces que los k (minúscula) anotadores coinciden (pudiendo ser entre 0.0 y 1.0), y $P(E)$ es la proporción de veces que se esperaría que los anotadores coincidieran por azar (también entre 0.0 y 1.0). Cuando hay un completo acuerdo entre los anotadores, $K=1$; de otro modo, $K<1$. Además, K podría ser negativa si $P(A) < P(E)$; es decir, si los etiquetadores tuvieran tan pocas coincidencias que los etiquetajes resultarían peores que si se hicieran aleatoriamente.

Instancia	Categoría						
	1	2	...	j	...	m	
1	n_{11}	n_{12}	...	n_{1j}	...	n_{1m}	S_1
2	n_{21}	n_{22}	...	n_{2j}	...	n_{2m}	S_2
.							.
.							.
i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{im}	S_i
.							.
.							.
N	n_{N1}	n_{N2}	...	n_{Nj}	...	n_{Nm}	S_N
	C_1	C_2	...	C_j	...	C_m	

Tabla II.3. Ejemplo de matriz para computar Kappa.

Para calcular K , los conteos de datos de etiquetaje se colocan en una matriz de tamaño $N \times m$, donde N es el número total de instancias a clasificar y m es el número de categorías predefinidas. En la matriz, n_{ij} es el número de anotadores que asignan una instancia i a una categoría j . La tabla II.3 presenta un ejemplo de matriz para computar K .

La medida de acuerdo bruto, $P(A)$, se calcula con la siguiente expresión:

$$P(A) = \frac{1}{N} \sum_{i=1}^N S_i = \left[\frac{1}{Nk(k-1)} \sum_{i=1}^N \sum_{j=1}^m n_{ij}^2 \right] - \frac{1}{k-1}$$

El término S_i se calcula con:

$$S_i = \frac{\sum_{j=1}^m \binom{n_{ij}}{2}}{\binom{k}{2}} = \frac{1}{k(k-1)} \sum_{j=1}^m n_{ij}(n_{ij} - 1)$$

La siguiente expresión describe el cómputo de la medida de acuerdo esperado, $P(E)$:

$$P(E) = \sum_{j=1}^m p_j^2$$

A continuación se presentan las fórmulas para calcular los términos p_j y C_j :

$$p_j = C_j / Nk$$

$$C_j = \sum_{i=1}^N n_{ij}$$

Existen diversos criterios para determinar la confiabilidad del etiquetaje expresada mediante Kappa; uno de los más conocidos es presentado por Carletta (1996), refiriendo a Krippendorff (1980), quien propone la siguiente escala:

Si $K > 0.8$, la consistencia del etiquetaje es confiable.

Si $0.67 < K < 0.8$, la consistencia ofrece una confiabilidad que sólo permite generar conclusiones tentativas.

Si $K < 0.7$, la consistencia del etiquetaje no es confiable.

La determinación mediante Kappa de la confiabilidad de los etiquetajes permite saber si éstos pueden usarse o no en análisis o modelaciones. Dado que uno de los principales objetivos del presente trabajo es el análisis y modelación de los actos de diálogo, esta evaluación cobra especial importancia.

Apéndice III. Análisis estadístico de la modalidad del enunciado

Diálogo	Modalidad						Subtotal
	dec	int	int_sn	int_pron	imp	otra	
d01	86	3	21	3	0	3	116
d03	110	4	29	4	3	18	168
d12	70	0	36	5	0	6	117
d13	112	4	49	4	14	8	191
d14	92	1	24	10	0	10	137
d15	59	4	19	1	0	7	90
d17	168	1	44	3	1	20	237
d19	79	2	17	6	0	1	105
d21	50	6	13	0	0	0	69
d22	132	0	42	2	1	4	181
d23	57	3	16	3	0	2	81
d26	129	4	43	4	4	26	210
Total	1,144	32	353	45	23	105	1,702

Tabla III.1 Conteo de enunciados por tipo de modalidad por diálogo.

Diálogo	Modalidad					
	dec	int	int_sn	int_pron	imp	other
d01	74.1%	2.6%	18.1%	2.6%	0.0%	2.6%
d03	65.5%	2.4%	17.3%	2.4%	1.8%	10.7%
d12	59.8%	0.0%	30.8%	4.3%	0.0%	5.1%
d13	58.6%	2.1%	25.7%	2.1%	7.3%	4.2%
d14	67.2%	0.7%	17.5%	7.3%	0.0%	7.3%
d15	65.6%	4.4%	21.1%	1.1%	0.0%	7.8%
d17	70.9%	0.4%	18.6%	1.3%	0.4%	8.4%
d19	75.2%	1.9%	16.2%	5.7%	0.0%	1.0%
d21	72.5%	8.7%	18.8%	0.0%	0.0%	0.0%
d22	72.9%	0.0%	23.2%	1.1%	0.6%	2.2%
d23	70.4%	3.7%	19.8%	3.7%	0.0%	2.5%
d26	61.4%	1.9%	20.5%	1.9%	1.9%	12.4%
General	67.2%	1.9%	20.7%	2.6%	1.4%	6.2%

Tabla III.2 Porcentajes de enunciados por tipo de modalidad por diálogo.

Nota: la categoría *int* se refiere a las interrogativas que no fueron etiquetables como *int_sn* ni *int_pron*.

Apéndice IV. Análisis estadístico del rol del hablante

Diálogo	Rol		Total
	Sistema	Usuario	
d01	48	68	116
d03	72	96	168
d12	65	52	117
d13	94	97	191
d14	69	68	137
d15	41	49	90
d17	95	142	237
d19	52	53	105
d21	37	32	69
d22	97	84	181
d23	42	39	81
d26	82	128	210
Total	794	908	1,702

Tabla IV.1 Conteos de enunciados por rol de hablante por diálogo.

Diálogo	Rol	
	Sistema	Usuario
d01	41.4%	58.6%
d03	42.9%	57.1%
d12	55.6%	44.4%
d13	49.2%	50.8%
d14	50.4%	49.6%
d15	45.6%	54.4%
d17	40.1%	59.9%
d19	49.5%	50.5%
d21	53.6%	46.4%
d22	53.6%	46.4%
d23	51.9%	48.1%
d26	39.0%	61.0%
Global	46.7%	53.3%

Tabla IV.2 Porcentajes de enunciados por rol de hablante por diálogo.

Apéndice V. Análisis estadístico de etiquetaciones INTSINT

Diálogo	Núm. Total de Enunciados por Diálogo	Núm. de enunciados con INTSINT	Disponibilidad de etiquetación INTSINT (%)
d01	116	55	47.4
d03	168	134	79.8
d12	117	94	80.3
d13	191	103	53.9
d14	137	50	36.5
d15	90	41	45.6
d17	237	176	74.3
d19	105	38	36.2
d21	69	37	53.6
d22	181	109	60.2
d23	81	64	79.0
d26	210	142	67.6
Total	1,702	1,043	59.5%

Tabla V.1 Producción de etiquetaciones INTSINT en 12 diálogos del Corpus DIME.

Longitud de la secuencia INTSINT				
Límite inferior	Límite superior	Frecuencia	%	% Acum.
2	7	728	69.8	69.8
8	13	206	19.8	89.5
14	19	65	6.2	95.8
20	25	25	2.4	98.2
26	31	11	1.1	99.2
32	37	6	0.6	99.8
38	45	2	0.2	100.0
Total		1,043	100.0%	

Máximo	Mínimo	Promedio	Desv. Estándar	Rango
45	2	7.1	5.5	43

Tabla V.2. Análisis de longitud de secuencia INTSINT: número de etiquetas por enunciado

Apéndice VI. Análisis estadístico de las etiquetaciones de DIME-DAMSL preliminar

	d01	d03	d12	d13	d14	d15	d17	d19	d21	d22	d23	d26	GLOBAL
sol-inf	24	31	33	46	34	24	59	24	19	53	20	55	422
dir-accion	11	11	11	24	11	10	38	6	7	19	13	23	184
compromiso	11	10	11	24	11	10	27	6	7	18	12	18	165
oferta	2	2	1	1	1	1	0	1	2	1	1	1	14
resp	23	29	38	44	32	21	51	23	18	51	20	43	393
opcion-abierta	2	2	9	0	10	2	6	7	3	10	2	6	59
afirm	19	26	12	14	22	8	13	19	11	13	6	21	184
reafirm	0	1	1	2	3	3	0	1	4	1	2	7	25
acepta	37	52	50	80	64	33	81	38	30	85	34	72	656
acepta-parte	1	1	2	6	1	1	7	0	0	3	2	1	25
rechaza	3	2	5	5	1	1	7	5	0	2	1	4	36
rechaza-parte	1	0	0	1	0	0	0	0	0	0	0	2	4
espera	5	13	10	8	13	7	6	5	5	10	3	16	101
quiza	1	0	1	0	0	0	0	0	0	0	0	0	2
ack	4	8	2	0	1	2	10	3	3	5	4	4	46
sne	2	0	0	3	0	2	0	1	3	0	2	4	17
back-channel	0	0	0	0	0	0	0	0	1	1	0	0	2
repet-refr	6	13	8	8	11	6	2	5	4	7	3	10	83
complementacion	0	0	0	0	0	0	0	0	0	0	0	0	0
correccion	0	0	0	0	0	0	1	0	0	0	0	1	2
conv-abre	0	0	1	1	1	0	1	1	1	1	0	1	8
conv-cierra	2	1	2	1	1	1	3	2	1	1	1	2	18
perform	1	4	0	1	1	2	1	3	1	1	1	0	16
abandonada	4	3	0	0	2	4	11	1	0	0	2	11	38
mono	1	15	0	3	2	1	1	0	0	0	0	15	38
accion-graf	12	6	11	20	10	10	32	9	6	23	11	23	173
Total	172	230	208	292	232	149	357	160	126	305	140	340	2,711

Tabla VI.1 Conteo de etiquetas de acto de diálogo por diálogo y general.

Doce Diálogos: Análisis de Etiquetas de Actos de Diálogo

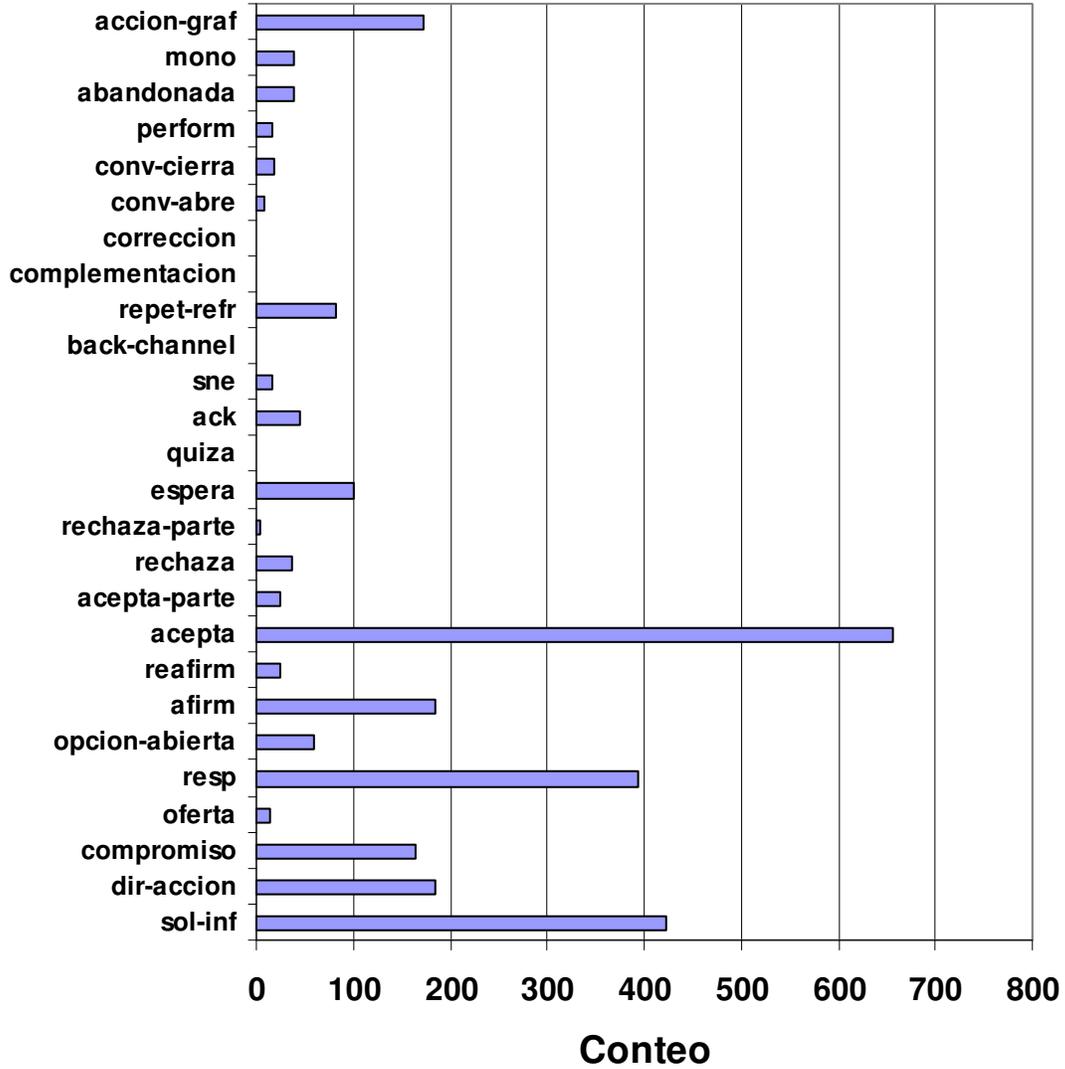


Figura VI.1 Conteo de etiquetas de acto de diálogo por diálogo y general.

	d01	d03	d12	d13	d14	d15	d17	d19	d21	d22	d23	d26	Promedio
sol-inf	14.0	13.5	15.9	15.8	14.7	16.1	16.5	15.0	15.1	17.4	14.3	16.2	15.4
dir-accion	6.4	4.8	5.3	8.2	4.7	6.7	10.6	3.8	5.6	6.2	9.3	6.8	6.5
compromiso	6.4	4.3	5.3	8.2	4.7	6.7	7.6	3.8	5.6	5.9	8.6	5.3	6.0
oferta	1.2	0.9	0.5	0.3	0.4	0.7	0.0	0.6	1.6	0.3	0.7	0.3	0.6
resp	13.4	12.6	18.3	15.1	13.8	14.1	14.3	14.4	14.3	16.7	14.3	12.6	14.5
opcion-abierta	1.2	0.9	4.3	0.0	4.3	1.3	1.7	4.4	2.4	3.3	1.4	1.8	2.2
afirm	11.0	11.3	5.8	4.8	9.5	5.4	3.6	11.9	8.7	4.3	4.3	6.2	7.2
reafirm	0.0	0.4	0.5	0.7	1.3	2.0	0.0	0.6	3.2	0.3	1.4	2.1	1.0
acepta	21.5	22.6	24.0	27.4	27.6	22.1	22.7	23.8	23.8	27.9	24.3	21.2	24.1
acepta-parte	0.6	0.4	1.0	2.1	0.4	0.7	2.0	0.0	0.0	1.0	1.4	0.3	0.8
rechaza	1.7	0.9	2.4	1.7	0.4	0.7	2.0	3.1	0.0	0.7	0.7	1.2	1.3
rechaza-parte	0.6	0.0	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.6	0.1
espera	2.9	5.7	4.8	2.7	5.6	4.7	1.7	3.1	4.0	3.3	2.1	4.7	3.8
quizá	0.6	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
ack	2.3	3.5	1.0	0.0	0.4	1.3	2.8	1.9	2.4	1.6	2.9	1.2	1.8
sne	1.2	0.0	0.0	1.0	0.0	1.3	0.0	0.6	2.4	0.0	1.4	1.2	0.8
back-channel	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	0.3	0.0	0.0	0.1
repet-refr	3.5	5.7	3.8	2.7	4.7	4.0	0.6	3.1	3.2	2.3	2.1	2.9	3.2
complementacion	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
correccion	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.0	0.3	0.0
conv-abre	0.0	0.0	0.5	0.3	0.4	0.0	0.3	0.6	0.8	0.3	0.0	0.3	0.3
conv-cierra	1.2	0.4	1.0	0.3	0.4	0.7	0.8	1.3	0.8	0.3	0.7	0.6	0.7
perform	0.6	1.7	0.0	0.3	0.4	1.3	0.3	1.9	0.8	0.3	0.7	0.0	0.7
abandonada	2.3	1.3	0.0	0.0	0.9	2.7	3.1	0.6	0.0	0.0	1.4	3.2	1.3
mono	0.6	6.5	0.0	1.0	0.9	0.7	0.3	0.0	0.0	0.0	0.0	4.4	1.2
accion-graf	7.0	2.6	5.3	6.8	4.3	6.7	9.0	5.6	4.8	7.5	7.9	6.8	6.2

Tabla VI.2 Porcentajes detallados y promedios de etiquetas de acto de diálogo.

Porcentajes de Etiquetas DIME-DAMSL en Doce Diálogos

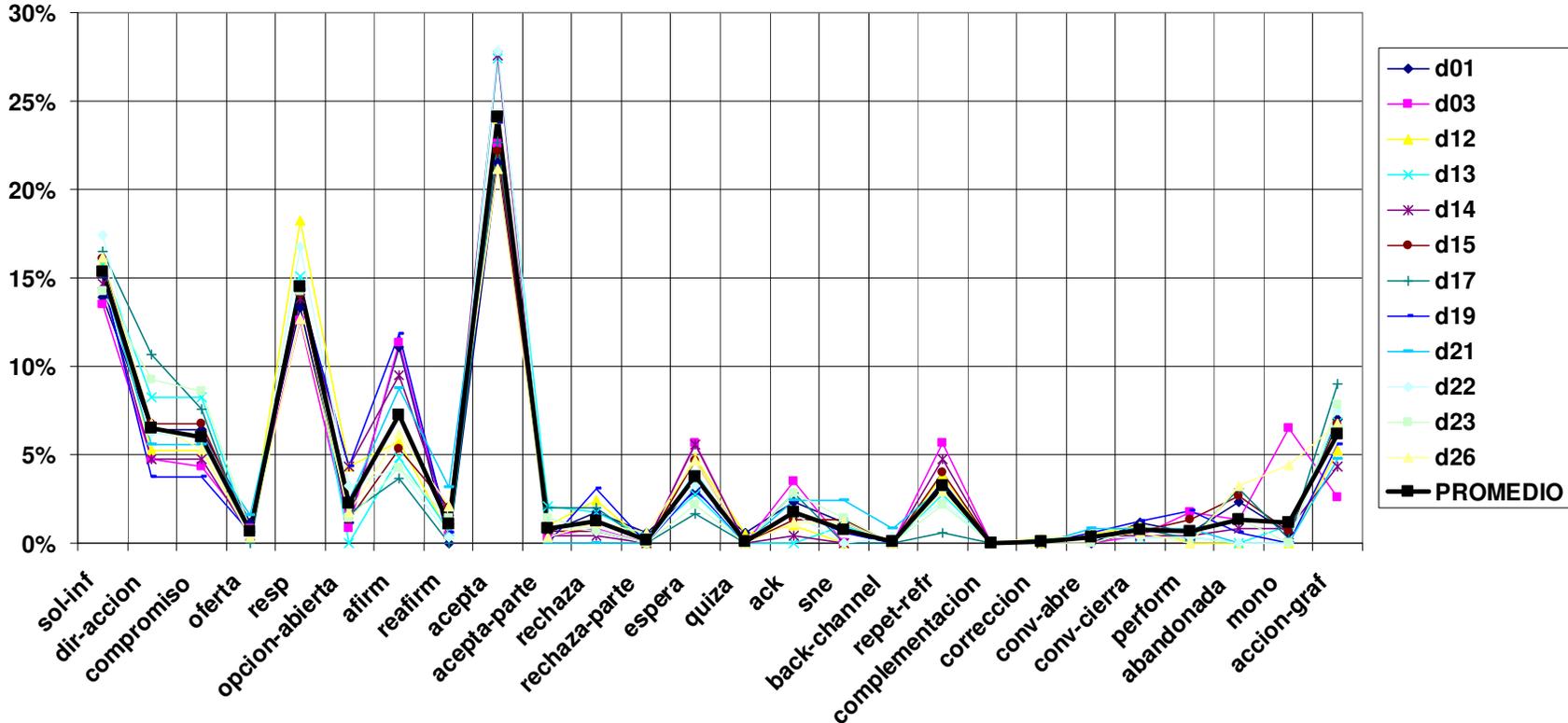


Figura VI.2 Porcentajes de etiquetas DIME-DAMSL en doce diálogos.

Etiqueta de Acto de Diálogo	Conteo	%	% Acum
acepta	656	24.2	24.2
sol-inf	422	15.6	39.8
resp	393	14.5	54.3
dir-accion	184	6.8	61.0
afirm	184	6.8	67.8
accion-graf	173	6.4	74.2
compromiso	165	6.1	80.3
espera	101	3.7	84.0
repet-refr	83	3.1	87.1
opcion-abierta	59	2.2	89.3
ack	46	1.7	91.0
abandonada	38	1.4	92.4
mono	38	1.4	93.8
rechaza	36	1.3	95.1
Otras etiquetas	133	4.9	100.0
Total	2,711	100.0%	

Tabla VI.3 Conteos y porcentajes globales de etiquetas de actos de diálogo en doce diálogos.

Nota: La categoría *otras etiquetas* incluye: reafirm, acepta-parte, conv-cierra, S.N.E. (Señal de No Entendimiento), perform, otra, conv-abre, rechaza-parte, quizá, back-channel, corrección, y complementación. No confundir con *otra*, que es una etiqueta por sí misma.

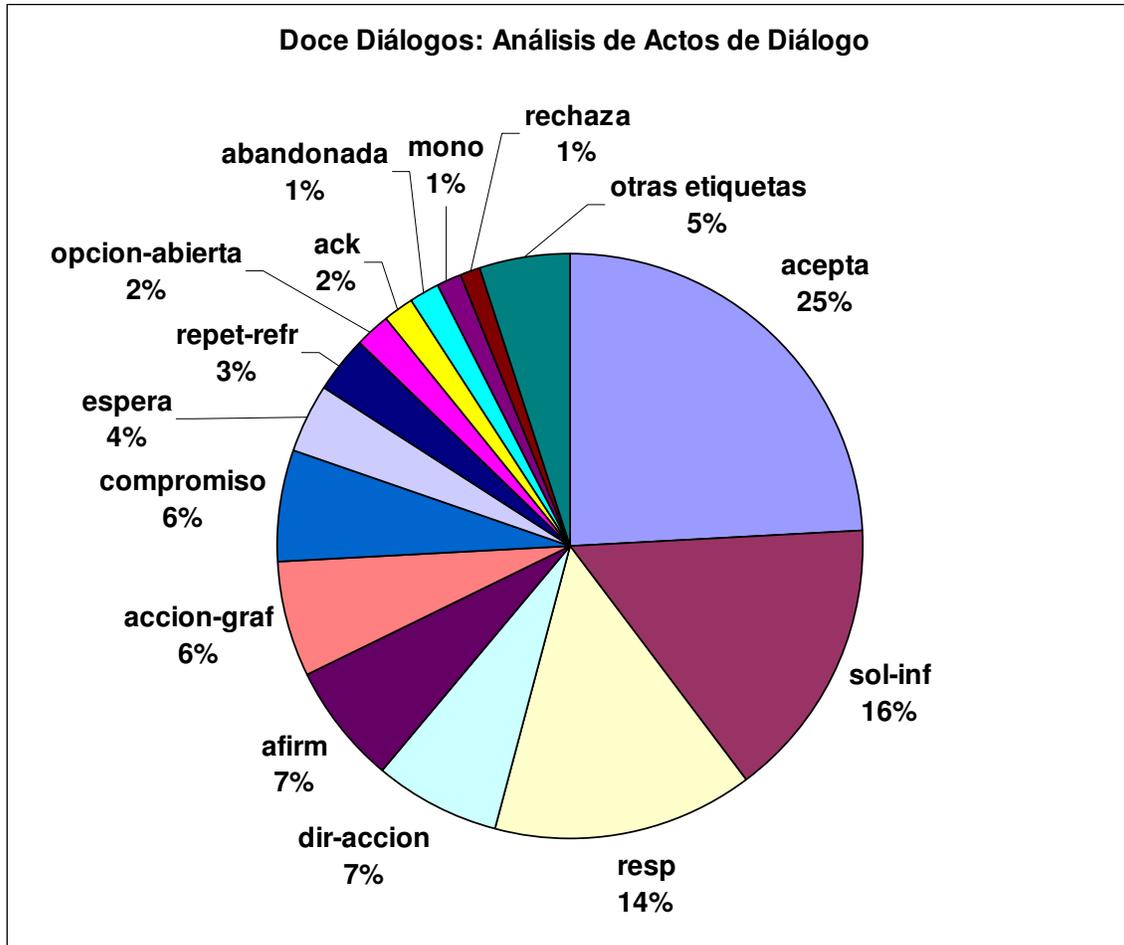


Figura VI.3 Porcentajes de actos de diálogo en doce diálogos.

Nota: La categoría *otras etiquetas* incluye: reafirm, acepta-parte, conv-cierra, S.N.E. (Señal de No Entendimiento), perform, otra, conv-abre, rechaza-parte, quizá, back-channel, corrección, y complementación. No confundir con *otra*, que es una etiqueta por sí misma.

Etiqueta de acto de diálogo	Conteo	%	% Acum
acepta	656	24.2	24.2
sol-inf	422	15.6	39.8
resp	393	14.5	54.3
dir-accion	184	6.8	61.0
afirm	184	6.8	67.8
accion-graf	173	6.4	74.2
compromiso	165	6.1	80.3
espera	101	3.7	84.0
repet-refr	83	3.1	87.1
opcion-abierta	59	2.2	89.3
ack	46	1.7	91.0
abandonada	38	1.4	92.4
mono	38	1.4	93.8
rechaza	36	1.3	95.1
reafirm	25	0.9	96.0
acepta-parte	25	0.9	96.9
conv-cierra	18	0.7	97.6
sne	17	0.6	98.2
perform	16	0.6	98.8
oferta	14	0.5	99.3
conv-abre	8	0.3	99.6
rechaza-parte	4	0.1	99.8
quiza	2	0.1	99.9
back-channel	2	0.1	99.9
correccion	2	0.1	100.0
complementacion	0	0.0	100.0
Total	2,711	100.0%	

Tabla VI.4 Pareto de etiquetas de actos de diálogo.

El análisis de Pareto muestra que la mayoría de los actos de diálogo (80.3%) contienen solamente siete etiquetas DIME-DAMSL: *acepta*, *sol-info*, *respuesta*, *dir-accion*, *afirm*, *accion-graf* y *compromiso*.

Apéndice VII. Análisis de correlación entre actos de diálogo y modalidades de enunciado

Actos de diálogo de obligaciones	Modalidad del enunciado				Totales
	int	dec	imp	otra	
sol-info	228	33	6	6	273
resp	4	214	1	6	225
dir-acción	21	88	5	6	120
compromiso	0	112	0	0	112
dir-acción_resp	3	13	0	0	16
sol-info_resp	12	3	1	0	16
oferta	0	5	0	0	5
sol-info_oferta	0	1	0	0	1
dir-acción_oferta	0	1	0	0	1
<i>Sin etiqueta de obligs.</i>	23	222	1	28	274
Totales	291	692	14	46	1,043

Tabla VII.1 Número de enunciados por modalidad por acto de diálogo de obligaciones.

Actos de diálogo de obligaciones	Modalidad del enunciado			
	int	dec	imp	otra
sol-info	83.5	12.1	2.2	2.2
resp	1.8	95.1	0.4	2.7
dir-acción	17.5	73.3	4.2	5.0
compromiso	0.0	100.0	0.0	0.0
dir-acción_resp	18.8	81.3	0.0	0.0
sol-info_resp	75.0	18.8	6.3	0.0
oferta	0.0	100.0	0.0	0.0
sol-info_oferta	0.0	100.0	0.0	0.0
dir-acción_oferta	0.0	100.0	0.0	0.0
<i>Sin etiqueta de obligs.</i>	8.4	81.0	0.4	10.2

Tabla VII.2 Porcentajes de enunciados por modalidad por acto de diálogo de obligaciones.

Actos de diálogo del acuerdo	Modalidad del enunciado				Totales
	int	dec	imp	otra	
acepta_espera (*)	2	0	0	0	2
acepta	29	344	3	10	386
espera	58	5	2	4	69
afirm	2	75	1	0	78
afirm_acepta- parte_exclama	0	1	0	0	1
afirm_quizá	0	2	0	0	2
afirm_acepta	0	9	0	0	9
opción_abierta	3	44	0	0	47

oferta_acepta	2	0	0	0	2
afirm_acepta_exclama	0	1	0	0	1
conv-cierra	0	5	0	0	5
perform	0	4	0	0	4
acepta-parte	3	17	0	0	20
reafirm	3	14	0	0	17
afirm_rechaza	0	4	0	0	4
oferta_conv-abre	5	0	0	0	5
rechaza	1	10	0	0	11
quizá	0	1	0	0	1
reafirm_espera	1	0	0	0	1
opción-abierta_acepta	0	4	0	0	4
rechaza-parte	0	4	0	0	4
oferta	2	0	0	0	2
otra	0	1	0	0	1
conv-abre	0	1	0	0	1
afirm_espera	0	2	0	0	2
opción-abierta_rechaza	0	1	0	0	1
afirm_perform_conv-cierra	0	1	0	0	1
afirm_conv-cierra	0	1	0	0	1
<i>Sin etiqueta de acuerdo</i>	180	141	8	32	361
Totales	291	692	14	46	1,043

Tabla VII.3 Número de enunciados por modalidad por acto de diálogo del acuerdo. (*) Error de anotación.

Actos de diálogo del acuerdo	Modalidad del enunciado (%)				Totales %
	int	dec	imp	otra	
acepta_espera (*)	100.0	0.0	0.0	0.0	100.0
acepta	7.5	89.1	0.8	2.6	100.0
espera	84.1	7.2	2.9	5.8	100.0
afirm	2.6	96.2	1.3	0.0	100.0
afirm_acepta-parte_exclama	0.0	100.0	0.0	0.0	100.0
afirm_quizá	0.0	100.0	0.0	0.0	100.0
afirm_acepta	0.0	100.0	0.0	0.0	100.0
opción_abierta	6.4	93.6	0.0	0.0	100.0
oferta_acepta	100.0	0.0	0.0	0.0	100.0
afirm_acepta_exclama	0.0	100.0	0.0	0.0	100.0
conv-cierra	0.0	100.0	0.0	0.0	100.0
perform	0.0	100.0	0.0	0.0	100.0
acepta-parte	15.0	85.0	0.0	0.0	100.0
reafirm	17.6	82.4	0.0	0.0	100.0
afirm_rechaza	0.0	100.0	0.0	0.0	100.0
oferta_conv-abre	100.0	0.0	0.0	0.0	100.0
rechaza	9.1	90.9	0.0	0.0	100.0
quizá	0.0	100.0	0.0	0.0	100.0
reafirm_espera	100.0	0.0	0.0	0.0	100.0
opción-abierta_acepta	0.0	100.0	0.0	0.0	100.0
rechaza-parte	0.0	100.0	0.0	0.0	100.0

oferta	100.0	0.0	0.0	0.0	100.0
otra	0.0	100.0	0.0	0.0	100.0
conv-abre	0.0	100.0	0.0	0.0	100.0
afirm_espera	0.0	100.0	0.0	0.0	100.0
opción-abierta_rechaza	0.0	100.0	0.0	0.0	100.0
afirm_perform_conv-cierra	0.0	100.0	0.0	0.0	100.0
afirm_conv-cierra	0.0	100.0	0.0	0.0	100.0
<i>Sin etiqueta de acuerdo</i>	49.9	39.1	2.2	8.9	100.0
Totales	27.9	66.3	1.3	4.4	100.0

Tabla VII.4 Porcentajes de enunciados por modalidad por acto de diálogo de acuerdo. (*) Error de anotación.

Actos de diálogo de obligaciones	Modalidad Interrogativa		
	Frec.	%	% Acum.
sol-info	228	78.4	78.4
<i>Sin etiqueta de obligaciones</i>	23	7.9	86.3
dir-accion	21	7.2	93.5
sol-info_resp	12	4.1	97.6
resp	4	1.4	99.0
dir-accion_resp	3	1.0	100.0
compromiso	0	0.0	100.0
oferta	0	0.0	100.0
sol-info_oferta	0	0.0	100.0
dir-accion_oferta	0	0.0	100.0
Total	291		

Tabla VII.5 Actos de diálogo de obligaciones con modalidad interrogativa.

Actos de diálogo de Obligaciones	Modalidad Declarativa		
	Nr.	%	% Acum.
<i>Sin etiqueta de obligaciones</i>	222	32.1	32.1
resp	214	30.9	63.0
compromiso	112	16.2	79.2
dir-accion	88	12.7	91.9
sol-info	33	4.8	96.7
dir-accion_resp	13	1.9	98.6
oferta	5	0.7	99.3
sol-info_resp	3	0.4	99.7
sol-info_oferta	1	0.1	99.9
dir-accion_oferta	1	0.1	100.0
Total	692		

Tabla VII.6. Actos de diálogo de obligaciones con modalidad declarativa.

Actos de diálogo de obligaciones	Modalidad Imperativa		
	Frec.	%	% Acum.
sol-info	6	42.9	42.9
dir-accion	5	35.7	78.6
resp	1	7.1	85.7
sol-info_resp	1	7.1	92.9
<i>Sin etiqueta en obligaciones</i>	1	7.1	100.0
compromiso	0	0.0	100.0
dir-accion_resp	0	0.0	100.0
oferta	0	0.0	100.0
sol-info_oferta	0	0.0	100.0
dir-accion_oferta	0	0.0	100.0
Total	14		

Tabla VII.7 Actos de diálogo de obligaciones con modalidad imperativa.

Actos de diálogo de obligaciones	Modalidad <i>otra</i>		
	Frec.	%	% Acum.
<i>Sin etiqueta de obligaciones</i>	28	60.9	60.9
sol-info	6	13.0	73.9
resp	6	13.0	87.0
dir-accion	6	13.0	100.0
compromiso	0	0.0	100.0
dir-accion_resp	0	0.0	100.0
sol-info_resp	0	0.0	100.0
oferta	0	0.0	100.0
sol-info_oferta	0	0.0	100.0
dir-accion_oferta	0	0.0	100.0
Total	46		

Tabla VII.8 Actos de diálogo de obligaciones con modalidad *Otra*.

Actos de diálogo del acuerdo	Modalidad Interrogativa		
	Frec.	%	% Acum.
<i>Sin etiqueta de acuerdo</i>	180	61.9	61.9
espera	58	19.9	81.8
acepta	29	10.0	91.8
oferta_conv-abre	5	1.7	93.5
opción-abierta	3	1.0	94.5
acepta-parte	3	1.0	95.5
reafirm	3	1.0	96.6
acepta, espera (*)	2	0.7	97.3
afirm	2	0.7	97.9

oferta_acepta	2	0.7	98.6
oferta	2	0.7	99.3
rechaza	1	0.3	99.7
reafirm_espera	1	0.3	100.0
afirm_acepta- parte_exclama	0	0.0	100.0
afirm_quiza	0	0.0	100.0
afirm_acepta	0	0.0	100.0
afirm_acepta_ exclama	0	0.0	100.0
conv-cierra	0	0.0	100.0
perform	0	0.0	100.0
afirm_rechaza	0	0.0	100.0
quiza	0	0.0	100.0
opción- abierta_acepta	0	0.0	100.0
rechaza-parte	0	0.0	100.0
otra	0	0.0	100.0
conv-abre	0	0.0	100.0
afirm_espera	0	0.0	100.0
opción-abierta_ rechaza	0	0.0	100.0
afirm_perform_conv- cierra	0	0.0	100.0
afirm_conv-cierra	0	0.0	100.0
Total	291		

Tabla VII.9 Actos de diálogo de acuerdo con modalidad interrogativa. (*) Error de anotación.

Actos de diálogo del <i>acuerdo</i>	Modalidad Declarativa		
	Frec.	%	% Acum.
acepta	344	49.7	49.7
<i>Sin etiqueta de acuerdo</i>	141	20.4	70.1
afirm	75	10.8	80.9
opción-abierta	44	6.4	87.3
acepta-parte	17	2.5	89.7
reafirm	14	2.0	91.8
rechaza	10	1.4	93.2
afirm_acepta	9	1.3	94.5
espera	5	0.7	95.2
conv-cierra	5	0.7	96.0
perform	4	0.6	96.5
afirm_rechaza	4	0.6	97.1
opción- abierta_acepta	4	0.6	97.7
rechaza-parte	4	0.6	98.3
afirm_quiza	2	0.3	98.6
afirm_espera	2	0.3	98.8
afirm_acepta- parte_exclama	1	0.1	99.0
afirm, acepta_exclama	1	0.1	99.1
quiza	1	0.1	99.3

otra	1	0.1	99.4
conv-abre	1	0.1	99.6
opción-abierta_rechaza	1	0.1	99.7
afirm_perform_conv-cierra	1	0.1	99.9
afirm_conv-cierra	1	0.1	100.0
acepta, espera	0	0.0	100.0
oferta_acepta	0	0.0	100.0
oferta_conv-abre	0	0.0	100.0
reafirm_espera	0	0.0	100.0
oferta	0	0.0	100.0
Total	692		

Tabla VII.10 Actos de diálogo de acuerdo con modalidad declarativa.

Actos de diálogo del acuerdo	Modalidad Imperativa		
	Frec.	%	% Acum.
<i>Sin etiqueta de acuerdo</i>	8	57.1	57.1
acepta	3	21.4	78.6
espera	2	14.3	92.9
afirm	1	7.1	100.0
acepta, espera	0	0.0	100.0
afirm_acepta-parte_exclama	0	0.0	100.0
afirm_quiza	0	0.0	100.0
afirm_acepta	0	0.0	100.0
opción-abierta	0	0.0	100.0
oferta_acepta	0	0.0	100.0
afirm_acepta_exclama	0	0.0	100.0
conv-cierra	0	0.0	100.0
perform	0	0.0	100.0
acepta-parte	0	0.0	100.0
reafirm	0	0.0	100.0
afirm_rechaza	0	0.0	100.0
oferta_conv-abre	0	0.0	100.0
rechaza	0	0.0	100.0
quiza	0	0.0	100.0
reafirm_espera	0	0.0	100.0
opción-abierta_acepta	0	0.0	100.0
rechaza-parte	0	0.0	100.0
oferta	0	0.0	100.0
otra	0	0.0	100.0
conv-abre	0	0.0	100.0
afirm_espera	0	0.0	100.0
opción-abierta_rechaza	0	0.0	100.0
afirm_perform_conv-cierra	0	0.0	100.0
afirm_conv-cierra	0	0.0	100.0
Total	14		

Tabla VII.11 Actos de diálogo de acuerdo con modalidad imperativa.

Actos de diálogo del acuerdo	Modalidad <i>otra</i>		
	Frec.	%	% Acum.
<i>Sin etiqueta de acuerdo</i>	32	69.6	70
acepta	10	21.7	91
espera	4	8.7	100
acepta, espera	0	0.0	100
afirm	0	0.0	100
afirm_acepta- parte_exclama	0	0.0	100
afirm_quiza	0	0.0	100
afirm_acepta	0	0.0	100
opción-abierta	0	0.0	100
oferta_acepta	0	0.0	100
afirm_acepta_exclama	0	0.0	100
conv-cierra	0	0.0	100
perform	0	0.0	100
acepta-parte	0	0.0	100
reafirm	0	0.0	100
afirm_rechaza	0	0.0	100
oferta_conv-abre	0	0.0	100
rechaza	0	0.0	100
quiza	0	0.0	100
reafirm_espera	0	0.0	100
opción-abierta_acepta	0	0.0	100
rechaza-parte	0	0.0	100
oferta	0	0.0	100
otra	0	0.0	100
conv-abre	0	0.0	100
afirm_espera	0	0.0	100
opción- abierta_rechaza	0	0.0	100
afirm_perform_conv- cierra	0	0.0	100
afirm_conv-cierra	0	0.0	100
Total	46		

Tabla VII.12 Actos de diálogo de acuerdo con modalidad *Otra*.

Apéndice VIII. Análisis de correlación entre rol de hablante y acto de diálogo

Obligaciones	Rol		Total
	Sistema	Usuario	
<i>Sin etiqueta de obligaciones</i>	148	126	274
sol-info	197	76	273
resp	60	165	225
dir-acción	0	120	120
compromiso	112	0	112
dir-acción_resp	0	16	16
sol-info_resp	10	6	16
oferta	0	5	5
dir-acción_oferta	0	1	1
sol-info_oferta	0	1	1
Total	527	516	1,043

Tabla VIII.1 Conteos de etiquetas de *obligaciones* por rol de hablante.

Obligaciones	Rol		Global
	Sistema	Usuario	
<i>Sin etiqueta de obligaciones</i>	54.0	46.0	26.3
sol-info	72.2	27.8	26.2
resp	26.7	73.3	21.6
dir-acción	0.0	100.0	11.5
compromiso	100.0	0.0	10.7
dir-acción_resp	0.0	100.0	1.5
sol-info_resp	62.5	37.5	1.5
oferta	0.0	100.0	0.5
dir-acción_oferta	0.0	100.0	0.1
sol-info_oferta	0.0	100.0	0.1

Tabla VIII.2 Porcentajes de etiquetas de *obligaciones* por rol de hablante.

Acuerdo	Rol		Total
	Sistema	Usuario	
acepta	205	181	386
<i>Sin etiqueta de acuerdo</i>	164	197	361
afirm	22	56	78
espera	59	10	69
opción-abierta	47	0	47
acepta-parte	0	20	20
reafirm	1	16	17
rechaza	1	10	11
afirm_acepta	2	7	9
conv-cierra	3	2	5
oferta_conv-abre	5	0	5
afirma_rechaza	1	3	4
opción_abierta_acepta	4	0	4
perform	2	2	4
rechaza-parte	0	4	4
acepta_espera	2	0	2
afirm_espera	0	2	2
afirm_quizá	0	2	2
oferta	2	0	2
oferta_acepta	2	0	2
afirm_acepta_exclama	0	1	1
afirm_acepta-parte_exclama	0	1	1
afirm_conv-cierra	1	0	1
afirm_perform_conv-cierra	1	0	1
conv-abre	0	1	1
opción-abierta_rechaza	1	0	1
otra	0	1	1
quizá	1	0	1
reafirm_espera	1	0	1
Total	527	516	1,043

Tabla VIII.3 Conteos de etiquetas del plano de *acuerdo* por rol de hablante.

Acuerdo	Rol		Global
	Sistema	Usuario	
acepta	53.1%	46.9%	37.0%
sin-etiqueta	45.4%	54.6%	34.6%
afirm	28.2%	71.8%	7.5%
espera	85.5%	14.5%	6.6%
opción-abierta	100.0%	0.0%	4.5%
acepta-parte	0.0%	100.0%	1.9%
reafirm	5.9%	94.1%	1.6%
rechaza	9.1%	90.9%	1.1%
afirm_acepta	22.2%	77.8%	0.9%
conv-cierra	60.0%	40.0%	0.5%
oferta_conv-abre	100.0%	0.0%	0.5%
afirma_rechaza	25.0%	75.0%	0.4%
opción_abierta_acepta	100.0%	0.0%	0.4%
perform	50.0%	50.0%	0.4%
rechaza-parte	0.0%	100.0%	0.4%
acepta_espera	100.0%	0.0%	0.2%
afirm_espera	0.0%	100.0%	0.2%
afirm_quizá	0.0%	100.0%	0.2%
oferta	100.0%	0.0%	0.2%
oferta_acepta	100.0%	0.0%	0.2%
afirm_acepta_exclama	0.0%	100.0%	0.1%
afirm_acepta-parte_exclama	0.0%	100.0%	0.1%
afirm_conv-cierra	100.0%	0.0%	0.1%
afirm_perform_conv-cierra	100.0%	0.0%	0.1%
conv-abre	0.0%	100.0%	0.1%
opción-abierta_rechaza	100.0%	0.0%	0.1%
otra	0.0%	100.0%	0.1%
quizá	100.0%	0.0%	0.1%
reafirm_espera	100.0%	0.0%	0.1%

Tabla VIII.4 Porcentajes de etiquetas del plano de *acuerdo* por rol de hablante.

Apéndice IX. Análisis de correlaciones entre actos de diálogo de dos planos.

Actos de diálogo del acuerdo	Actos de diálogo de obligaciones										Totales
	dir-acción	dir-acción_resp	dir-acción_oferta	resp	compro-miso	sol-info	sol-info_resp	sol-info_oferta	oferta	sin-etiqueta	
acepta	30	0	1	126	109	26	1	1	5	87	386
sin-etiqueta	83	2	0	2	3	179	10	0	0	82	361
afirm	2	0	0	36	0	0	1	0	0	39	78
espera	0	0	0	0	0	63	1	0	0	5	69
opción-abierta	0	0	0	32	0	0	0	0	0	15	47
acepta-parte	4	13	0	1	0	0	2	0	0	0	20
reafirm	0	0	0	11	0	1	0	0	0	5	17
rechaza	1	0	0	3	0	1	1	0	0	5	11
afirm_acepta	0	1	0	2	0	0	0	0	0	6	9
conv-cierra	0	0	0	0	0	0	0	0	0	5	5
oferta_conv-abre	0	0	0	0	0	0	0	0	0	5	5
afirm_rechaza	0	0	0	4	0	0	0	0	0	0	4
opción-abierta_acepta	0	0	0	2	0	0	0	0	0	2	4
perform	0	0	0	0	0	0	0	0	0	4	4
rechaza-parte	0	0	0	4	0	0	0	0	0	0	4
acepta_espera	0	0	0	0	0	2	0	0	0	0	2
afirma_espera	0	0	0	0	0	0	0	0	0	2	2
afirm_quizá	0	0	0	0	0	0	0	0	0	2	2
oferta	0	0	0	0	0	0	0	0	0	2	2
oferta_acepta	0	0	0	0	0	0	0	0	0	2	2
afirm_acepta_exclama	0	0	0	0	0	0	0	0	0	1	1
afirm_acepta-parte_exclama	0	0	0	1	0	0	0	0	0	0	1
afirm_conv-cierra	0	0	0	1	0	0	0	0	0	0	1
afirm_perform_conv-cierra	0	0	0	0	0	0	0	0	0	1	1
conv-abre	0	0	0	0	0	0	0	0	0	1	1
quizá	0	0	0	0	0	0	0	0	0	1	1
opción-abierta_rechaza	0	0	0	0	0	0	0	0	0	1	1
otra	0	0	0	0	0	0	0	0	0	1	1
reafirm_espera	0	0	0	0	0	1	0	0	0	0	1
TOTALES	120	16	1	225	112	273	16	1	5	274	1,043

Tabla IX.1 Conteo de combinaciones de actos de diálogo de acuerdo con actos de obligaciones.

Actos de diálogo del acuerdo	Actos de diálogo de obligaciones										Globales
	dir-acción	dir-acción_resp	dir-acción_oferta	resp	compromiso	sol-info	sol-info_resp	sol-info_oferta	oferta	sin-etiqueta	
acepta	7.8	0.0	0.3	32.6	28.2	6.7	0.3	0.3	1.3	22.5	37.0
sin-etiqueta	23.0	0.6	0.0	0.6	0.8	49.6	2.8	0.0	0.0	22.7	34.6
afirm	2.6	0.0	0.0	46.2	0.0	0.0	1.3	0.0	0.0	50.0	7.5
espera	0.0	0.0	0.0	0.0	0.0	91.3	1.4	0.0	0.0	7.2	6.6
opción-abierta	0.0	0.0	0.0	68.1	0.0	0.0	0.0	0.0	0.0	31.9	4.5
acepta-parte	20.0	65.0	0.0	5.0	0.0	0.0	10.0	0.0	0.0	0.0	1.9
reafirm	0.0	0.0	0.0	64.7	0.0	5.9	0.0	0.0	0.0	29.4	1.6
rechaza	9.1	0.0	0.0	27.3	0.0	9.1	9.1	0.0	0.0	45.5	1.1
afirm_acepta	0.0	11.1	0.0	22.2	0.0	0.0	0.0	0.0	0.0	66.7	0.9
conv-cierra	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.5
oferta_conv-abre	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.5
afirm_rechaza	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4
opción-abierta_acepta	0.0	0.0	0.0	50.0	0.0	0.0	0.0	0.0	0.0	50.0	0.4
perform	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.4
rechaza-parte	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4
acepta_espera	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.2
afirma_espera	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.2
afirm_quizá	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.2
oferta	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.2
oferta_acepta	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.2
afirm_acepta_exclama	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.1
afirm_acepta-parte_exclama	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
afirm_conv-cierra	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
afirm_perform_conv-cierra	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.1
conv-abre	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.1
quizá	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.1
opción-abierta_rechaza	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.1
otra	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.1
reafirm_espera	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.1
Globales	11.5	1.5	0.1	21.6	10.7	26.2	1.5	0.1	0.5	26.3	100.0%

Tabla IX.2 Porcentajes de combinaciones de actos de diálogo de *acuerdo* con actos de *obligaciones*.

Apéndice X. Reglas del modelo 2.1 para modalidad predicha optimizada (ordenadas por soporte)

Núm. regla	Regla	Casos donde Ocurre la Premisa	Casos No Satisfactorios	Confianza %	Soporte %
1	IF last_2 = BS, THEN dec (136.0/16.0)	136	16	88.2	11.5
2	IF last_2 = DB, THEN dec (107.0/17.0)	107	17	84.1	8.6
3	IF last_2 = MB, THEN dec (87.0/2.0)	87	2	97.7	8.1
4	IF last_2 = TB, THEN dec (90.0/17.0)	90	17	81.1	7.0
5	IF last_2 = TS AND speaker_role = s, THEN int (58.0/9.0)	58	9	84.5	4.7
6	IF last_2 = UT AND speaker_role = s AND utt_duration > 659.3125, THEN int (47.0)	47	0	100.0	4.5
7	IF last_2 = US AND speaker_role = u, THEN dec (34.0/6.0)	34	6	82.4	2.7
8	IF last_2 = UT AND speaker_role = u, THEN dec (45.0/18.0)	45	18	60.0	2.6
9	IF last_2 = TS AND speaker_role = u, THEN dec (36.0/11.0)	36	11	69.4	2.4
10	IF last_2 = LS, THEN dec (31.0/9.0)	31	9	71.0	2.1
11	IF last_2 = TL, THEN dec (35.0/14.0)	35	14	60.0	2.0
12	IF last_2 = SB, THEN dec (28.0/7.0)	28	7	75.0	2.0
13	IF last_2 = US AND speaker_role = s AND utt_duration <= 1178.8125, THEN int (22.0/1.0)	22	1	95.5	2.0
14	IF last_2 = BH, THEN dec (36.0/16.0)	36	16	55.6	1.9
15	IF last_2 = BT AND first_2 = MB, THEN dec (32.0/12.0)	32	12	62.5	1.9
16	IF last_2 = DS, THEN dec (20.0/2.0)	20	2	90.0	1.7
17	IF last_2 = LT AND speaker_role = s, THEN int (15.0/1.0)	15	1	93.3	1.3
18	IF last_2 = HB, THEN dec (14.0/2.0)	14	2	85.7	1.2
19	IF last_2 = BT AND first_2 = MS AND utt_duration <= 737.75, THEN dec (11.0)	11	0	100.0	1.1
20	IF last_2 = ST AND speaker_role = u, THEN dec (11.0/1.0)	11	1	90.9	1.0
21	IF last_2 = DL, THEN dec (10.0)	10	0	100.0	1.0
22	IF last_2 = HS AND speaker_role = u, THEN dec (11.0/2.0)	11	2	81.8	0.9
23	IF last_2 = BT AND first_2 = MH, THEN int (9.0)	9	0	100.0	0.9
24	IF last_2 = ST AND speaker_role = s, THEN int (9.0/1.0)	9	1	88.9	0.8
25	IF last_2 = BT AND first_2 = MS AND utt_duration > 737.75 AND utt_duration > 786.25, THEN dec (9.0/2.0)	9	2	77.8	0.7
26	IF last_2 = HL, THEN dec (10.0/4.0)	10	4	60.0	0.6

27	IF last_2 = LH, THEN dec (8.0/3.0)	8	3	62.5	0.5
28	IF last_2 = US AND speaker_role = s AND utt_duration > 1178.8125, THEN dec (8.0/3.0)	8	3	62.5	0.5
29	IF last_2 = BT AND first_2 = BT AND speaker_role = u, THEN dec (6.0/1.0)	6	1	83.3	0.5
30	IF last_2 = BT AND first_2 = MD AND speaker_role = s, THEN int (6.0/1.0)	6	1	83.3	0.5
31	IF last_2 = LT AND speaker_role = u, THEN dec (7.0/3.0)	7	3	57.1	0.4
32	IF last_2 = SH, THEN dec (7.0/3.0)	7	3	57.1	0.4
33	IF last_2 = SL, THEN dec (6.0/2.0)	6	2	66.7	0.4
34	IF last_2 = HS AND speaker_role = s, THEN int (6.0/2.0)	6	2	66.7	0.4
35	IF last_2 = BT AND first_2 = MD AND speaker_role = u, THEN dec (5.0/2.0)	5	2	60.0	0.3
36	IF last_2 = DD, THEN dec (4.0/1.0)	4	1	75.0	0.3
37	IF last_2 = UT AND speaker_role = s AND utt_duration <= 659.3125 AND utt_duration <= 520.5625, THEN int (4.0/1.0)	4	1	75.0	0.3
38	IF last_2 = UH, THEN dec (3.0/1.0)	3	1	66.7	0.2
39	IF last_2 = UT AND speaker_role = s AND utt_duration <= 659.3125 AND utt_duration > 520.5625, THEN dec (2.0)	2	0	100.0	0.2
40	IF last_2 = BT AND first_2 = BT AND speaker_role = s, THEN int (3.0/1.0)	3	1	66.7	0.2
41	IF last_2 = BT AND first_2 = MS AND utt_duration > 737.75 AND utt_duration <= 786.25, THEN int (2.0)	2	0	100.0	0.2
42	IF last_2 = BT AND first_2 = ML, THEN otra (3.0/1.0)	3	1	66.7	0.2
43	IF last_2 = UM, THEN dec (1.0)	1	0	100.0	0.1
44	IF last_2 = DU, THEN dec (1.0)	1	0	100.0	0.1
45	IF last_2 = SU, THEN dec (1.0)	1	0	100.0	0.1
46	IF last_2 = HD, THEN int (2.0/1.0)	2	1	50.0	0.1
47	IF last_2 = BT AND first_2 = MT, THEN int (1.0)	1	0	100.0	0.1
48	IF last_2 = BT AND first_2 = MU, THEN int (1.0)	1	0	100.0	0.1
49	IF last_2 = LU, THEN int (1.0)	1	0	100.0	0.1
50	IF last_2 = DM, THEN int (1.0)	1	0	100.0	0.1
51	IF last_2 = BU, THEN imp (1.0)	1	0	100.0	0.1

Apéndice XI. Reglas del modelo 3.1.A para actos de diálogo de *obligaciones* (ordenadas por soporte)

Núm. regla	Regla	Casos donde Ocurre la Premisa	Casos No Satisfactorios	Confianza %	Soporte %
1	IF commgr = sin-etiqueta AND commgr_minus1 = acepta AND number_of_tones <= 15, THEN sol-inf (91.0/52.0)	91	52	42.9	8.7
2	IF commgr = accion-graf AND obligations_minus1 = compromiso, THEN sol-inf_accion-graf (72.0/1.0)	72	1	98.6	6.9
3	IF commgr = acepta AND speaker_role = s AND obligations_minus1 = dir-accion, THEN compromiso (71.0/19.0)	71	19	73.2	6.8
4	IF commgr = espera_repet-refr, THEN sol-inf (54.0/1.0)	54	1	98.1	5.2
5	IF commgr = acepta AND speaker_role = u AND commgr_minus1 = accion-graf, THEN resp (51.0)	51	0	100.0	4.9
6	IF commgr = acepta AND speaker_role = u AND commgr_minus1 = espera_repet-refr, THEN resp (43.0/3.0)	43	3	93.0	4.1
7	IF commgr = acepta AND speaker_role = s AND obligations_minus1 = resp AND optimal_pred_mood = dec, THEN compromiso (43.0/7.0)	43	7	83.7	4.1
8	IF commgr = opcion-abierta_despliega, THEN resp (41.0/9.0)	41	9	78.0	3.9
9	IF commgr = acepta AND speaker_role = s AND obligations_minus1 = sol-inf, THEN sin-etiqueta (38.0/11.0)	38	11	71.1	3.6
10	IF commgr = acepta AND speaker_role = s AND obligations_minus1 = sin-etiqueta, THEN sin-etiqueta (34.0/15.0)	34	15	55.9	3.3
11	IF commgr = afirm AND obligations_minus1 = sol-inf, THEN resp (30.0/4.0)	30	4	86.7	2.9
12	IF commgr = ack, THEN sin-etiqueta (24.0)	24	0	100.0	2.3
13	IF commgr = afirm AND obligations_minus1 = sin-etiqueta, THEN sin-etiqueta (19.0/5.0)	19	5	73.7	1.8
14	IF commgr = sin-etiqueta AND commgr_minus1 = acepta AND number_of_tones > 15, THEN dir-accion (18.0/4.0)	18	4	77.8	1.7
15	IF commgr = acepta AND speaker_role = u AND commgr_minus1 = opcion-abierta_despliega, THEN dir-accion (18.0/8.0)	18	8	55.6	1.7
16	IF commgr = afirm AND obligations_minus1 = resp, THEN sin-etiqueta (16.0/3.0)	16	3	81.3	1.5
17	IF commgr = acepta-parte AND obligations_minus1 = sol-inf_accion-graf,	13	1	92.3	1.2

	THEN dir-accion_resp (13.0/1.0)				
18	IF commgr = espera, THEN sol-inf (13.0/4.0)	13	4	69.2	1.2
19	IF commgr = acepta AND speaker_role = u AND commgr_minus1 = acepta AND obligations_minus1 = resp, THEN resp (12.0)	12	0	100.0	1.2
20	IF commgr = SNE, THEN sol-inf (12.0)	12	0	100.0	1.2
21	IF commgr = accion-graf AND obligations_minus1 = sin-etiqueta AND commgr_minus1 = sin-etiqueta, THEN sol-inf_accion-graf (11.0)	11	0	100.0	1.1
22	IF commgr = rechaza, THEN sin-etiqueta (11.0/6.0)	11	6	45.5	1.1
23	IF commgr = reafirm AND obligations_minus1 = sol-inf, THEN resp (10.0)	10	0	100.0	1.0
24	IF commgr = sin-etiqueta AND commgr_minus1 = rechaza, THEN dir-accion (10.0/3.0)	10	3	70.0	1.0
25	IF commgr = acepta AND speaker_role = u AND commgr_minus1 = sin-etiqueta, THEN resp (10.0/4.0)	10	4	60.0	1.0
26	IF commgr = acepta AND speaker_role = s AND obligations_minus1 = dir-accion_resp, THEN compromiso (9.0/1.0)	9	1	88.9	0.9
27	IF commgr = sin-etiqueta AND commgr_minus1 = sin-etiqueta AND first_3 = MSH, THEN dir-accion (8.0/1.0)	8	1	87.5	0.8
28	IF commgr = accion-graf AND obligations_minus1 = sin-etiqueta AND commgr_minus1 = acepta, THEN sol-inf_accion-graf_resp (8.0/1.0)	8	1	87.5	0.8
29	IF commgr = acepta AND speaker_role = u AND commgr_minus1 = acepta AND obligations_minus1 = dir-accion, THEN dir-accion (7.0)	7	0	100.0	0.7
30	IF commgr = afirm AND obligations_minus1 = compromiso, THEN sin-etiqueta (7.0/1.0)	7	1	85.7	0.7
31	IF commgr = accion-graf AND obligations_minus1 = sin-etiqueta AND commgr_minus1 = ack, THEN sol-inf_accion-graf (6.0)	6	0	100.0	0.6
32	IF commgr = opcion-abierta, THEN sin-etiqueta (6.0)	6	0	100.0	0.6
33	IF commgr = accion-graf AND obligations_minus1 = dir-accion, THEN sol-inf_accion-graf (5.0)	5	0	100.0	0.5
34	IF commgr = repet-refr, THEN sin-etiqueta (5.0)	5	0	100.0	0.5
35	IF commgr = conv-cierra, THEN sin-etiqueta (5.0)	5	0	100.0	0.5
36	IF commgr = oferta_conv-abre, THEN sin-etiqueta (5.0)	5	0	100.0	0.5
37	IF commgr = sin-etiqueta AND commgr_minus1 = sin-etiqueta AND first_3 = MST AND number_of_tones <= 7, THEN sin-etiqueta (5.0/1.0)	5	1	80.0	0.5
38	IF commgr = acepta AND speaker_role = u AND commgr_minus1 = oferta_conv-abre, THEN oferta (5.0/1.0)	5	1	80.0	0.5

39	IF commgr = perform, THEN sin-etiqueta (4.0)	4	0	100.0	0.4
40	IF commgr = afirm_rechaza, THEN resp (4.0)	4	0	100.0	0.4
41	IF commgr = rechaza-parte, THEN resp (4.0)	4	0	100.0	0.4
42	IF commgr = sin-etiqueta AND commgr_minus1 = sin-etiqueta AND first_3 = MDB, THEN sin-etiqueta (4.0/1.0)	4	1	75.0	0.4
43	IF commgr = sin-etiqueta AND commgr_minus1 = accion-graf AND optimal_pred_mood = dec, THEN dir-accion (4.0/2.0)	4	2	50.0	0.4
44	IF commgr = acepta AND speaker_role = u AND commgr_minus1 = ack, THEN dir-accion (4.0/2.0)	4	2	50.0	0.4
45	IF commgr = acepta AND speaker_role = u AND commgr_minus1 = acepta AND obligations_minus1 = sin-etiqueta, THEN dir-accion (4.0/1.0)	4	1	75.0	0.4
46	IF commgr = acepta AND speaker_role = s AND obligations_minus1 = resp AND optimal_pred_mood = int, THEN sol-inf (4.0/1.0)	4	1	75.0	0.4
47	IF commgr = acepta AND speaker_role = u AND commgr_minus1 = afirm AND number_of_tones > 12, THEN dir-accion (3.0)	3	0	100.0	0.3
48	IF commgr = afirm_acepta AND obligations_minus1 = sin-etiqueta, THEN sin-etiqueta (3.0)	3	0	100.0	0.3
49	IF commgr = afirm_acepta AND obligations_minus1 = resp, THEN sin-etiqueta (3.0)	3	0	100.0	0.3
50	IF commgr = back-channel, THEN sin-etiqueta (3.0)	3	0	100.0	0.3
51	IF commgr = afirm_despliega, THEN resp (3.0)	3	0	100.0	0.3
52	IF commgr = sin-etiqueta AND commgr_minus1 = sin-etiqueta AND first_3 = MST AND number_of_tones > 7, THEN sol-inf (3.0/1.0)	3	1	66.7	0.3
53	IF commgr = sin-etiqueta AND commgr_minus1 = sin-etiqueta AND first_3 = BST, THEN sol-inf (3.0/1.0)	3	1	66.7	0.3
54	IF commgr = sin-etiqueta AND commgr_minus1 = sin-etiqueta AND first_3 = BUT, THEN sin-etiqueta (3.0/1.0)	3	1	66.7	0.3
55	IF commgr = sin-etiqueta AND commgr_minus1 = sin-etiqueta AND first_3 = MHD, THEN sol-inf (3.0/1.0)	3	1	66.7	0.3
56	IF commgr = sin-etiqueta AND commgr_minus1 = sin-etiqueta AND first_3 = MUU, THEN sol-inf (3.0/2.0)	3	2	33.3	0.3
57	IF commgr = sin-etiqueta AND commgr_minus1 = afirm, THEN dir-accion (3.0/1.0)	3	1	66.7	0.3
58	IF commgr = acepta AND speaker_role = u AND commgr_minus1 = afirm AND number_of_tones <= 12, THEN sin-etiqueta (3.0/1.0)	3	1	66.7	0.3
59	IF commgr = acepta AND speaker_role = u AND commgr_minus1 = acepta AND obligations_minus1 = sol-inf, THEN sol-inf (3.0/1.0)	3	1	66.7	0.3

60	IF commgr = reaffirm AND obligations_minus1 = sin-etiqueta, THEN sin-etiqueta (3.0/1.0)	3	1	66.7	0.3
61	IF commgr = accion-graf_acepta, THEN sol-inf_accion-graf (3.0/1.0)	3	1	66.7	0.3
62	IF commgr = opcion-abierta_despliega_acepta, THEN sin-etiqueta (3.0/1.0)	3	1	66.7	0.3
63	IF commgr = sin-etiqueta AND commgr_minus1 = sin-etiqueta AND first_3 = MTB AND obligations_minus1 = dir-accion, THEN sin-etiqueta (2.0)	2	0	100.0	0.2
64	IF commgr = sin-etiqueta AND commgr_minus1 = sin-etiqueta AND first_3 = MBT, THEN sol-inf (2.0)	2	0	100.0	0.2
65	IF commgr = sin-etiqueta AND commgr_minus1 = sin-etiqueta AND first_3 = MSU, THEN dir-accion (2.0)	2	0	100.0	0.2
66	IF commgr = sin-etiqueta AND commgr_minus1 = ack AND number_of_tones <= 8, THEN sin-etiqueta (2.0)	2	0	100.0	0.2
67	IF commgr = sin-etiqueta AND commgr_minus1 = ack AND number_of_tones > 8, THEN dir-accion (2.0)	2	0	100.0	0.2
68	IF commgr = sin-etiqueta AND commgr_minus1 = accion-graf AND optimal_pred_mood = int, THEN sol-inf (2.0)	2	0	100.0	0.2
69	IF commgr = sin-etiqueta AND commgr_minus1 = repet-refr, THEN dir-accion (2.0)	2	0	100.0	0.2
70	IF commgr = acepta_espera_repet-refr, THEN sol-inf (2.0)	2	0	100.0	0.2
71	IF commgr = acepta AND speaker_role = u AND commgr_minus1 = acepta_espera_repet-refr, THEN resp (2.0)	2	0	100.0	0.2
72	IF commgr = acepta AND speaker_role = u AND commgr_minus1 = acepta-parte, THEN sin-etiqueta (2.0)	2	0	100.0	0.2
73	IF commgr = acepta AND speaker_role = u AND commgr_minus1 = opcion-abierta AND number_of_tones <= 5, THEN sin-etiqueta (2.0)	2	0	100.0	0.2
74	IF commgr = afirm_quiza, THEN sin-etiqueta (2.0)	2	0	100.0	0.2
75	IF commgr = accion-graf AND obligations_minus1 = resp, THEN sol-inf_accion-graf (2.0)	2	0	100.0	0.2
76	IF commgr = afirm_acepta AND obligations_minus1 = sol-inf, THEN resp (2.0)	2	0	100.0	0.2
77	IF commgr = oferta_acepta, THEN sin-etiqueta (2.0)	2	0	100.0	0.2
78	IF commgr = acepta-parte AND obligations_minus1 = compromiso, THEN dir-accion (2.0)	2	0	100.0	0.2
79	IF commgr = acepta-parte AND obligations_minus1 = sol-inf_accion-graf_resp, THEN sol-inf_resp (2.0)	2	0	100.0	0.2
80	IF commgr = reaffirm AND obligations_minus1 = resp, THEN sin-etiqueta (2.0)	2	0	100.0	0.2
81	IF commgr = oferta, THEN sin-etiqueta (2.0)	2	0	100.0	0.2
82	IF commgr = afirm_espera, THEN sin-etiqueta	2	0	100.0	0.2

	(2.0)				
83	IF commgr = sin-etiqueta AND commgr_minus1 = sin-etiqueta AND first_3 = MTB AND obligations_minus1 = sin-etiqueta, THEN dir-accion (2.0/1.0)	2	1	50.0	0.2
84	IF commgr = sin-etiqueta AND commgr_minus1 = sin-etiqueta AND first_3 = MUT, THEN dir-accion (2.0/1.0)	2	1	50.0	0.2
85	IF commgr = sin-etiqueta AND commgr_minus1 = espera_repet-refr, THEN sol-inf (2.0/1.0)	2	1	50.0	0.2
86	IF commgr = sin-etiqueta AND commgr_minus1 = opcion-abierta_despliega, THEN sol-inf (2.0/1.0)	2	1	50.0	0.2
87	IF commgr = sin-etiqueta AND commgr_minus1 = back-channel, THEN sol-inf (2.0/1.0)	2	1	50.0	0.2
88	IF commgr = sin-etiqueta AND commgr_minus1 = acepta-parte, THEN dir-accion (2.0/1.0)	2	1	50.0	0.2
89	IF commgr = sin-etiqueta AND commgr_minus1 = reafirm, THEN sol-inf (2.0/1.0)	2	1	50.0	0.2
90	IF commgr = acepta AND speaker_role = u AND commgr_minus1 = opcion-abierta AND number_of_tones > 5, THEN sol-inf (2.0/1.0)	2	1	50.0	0.2
91	IF commgr = acepta AND speaker_role = u AND commgr_minus1 = oferta, THEN oferta_sol-inf (2.0/1.0)	2	1	50.0	0.2
92	IF commgr = acepta AND speaker_role = s AND obligations_minus1 = sol-inf_resp, THEN compromiso (2.0/1.0)	2	1	50.0	0.2
93	IF commgr = despliega, THEN resp (2.0/1.0)	2	1	50.0	0.2
94	IF commgr = sin-etiqueta AND commgr_minus1 = sin-etiqueta AND first_3 = MBU, THEN dir-accion (1.0)	1	0	100.0	0.1
95	IF commgr = sin-etiqueta AND commgr_minus1 = sin-etiqueta AND first_3 = BSU, THEN dir-accion (1.0)	1	0	100.0	0.1
96	IF commgr = sin-etiqueta AND commgr_minus1 = sin-etiqueta AND first_3 = MTD, THEN sin-etiqueta (1.0)	1	0	100.0	0.1
97	IF commgr = sin-etiqueta AND commgr_minus1 = sin-etiqueta AND first_3 = MTL, THEN sin-etiqueta (1.0)	1	0	100.0	0.1
98	IF commgr = sin-etiqueta AND commgr_minus1 = sin-etiqueta AND first_3 = MMS, THEN dir-accion (1.0)	1	0	100.0	0.1
99	IF commgr = sin-etiqueta AND commgr_minus1 = sin-etiqueta AND first_3 = MSL, THEN sin-etiqueta (1.0)	1	0	100.0	0.1
100	IF commgr = sin-etiqueta AND commgr_minus1 = sin-etiqueta AND first_3 = MML, THEN sin-etiqueta (1.0)	1	0	100.0	0.1
101	IF commgr = sin-etiqueta AND commgr_minus1 = sin-etiqueta AND first_3 = MLS, THEN sin-etiqueta (1.0)	1	0	100.0	0.1
102	IF commgr = sin-etiqueta AND commgr_minus1 = sin-etiqueta AND first_3 =	1	0	100.0	0.1

	MLT, THEN sol-inf (1.0)				
103	IF commgr = sin-etiqueta AND commgr_minus1 = sin-etiqueta AND first_3 = MMT, THEN dir-accion (1.0)	1	0	100.0	0.1
104	IF commgr = sin-etiqueta AND commgr_minus1 = sin-etiqueta AND first_3 = MLH, THEN sin-etiqueta (1.0)	1	0	100.0	0.1
105	IF commgr = sin-etiqueta AND commgr_minus1 = sin-etiqueta AND first_3 = MBH, THEN dir-accion (1.0)	1	0	100.0	0.1
106	IF commgr = sin-etiqueta AND commgr_minus1 = sin-etiqueta AND first_3 = BTL, THEN sin-etiqueta (1.0)	1	0	100.0	0.1
107	IF commgr = sin-etiqueta AND commgr_minus1 = sin-etiqueta AND first_3 = MSD, THEN sol-inf (1.0)	1	0	100.0	0.1
108	IF commgr = sin-etiqueta AND commgr_minus1 = sin-etiqueta AND first_3 = MDL, THEN sin-etiqueta (1.0)	1	0	100.0	0.1
109	IF commgr = sin-etiqueta AND commgr_minus1 = afirm_acepta, THEN sin-etiqueta (1.0)	1	0	100.0	0.1
110	IF commgr = sin-etiqueta AND commgr_minus1 = SNE, THEN dir-accion_resp (1.0)	1	0	100.0	0.1
111	IF commgr = sin-etiqueta AND commgr_minus1 = perform, THEN sol-inf (1.0)	1	0	100.0	0.1
112	IF commgr = sin-etiqueta AND commgr_minus1 = accion-graf_acepta, THEN sin-etiqueta (1.0)	1	0	100.0	0.1
113	IF commgr = sin-etiqueta AND commgr_minus1 = opcion-abierta, THEN sin-etiqueta (1.0)	1	0	100.0	0.1
114	IF commgr = sin-etiqueta AND commgr_minus1 = rechaza-parte, THEN dir-accion (1.0)	1	0	100.0	0.1
115	IF commgr = sin-etiqueta AND commgr_minus1 = afirm_accion-graf, THEN sin-etiqueta (1.0)	1	0	100.0	0.1
116	IF commgr = sin-etiqueta AND commgr_minus1 = opcion-abierta_despliega_acepta, THEN sol-inf (1.0)	1	0	100.0	0.1
117	IF commgr = sin-etiqueta AND commgr_minus1 = afirm_espera, THEN sol-inf (1.0)	1	0	100.0	0.1
118	IF commgr = acepta AND speaker_role = u AND commgr_minus1 = back-channel, THEN dir-accion (1.0)	1	0	100.0	0.1
119	IF commgr = acepta AND speaker_role = u AND commgr_minus1 = reafirm, THEN sin-etiqueta (1.0)	1	0	100.0	0.1
120	IF commgr = acepta AND speaker_role = u AND commgr_minus1 = oferta_acepta, THEN sol-inf (1.0)	1	0	100.0	0.1
121	IF commgr = acepta AND speaker_role = u AND commgr_minus1 = accion-graf_acepta, THEN resp (1.0)	1	0	100.0	0.1

122	IF commgr = acepta AND speaker_role = u AND commgr_minus1 = afirm_accion-graf, THEN sin-etiqueta (1.0)	1	0	100.0	0.1
123	IF commgr = acepta AND speaker_role = u AND commgr_minus1 = otra, THEN dir-accion (1.0)	1	0	100.0	0.1
124	IF commgr = acepta AND speaker_role = u AND commgr_minus1 = afirm_correccion, THEN sol-inf (1.0)	1	0	100.0	0.1
125	IF commgr = acepta AND speaker_role = u AND commgr_minus1 = espera_correccion, THEN resp (1.0)	1	0	100.0	0.1
126	IF commgr = acepta AND speaker_role = s AND obligations_minus1 = dir-accion_oferta, THEN compromiso (1.0)	1	0	100.0	0.1
127	IF commgr = afirm AND obligations_minus1 = sol-inf_resp, THEN sin-etiqueta (1.0)	1	0	100.0	0.1
128	IF commgr = afirm_acepta-parte_esin-etiquetaclama, THEN resp (1.0)	1	0	100.0	0.1
129	IF commgr = accion-graf AND obligations_minus1 = sol-inf_resp, THEN sol-inf_accion-graf_resp (1.0)	1	0	100.0	0.1
130	IF commgr = afirm_acepta AND obligations_minus1 = sol-inf_accion-graf, THEN dir-accion_resp (1.0)	1	0	100.0	0.1
131	IF commgr = afirm_acepta_esin-etiquetaclama, THEN sin-etiqueta (1.0)	1	0	100.0	0.1
132	IF commgr = espera_en_admon.-tarea, THEN sin-etiqueta (1.0)	1	0	100.0	0.1
133	IF commgr = acepta-parte AND obligations_minus1 = sol-inf, THEN dir-accion_resp (1.0)	1	0	100.0	0.1
134	IF commgr = acepta-parte AND obligations_minus1 = dir-accion, THEN dir-accion (1.0)	1	0	100.0	0.1
135	IF commgr = acepta-parte AND obligations_minus1 = dir-accion_resp, THEN dir-accion (1.0)	1	0	100.0	0.1
136	IF commgr = reafirm AND obligations_minus1 = dir-accion, THEN sin-etiqueta (1.0)	1	0	100.0	0.1
137	IF commgr = quiza, THEN sin-etiqueta (1.0)	1	0	100.0	0.1
138	IF commgr = reafirm_espera, THEN sol-inf (1.0)	1	0	100.0	0.1
139	IF commgr = opcion-abierta_acepta, THEN resp (1.0)	1	0	100.0	0.1
140	IF commgr = otra, THEN sin-etiqueta (1.0)	1	0	100.0	0.1
141	IF commgr = reafirm_complem, THEN resp (1.0)	1	0	100.0	0.1
142	IF commgr = conv-abre, THEN sin-etiqueta (1.0)	1	0	100.0	0.1
143	IF commgr = afirm_accion-graf, THEN accion-graf (1.0)	1	0	100.0	0.1
144	IF commgr = afirm_correccion, THEN sin-etiqueta (1.0)	1	0	100.0	0.1
145	IF commgr = opcion-abierta_rechaza, THEN sin-etiqueta (1.0)	1	0	100.0	0.1

146	IF commgr = afirm_perform_conv-cierra, THEN sin-etiqueta (1.0)	1	0	100.0	0.1
147	IF commgr = espera_SNE, THEN sol-inf (1.0)	1	0	100.0	0.1
148	IF commgr = afirm_conv-cierra, THEN resp (1.0)	1	0	100.0	0.1

**Apéndice XII. Reglas del modelo 3.1.B para actos de diálogo de *common ground*
(ordenadas por soporte)**

Núm. regla	Regla	Casos donde Ocurre la Premisa	Casos No Satisfactorios	Confianza %	Soporte %
1	IF obligations = compromiso, THEN acepta (112.0/3.0)	112	3	97.3	10.7
2	IF obligations = sol-inf AND speaker_role = s, THEN espera_repet-refr (99.0/47.0)	99	47	52.5	9.5
3	IF obligations = sol-inf_accion-graf, THEN accion-graf (98.0/2.0)	98	2	98.0	9.4
4	IF obligations = resp AND commgr_minus1 = accion-graf, THEN acepta (56.0/5.0)	56	5	91.1	5.4
5	IF obligations = resp AND commgr_minus1 = espera_repet-refr, THEN acepta (48.0/7.0)	48	7	85.4	4.6
6	IF obligations = sin-etiqueta AND commgr_minus1 = sin-etiqueta AND utt_duration > 811, THEN sin-etiqueta (40.0/18.0)	40	18	55.0	3.8
7	IF obligations = sol-inf AND speaker_role = u AND commgr_minus1 = acepta, THEN sin-etiqueta (34.0/2.0)	34	2	94.1	3.3
8	IF obligations = sin-etiqueta AND commgr_minus1 = sin-etiqueta AND utt_duration <= 811 AND speaker_role = s, THEN acepta (34.0/3.0)	34	3	91.2	3.3
9	IF obligations = dir-accion AND commgr_minus1 = acepta AND obligations_minus1 = resp, THEN sin-etiqueta (34.0)	34	0	100.0	3.3
10	IF obligations = resp AND commgr_minus1 = acepta AND speaker_role = s, THEN opcion-abierta_despliega (33.0/7.0)	33	7	78.8	3.2
11	IF obligations = sin-etiqueta AND commgr_minus1 = afirm, THEN acepta (30.0/20.0)	30	20	33.3	2.9
12	IF obligations = dir-accion AND commgr_minus1 = sin-etiqueta, THEN sin-etiqueta (22.0/2.0)	22	2	90.9	2.1
13	IF obligations = resp AND commgr_minus1 = sin-etiqueta AND number_of_tones > 5, THEN afirm (18.0/8.0)	18	8	55.6	1.7
14	IF obligations = resp AND commgr_minus1 = sin-etiqueta AND number_of_tones <= 5, THEN acepta (17.0/4.0)	17	4	76.5	1.6
15	IF obligations = dir-accion_resp, THEN acepta-parte (16.0/3.0)	16	3	81.3	1.5
16	IF obligations = resp AND commgr_minus1 = acepta AND speaker_role = u AND obligations_minus1 = resp, THEN acepta (13.0/1.0)	13	1	92.3	1.2
17	IF obligations = sol-inf AND speaker_role = u AND commgr_minus1 = sin-etiqueta, THEN sin-etiqueta (12.0/1.0)	12	1	91.7	1.2
18	IF obligations = dir-accion AND commgr_minus1 = acepta AND obligations_minus1 = sin-etiqueta, THEN sin-etiqueta (10.0/3.0)	10	3	70.0	1.0
19	IF obligations = dir-accion AND commgr_minus1 = opcion-abierta_despliega, THEN acepta (10.0)	10	0	100.0	1.0
20	IF obligations = resp AND commgr_minus1 = SNE, THEN reafirm (9.0)	9	0	100.0	0.9

21	IF obligations = sol-inf AND speaker_role = u AND commgr_minus1 = opcion-abierta_despliega, THEN acepta (8.0/2.0)	8	2	75.0	0.8
22	IF obligations = resp AND commgr_minus1 = acepta AND speaker_role = u AND obligations_minus1 = sol-inf, THEN afirm (8.0/1.0)	8	1	87.5	0.8
23	IF obligations = dir-accion AND commgr_minus1 = acepta AND obligations_minus1 = dir-accion, THEN acepta (8.0/1.0)	8	1	87.5	0.8
24	IF obligations = sin-etiqueta AND commgr_minus1 = acepta AND first_3 = MTB AND speaker_role = s, THEN acepta (8.0/3.0)	8	3	62.5	0.8
25	IF obligations = sol-inf_accion-graf_resp, THEN accion-graf (8.0)	8	0	100.0	0.8
26	IF obligations = resp AND commgr_minus1 = espera, THEN afirm (7.0/2.0)	7	2	71.4	0.7
27	IF obligations = sin-etiqueta AND commgr_minus1 = nil, THEN oferta_conv-abre (7.0/2.0)	7	2	71.4	0.7
28	IF obligations = dir-accion AND commgr_minus1 = rechaza, THEN sin-etiqueta (7.0)	7	0	100.0	0.7
29	IF obligations = sin-etiqueta AND commgr_minus1 = acepta-parte, THEN acepta (5.0/2.0)	5	2	60.0	0.5
30	IF obligations = sin-etiqueta AND commgr_minus1 = acepta AND first_3 = MTD AND obligations_minus1 = sin-etiqueta, THEN opcion-abierta_despliega (5.0)	5	0	100.0	0.5
31	IF obligations = oferta, THEN acepta (5.0)	5	0	100.0	0.5
32	IF obligations = sol-inf AND speaker_role = u AND commgr_minus1 = accion-graf, THEN espera (4.0/1.0)	4	1	75.0	0.4
33	IF obligations = dir-accion AND commgr_minus1 = acepta AND obligations_minus1 = compromiso AND number_of_tones > 5, THEN sin-etiqueta (4.0/1.0)	4	1	75.0	0.4
34	IF obligations = sin-etiqueta AND commgr_minus1 = sin-etiqueta AND utt_duration <= 811 AND speaker_role = u, THEN ack (4.0/2.0)	4	2	50.0	0.4
35	IF obligations = sin-etiqueta AND commgr_minus1 = acepta AND first_3 = MST AND utt_duration > 1356, THEN afirm (4.0/1.0)	4	1	75.0	0.4
36	IF obligations = sol-inf AND speaker_role = u AND commgr_minus1 = ack, THEN espera (3.0/1.0)	3	1	66.7	0.3
37	IF obligations = sol-inf AND speaker_role = u AND commgr_minus1 = afirm, THEN sin-etiqueta (3.0/2.0)	3	2	33.3	0.3
38	IF obligations = resp AND commgr_minus1 = afirm AND speaker_role = s, THEN acepta (3.0/2.0)	3	2	33.3	0.3
39	IF obligations = dir-accion AND commgr_minus1 = ack AND utt_duration > 5006, THEN acepta (3.0/1.0)	3	1	66.7	0.3
40	IF obligations = dir-accion AND commgr_minus1 = acepta-parte, THEN acepta-parte (3.0/1.0)	3	1	66.7	0.3
41	IF obligations = sin-etiqueta AND commgr_minus1 = accion-graf, THEN sin-etiqueta (3.0/2.0)	3	2	33.3	0.3
42	IF obligations = sin-etiqueta AND commgr_minus1 = acepta AND first_3 = MTB AND speaker_role = u, THEN afirm (3.0/1.0)	3	1	66.7	0.3
43	IF obligations = sin-etiqueta AND commgr_minus1 = acepta AND first_3 = MB AND speaker_role = u, THEN sin-etiqueta (3.0/2.0)	3	2	33.3	0.3

44	IF obligations = sin-etiqueta AND commgr_minus1 = acepta AND first_3 = MBT, THEN sin-etiqueta (3.0/2.0)	3	2	33.3	0.3
45	IF obligations = sin-etiqueta AND commgr_minus1 = acepta AND first_3 = MHL, THEN sin-etiqueta (3.0/2.0)	3	2	33.3	0.3
46	IF obligations = sin-etiqueta AND commgr_minus1 = acepta AND first_3 = MLT, THEN sin-etiqueta (3.0/2.0)	3	2	33.3	0.3
47	IF obligations = sin-etiqueta AND commgr_minus1 = acepta AND first_3 = MDS, THEN sin-etiqueta (3.0/1.0)	3	1	66.7	0.3
48	IF obligations = sin-etiqueta AND commgr_minus1 = opcion-abierta_despliega AND first_2 = MT, THEN sin-etiqueta (3.0/2.0)	3	2	33.3	0.3
49	IF obligations = sin-etiqueta AND commgr_minus1 = back-channel, THEN acepta (3.0/1.0)	3	1	66.7	0.3
50	IF obligations = sin-etiqueta AND commgr_minus1 = opcion-abierta, THEN acepta (3.0/1.0)	3	1	66.7	0.3
51	IF obligations = sin-etiqueta AND commgr_minus1 = conv-cierra AND number_of_tones <= 4, THEN acepta (3.0/1.0)	3	1	66.7	0.3
52	IF obligations = sol-inf_resp AND obligations_minus1 = sol-inf, THEN acepta (3.0/2.0)	3	2	33.3	0.3
53	IF obligations = sol-inf AND speaker_role = u AND commgr_minus1 = rechaza, THEN sin-etiqueta (3.0)	3	0	100.0	0.3
54	IF obligations = dir-accion AND commgr_minus1 = afirm AND number_of_tones <= 17, THEN acepta (3.0)	3	0	100.0	0.3
55	IF obligations = sin-etiqueta AND commgr_minus1 = acepta AND first_3 = MB AND speaker_role = s AND utt_duration > 401, THEN acepta (3.0)	3	0	100.0	0.3
56	IF obligations = sin-etiqueta AND commgr_minus1 = afirm_acepta AND number_of_tones > 3, THEN acepta (3.0)	3	0	100.0	0.3
57	IF obligations = sin-etiqueta AND commgr_minus1 = espera, THEN acepta (3.0)	3	0	100.0	0.3
58	IF obligations = sin-etiqueta AND commgr_minus1 = afirm_despliega, THEN opcion-abierta (3.0)	3	0	100.0	0.3
59	IF obligations = dir-accion AND commgr_minus1 = ack AND utt_duration <= 5006, THEN sin-etiqueta (2.0/1.0)	2	1	50.0	0.2
60	IF obligations = dir-accion AND commgr_minus1 = acepta AND obligations_minus1 = compromiso AND number_of_tones <= 5, THEN afirm (2.0/1.0)	2	1	50.0	0.2
61	IF obligations = sin-etiqueta AND commgr_minus1 = ack AND first_2 = MS, THEN sin-etiqueta (2.0/1.0)	2	1	50.0	0.2
62	IF obligations = sin-etiqueta AND commgr_minus1 = acepta AND first_3 = MST AND utt_duration <= 1356, THEN ack (2.0/1.0)	2	1	50.0	0.2
63	IF obligations = sin-etiqueta AND commgr_minus1 = acepta AND first_3 = MTD AND obligations_minus1 = resp AND last_2 = DB, THEN ack (2.0/1.0)	2	1	50.0	0.2
64	IF obligations = sin-etiqueta AND commgr_minus1 = acepta AND first_3 = MSH, THEN sin-etiqueta (2.0/1.0)	2	1	50.0	0.2
65	IF obligations = sin-etiqueta AND commgr_minus1 = acepta AND first_3 = BTS, THEN sin-etiqueta (2.0/1.0)	2	1	50.0	0.2
66	IF obligations = sin-etiqueta AND commgr_minus1 = acepta AND first_3 = MSL, THEN sin-etiqueta (2.0/1.0)	2	1	50.0	0.2
67	IF obligations = sin-etiqueta AND commgr_minus1 = acepta AND first_3 = BTM, THEN sin-etiqueta (2.0/1.0)	2	1	50.0	0.2

68	IF obligations = sin-etiqueta AND commgr_minus1 = acepta AND first_3 = MSD, THEN sin-etiqueta (2.0/1.0)	2	1	50.0	0.2
69	IF obligations = sin-etiqueta AND commgr_minus1 = repet-refr AND speaker_role = u, THEN repet-refr (2.0/1.0)	2	1	50.0	0.2
70	IF obligations = sin-etiqueta AND commgr_minus1 = afirm_acepta AND number_of_tones <= 3, THEN sin-etiqueta (2.0/1.0)	2	1	50.0	0.2
71	IF obligations = sin-etiqueta AND commgr_minus1 = opcion-abierta_despliega AND first_2 = MS, THEN opcion-abierta_despliega (2.0/1.0)	2	1	50.0	0.2
72	IF obligations = sin-etiqueta AND commgr_minus1 = rechaza AND first_3 = MTB, THEN acepta (2.0/1.0)	2	1	50.0	0.2
73	IF obligations = sin-etiqueta AND commgr_minus1 = reaffirm AND number_of_tones <= 5, THEN sin-etiqueta (2.0/1.0)	2	1	50.0	0.2
74	IF obligations = sin-etiqueta AND commgr_minus1 = reaffirm AND number_of_tones > 5, THEN reaffirm (2.0/1.0)	2	1	50.0	0.2
75	IF obligations = sin-etiqueta AND commgr_minus1 = afirm_rechaza, THEN acepta (2.0/1.0)	2	1	50.0	0.2
76	IF obligations = sin-etiqueta AND commgr_minus1 = accion-graf_acepta, THEN sin-etiqueta (2.0/1.0)	2	1	50.0	0.2
77	IF obligations = sin-etiqueta AND commgr_minus1 = afirm_accion-graf, THEN sin-etiqueta (2.0/1.0)	2	1	50.0	0.2
78	IF obligations = sin-etiqueta AND commgr_minus1 = afirm_conv-cierra, THEN acepta (2.0/1.0)	2	1	50.0	0.2
79	IF obligations = sol-inf_resp AND obligations_minus1 = sol-inf_accion-graf, THEN espera (2.0/1.0)	2	1	50.0	0.2
80	IF obligations = resp AND commgr_minus1 = acepta_espera_repet-refr, THEN acepta (2.0)	2	0	100.0	0.2
81	IF obligations = resp AND commgr_minus1 = afirm AND speaker_role = u, THEN afirm (2.0)	2	0	100.0	0.2
82	IF obligations = resp AND commgr_minus1 = acepta AND speaker_role = u AND obligations_minus1 = sin-etiqueta, THEN afirm (2.0)	2	0	100.0	0.2
83	IF obligations = dir-accion AND commgr_minus1 = afirm AND number_of_tones > 17, THEN sin-etiqueta (2.0)	2	0	100.0	0.2
84	IF obligations = dir-accion AND commgr_minus1 = accion-graf, THEN sin-etiqueta (2.0)	2	0	100.0	0.2
85	IF obligations = dir-accion AND commgr_minus1 = repet-refr, THEN sin-etiqueta (2.0)	2	0	100.0	0.2
86	IF obligations = sin-etiqueta AND commgr_minus1 = ack AND first_2 = MT, THEN conv-cierra (2.0)	2	0	100.0	0.2
87	IF obligations = sin-etiqueta AND commgr_minus1 = espera_repet-refr, THEN acepta (2.0)	2	0	100.0	0.2
88	IF obligations = sin-etiqueta AND commgr_minus1 = afirm_quiza, THEN afirm_quiza (2.0)	2	0	100.0	0.2
89	IF obligations = sin-etiqueta AND commgr_minus1 = acepta AND first_3 = MTD AND obligations_minus1 = resp AND last_2 = BS, THEN acepta (2.0)	2	0	100.0	0.2
90	IF obligations = sin-etiqueta AND commgr_minus1 = acepta AND first_3 = MTD AND obligations_minus1 = resp AND last_2 = BH, THEN afirm (2.0)	2	0	100.0	0.2

91	IF obligations = sin-etiqueta AND commgr_minus1 = acepta AND first_3 = MTD AND obligations_minus1 = compromiso, THEN afirm (2.0)	2	0	100.0	0.2
92	IF obligations = sin-etiqueta AND commgr_minus1 = acepta AND first_3 = MTL, THEN sin-etiqueta (2.0)	2	0	100.0	0.2
93	IF obligations = sin-etiqueta AND commgr_minus1 = acepta AND first_3 = MB AND speaker_role = s AND utt_duration <= 401, THEN ack (2.0)	2	0	100.0	0.2
94	IF obligations = sin-etiqueta AND commgr_minus1 = acepta AND first_3 = MHS, THEN afirm (2.0)	2	0	100.0	0.2
95	IF obligations = sin-etiqueta AND commgr_minus1 = acepta AND first_3 = MLH, THEN afirm (2.0)	2	0	100.0	0.2
96	IF obligations = sin-etiqueta AND commgr_minus1 = repet-refr AND speaker_role = s, THEN acepta (2.0)	2	0	100.0	0.2
97	IF obligations = sin-etiqueta AND commgr_minus1 = opcion-abierta_despliega AND first_2 = MD, THEN afirm_acepta (2.0)	2	0	100.0	0.2
98	IF obligations = sin-etiqueta AND commgr_minus1 = rechaza AND first_3 = MSU, THEN afirm (2.0)	2	0	100.0	0.2
99	IF obligations = sin-etiqueta AND commgr_minus1 = conv-cierra AND number_of_tones > 4, THEN conv-cierra (2.0)	2	0	100.0	0.2
100	IF obligations = sol-inf_resp AND obligations_minus1 = sol-inf_accion-graf_resp, THEN acepta-parte (2.0)	2	0	100.0	0.2
101	IF obligations = sol-inf AND speaker_role = u AND commgr_minus1 = espera_repet-refr, THEN sin-etiqueta (1.0)	1	0	100.0	0.1
102	IF obligations = sol-inf AND speaker_role = u AND commgr_minus1 = perform, THEN sin-etiqueta (1.0)	1	0	100.0	0.1
103	IF obligations = sol-inf AND speaker_role = u AND commgr_minus1 = reafirm, THEN sin-etiqueta (1.0)	1	0	100.0	0.1
104	IF obligations = sol-inf AND speaker_role = u AND commgr_minus1 = afirm_rechaza, THEN espera_repet-refr (1.0)	1	0	100.0	0.1
105	IF obligations = sol-inf AND speaker_role = u AND commgr_minus1 = espera, THEN espera (1.0)	1	0	100.0	0.1
106	IF obligations = sol-inf AND speaker_role = u AND commgr_minus1 = oferta_acepta, THEN acepta (1.0)	1	0	100.0	0.1
107	IF obligations = sol-inf AND speaker_role = u AND commgr_minus1 = opcion-abierta, THEN acepta (1.0)	1	0	100.0	0.1
108	IF obligations = sol-inf AND speaker_role = u AND commgr_minus1 = afirm_correccion, THEN acepta (1.0)	1	0	100.0	0.1
109	IF obligations = sol-inf AND speaker_role = u AND commgr_minus1 = afirm_espera, THEN sin-etiqueta (1.0)	1	0	100.0	0.1
110	IF obligations = resp AND commgr_minus1 = ack, THEN acepta (1.0)	1	0	100.0	0.1
111	IF obligations = resp AND commgr_minus1 = reafirm, THEN afirm (1.0)	1	0	100.0	0.1
112	IF obligations = resp AND commgr_minus1 = accion-graf_acepta, THEN acepta (1.0)	1	0	100.0	0.1
113	IF obligations = resp AND commgr_minus1 = reafirm_espera, THEN rechaza (1.0)	1	0	100.0	0.1
114	IF obligations = resp AND commgr_minus1 = rechaza-parte, THEN rechaza-parte (1.0)	1	0	100.0	0.1

115	IF obligations = resp AND commgr_minus1 = espera_correccion, THEN acepta (1.0)	1	0	100.0	0.1
116	IF obligations = resp AND commgr_minus1 = espera_SNE, THEN reafirm (1.0)	1	0	100.0	0.1
117	IF obligations = dir-accion AND commgr_minus1 = acepta AND obligations_minus1 = oferta, THEN sin-etiqueta (1.0)	1	0	100.0	0.1
118	IF obligations = dir-accion AND commgr_minus1 = back-channel, THEN acepta (1.0)	1	0	100.0	0.1
119	IF obligations = dir-accion AND commgr_minus1 = oferta_conv-abre, THEN acepta (1.0)	1	0	100.0	0.1
120	IF obligations = dir-accion AND commgr_minus1 = opcion-abierta, THEN acepta (1.0)	1	0	100.0	0.1
121	IF obligations = dir-accion AND commgr_minus1 = rechaza-parte, THEN sin-etiqueta (1.0)	1	0	100.0	0.1
122	IF obligations = dir-accion AND commgr_minus1 = otra, THEN acepta (1.0)	1	0	100.0	0.1
123	IF obligations = sin-etiqueta AND commgr_minus1 = ack AND first_2 = BU, THEN afirm (1.0)	1	0	100.0	0.1
124	IF obligations = sin-etiqueta AND commgr_minus1 = ack AND first_2 = MB, THEN sin-etiqueta (1.0)	1	0	100.0	0.1
125	IF obligations = sin-etiqueta AND commgr_minus1 = ack AND first_2 = MH, THEN afirm (1.0)	1	0	100.0	0.1
126	IF obligations = sin-etiqueta AND commgr_minus1 = ack AND first_2 = BH, THEN ack (1.0)	1	0	100.0	0.1
127	IF obligations = sin-etiqueta AND commgr_minus1 = acepta AND first_3 = BST, THEN perform (1.0)	1	0	100.0	0.1
128	IF obligations = sin-etiqueta AND commgr_minus1 = acepta AND first_3 = MBU, THEN otra (1.0)	1	0	100.0	0.1
129	IF obligations = sin-etiqueta AND commgr_minus1 = acepta AND first_3 = BSU, THEN sin-etiqueta (1.0)	1	0	100.0	0.1
130	IF obligations = sin-etiqueta AND commgr_minus1 = acepta AND first_3 = MSB, THEN ack (1.0)	1	0	100.0	0.1
131	IF obligations = sin-etiqueta AND commgr_minus1 = acepta AND first_3 = BT, THEN acepta (1.0)	1	0	100.0	0.1
132	IF obligations = sin-etiqueta AND commgr_minus1 = acepta AND first_3 = MUT, THEN acepta (1.0)	1	0	100.0	0.1
133	IF obligations = sin-etiqueta AND commgr_minus1 = acepta AND first_3 = MTS, THEN sin-etiqueta (1.0)	1	0	100.0	0.1
134	IF obligations = sin-etiqueta AND commgr_minus1 = acepta AND first_3 = MBS, THEN sin-etiqueta (1.0)	1	0	100.0	0.1
135	IF obligations = sin-etiqueta AND commgr_minus1 = acepta AND first_3 = MLS, THEN afirm (1.0)	1	0	100.0	0.1
136	IF obligations = sin-etiqueta AND commgr_minus1 = acepta AND first_3 = MLU, THEN afirm (1.0)	1	0	100.0	0.1
137	IF obligations = sin-etiqueta AND commgr_minus1 = acepta AND first_3 = BTL, THEN acepta (1.0)	1	0	100.0	0.1
138	IF obligations = sin-etiqueta AND commgr_minus1 = acepta AND first_3 = MDL, THEN afirm (1.0)	1	0	100.0	0.1
139	IF obligations = sin-etiqueta AND commgr_minus1 = acepta AND first_3 = MDD, THEN afirm (1.0)	1	0	100.0	0.1
140	IF obligations = sin-etiqueta AND commgr_minus1 = opcion-abierta_despliega AND first_2 = MB, THEN rechaza (1.0)	1	0	100.0	0.1

141	IF obligations = sin-etiqueta AND commgr_minus1 = opcion-abierta_despliega AND first_2 = BT, THEN acepta (1.0)	1	0	100.0	0.1
142	IF obligations = sin-etiqueta AND commgr_minus1 = opcion-abierta_despliega AND first_2 = ML, THEN afirm_acepta (1.0)	1	0	100.0	0.1
143	IF obligations = sin-etiqueta AND commgr_minus1 = rechaza AND first_3 = MST, THEN afirm (1.0)	1	0	100.0	0.1
144	IF obligations = sin-etiqueta AND commgr_minus1 = rechaza AND first_3 = MTL, THEN rechaza (1.0)	1	0	100.0	0.1
145	IF obligations = sin-etiqueta AND commgr_minus1 = rechaza AND first_3 = MSD, THEN acepta (1.0)	1	0	100.0	0.1
146	IF obligations = sin-etiqueta AND commgr_minus1 = oferta_conv-abre, THEN rechaza (1.0)	1	0	100.0	0.1
147	IF obligations = sin-etiqueta AND commgr_minus1 = rechaza-parte, THEN afirm (1.0)	1	0	100.0	0.1
148	IF obligations = sin-etiqueta AND commgr_minus1 = opcion-abierta_despliega_acepta, THEN afirm_acepta (1.0)	1	0	100.0	0.1
149	IF obligations = sin-etiqueta AND commgr_minus1 = afirm_espera, THEN afirm (1.0)	1	0	100.0	0.1
150	IF obligations = sol-inf_resp AND obligations_minus1 = sin-etiqueta, THEN sin-etiqueta (1.0)	1	0	100.0	0.1
151	IF obligations = accion-graf, THEN afirm_accion-graf (1.0)	1	0	100.0	0.1
152	IF obligations = oferta_sol-inf, THEN acepta (1.0)	1	0	100.0	0.1
153	IF obligations = dir-accion_oferta, THEN acepta (1.0)	1	0	100.0	0.1

Apéndice XIII. Árboles de decisión generados con WEKA para la configuración 1

Modelo 1.1.A

Target: *obligations*

=== Run information ===

```
Scheme:          weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:
univ_data_one_int_firstintsint_commgr_with_optimal_predmood_numoftones_
USER_21dic2006-weka.filters.unsupervised.attribute.Remove-R1-4,8-10,12-
14,16-17,19,21-weka.filters.unsupervised.attribute.Remove-R5
Instances:      516
Attributes:     6
                first_1
                first_2
                first_3
                last_2
                utt_duration
                obligations
Test mode:     10-fold cross-validation
```

=== Classifier model (full training set) ===

J48 pruned tree

```
first_3 = MST
|  utt_duration <= 1692: x (12.0/5.0)
|  utt_duration > 1692: dir-accion (28.0/14.0)
first_3 = BST: resp (8.0/3.0)
first_3 = MBU
|  utt_duration <= 2629.5: resp (7.0/2.0)
|  utt_duration > 2629.5: sol-inf (3.0/1.0)
first_3 = BSU: resp (14.0/7.0)
first_3 = MTB
|  utt_duration <= 526.8125: resp (20.0/7.0)
|  utt_duration > 526.8125: x (34.0/20.0)
first_3 = MTL
|  last_2 = DB: resp (4.0/2.0)
|  last_2 = LH: resp (0.0)
|  last_2 = UT: sol-inf (1.0)
|  last_2 = LT: resp (0.0)
|  last_2 = BS: resp (2.0/1.0)
|  last_2 = TS: resp (0.0)
|  last_2 = LS: resp (0.0)
|  last_2 = ST: resp (0.0)
|  last_2 = DS: x (1.0)
|  last_2 = BT: resp (0.0)
|  last_2 = TL: resp (0.0)
|  last_2 = US: resp (1.0)
|  last_2 = HL: resp (0.0)
|  last_2 = BH: x (1.0)
|  last_2 = MB: resp (0.0)
|  last_2 = SH: resp (0.0)
```

```
| last_2 = HS: resp (0.0)
| last_2 = TB: resp (0.0)
| last_2 = SB: sol-inf,_resp (1.0)
| last_2 = UH: resp (0.0)
| last_2 = DL: resp (0.0)
| last_2 = BU: resp (0.0)
| last_2 = HD: dir-accion,_resp (1.0)
| last_2 = LU: resp (0.0)
| last_2 = HB: x (1.0)
| last_2 = UM: resp (0.0)
| last_2 = DD: x (1.0)
| last_2 = SL: resp (0.0)
| last_2 = DM: resp (0.0)
| last_2 = DU: resp (0.0)
| last_2 = SU: resp (0.0)
first_3 = MSH: dir-accion (38.0/17.0)
first_3 = MSB
| last_2 = DB: dir-accion (0.0)
| last_2 = LH: dir-accion (0.0)
| last_2 = UT: dir-accion (3.0/1.0)
| last_2 = LT: dir-accion (0.0)
| last_2 = BS: dir-accion (1.0)
| last_2 = TS: dir-accion (1.0)
| last_2 = LS: x (1.0)
| last_2 = ST: resp (1.0)
| last_2 = DS: x (1.0)
| last_2 = BT
| | utt_duration <= 852.375: x (3.0/1.0)
| | utt_duration > 852.375: resp (2.0)
| last_2 = TL: dir-accion (0.0)
| last_2 = US: dir-accion (1.0)
| last_2 = HL: sol-inf (1.0)
| last_2 = BH: dir-accion (0.0)
| last_2 = MB: dir-accion (0.0)
| last_2 = SH: dir-accion (0.0)
| last_2 = HS: dir-accion (0.0)
| last_2 = TB: dir-accion (0.0)
| last_2 = SB: dir-accion (0.0)
| last_2 = UH: dir-accion (0.0)
| last_2 = DL: dir-accion (0.0)
| last_2 = BU: dir-accion (0.0)
| last_2 = HD: dir-accion (0.0)
| last_2 = LU: dir-accion (0.0)
| last_2 = HB: dir-accion (0.0)
| last_2 = UM: dir-accion (0.0)
| last_2 = DD: dir-accion (0.0)
| last_2 = SL: dir-accion (0.0)
| last_2 = DM: dir-accion (0.0)
| last_2 = DU: dir-accion (0.0)
| last_2 = SU: dir-accion (0.0)
first_3 = MHD
| utt_duration <= 3469.125: sol-inf (4.0)
| utt_duration > 3469.125: dir-accion (2.0/1.0)
first_3 = MMS
| utt_duration <= 4183.5: resp (2.0/1.0)
| utt_duration > 4183.5: sol-inf (3.0/1.0)
first_3 = MBT: resp (27.0/11.0)
```

```

first_3 = MSU: dir-accion (13.0/6.0)
first_3 = BT
|   utt_duration <= 466.9375: resp (3.0)
|   utt_duration > 466.9375: x (3.0/1.0)
first_3 = MUT: sol-inf (9.0/5.0)
first_3 = MB: resp (33.0/7.0)
first_3 = MHS
|   utt_duration <= 4769.875: sol-inf (4.0/2.0)
|   utt_duration > 4769.875: dir-accion (3.0/1.0)
first_3 = BSH: dir-accion (3.0/1.0)
first_3 = MSL: dir-accion (21.0/14.0)
first_3 = MTS: resp (9.0/6.0)
first_3 = MHL
|   last_2 = DB: x (3.0/1.0)
|   last_2 = LH: dir-accion (1.0)
|   last_2 = UT: dir-accion (0.0)
|   last_2 = LT: dir-accion (0.0)
|   last_2 = BS: dir-accion (0.0)
|   last_2 = TS: x (1.0)
|   last_2 = LS: dir-accion (0.0)
|   last_2 = ST: dir-accion (0.0)
|   last_2 = DS: dir-accion (0.0)
|   last_2 = BT: sol-inf (1.0)
|   last_2 = TL: dir-accion (0.0)
|   last_2 = US: dir-accion (0.0)
|   last_2 = HL: dir-accion (0.0)
|   last_2 = BH: dir-accion (2.0/1.0)
|   last_2 = MB: dir-accion (0.0)
|   last_2 = SH: dir-accion (0.0)
|   last_2 = HS: dir-accion,_resp (1.0)
|   last_2 = TB: dir-accion,_resp (1.0)
|   last_2 = SB: dir-accion (2.0)
|   last_2 = UH: dir-accion (0.0)
|   last_2 = DL: dir-accion (0.0)
|   last_2 = BU: dir-accion (0.0)
|   last_2 = HD: dir-accion (0.0)
|   last_2 = LU: dir-accion (0.0)
|   last_2 = HB: dir-accion (0.0)
|   last_2 = UM: dir-accion (0.0)
|   last_2 = DD: dir-accion (0.0)
|   last_2 = SL: dir-accion (0.0)
|   last_2 = DM: dir-accion (0.0)
|   last_2 = DU: dir-accion (0.0)
|   last_2 = SU: dir-accion (0.0)
first_3 = MML: x (1.0)
first_3 = MLS
|   utt_duration <= 6930.25
|   |   utt_duration <= 3726.375: sol-inf (2.0/1.0)
|   |   utt_duration > 3726.375: x (3.0/1.0)
|   utt_duration > 6930.25: dir-accion (2.0)
first_3 = BHL
|   utt_duration <= 1410.25: resp (4.0)
|   utt_duration > 1410.25: sol-inf (3.0/1.0)
first_3 = MMT
|   utt_duration <= 3971.1875: dir-accion (2.0)
|   utt_duration > 3971.1875: sol-inf (2.0)
first_3 = MDS

```

```
|   utt_duration <= 4535.875: x (5.0/2.0)
|   utt_duration > 4535.875: dir-accion (2.0)
first_3 = MLH
|   utt_duration <= 2241.25: resp (2.0)
|   utt_duration > 2241.25
|   |   utt_duration <= 5648.625: dir-accion (3.0/1.0)
|   |   utt_duration > 5648.625: x (2.0)
first_3 = MLU
|   utt_duration <= 3414.8125: x (2.0/1.0)
|   utt_duration > 3414.8125: dir-accion (2.0)
first_3 = MHB
|   utt_duration <= 2310: sol-inf (2.0/1.0)
|   utt_duration > 2310: dir-accion (2.0/1.0)
first_3 = MDB: x (8.0/4.0)
first_3 = MBH: dir-accion (1.0)
first_3 = MDH: sol-inf (2.0/1.0)
first_3 = MLT: resp (20.0/9.0)
first_3 = BUU: resp (3.0/1.0)
first_3 = BTD: resp (5.0/3.0)
first_3 = MUH: x (4.0/2.0)
first_3 = MSD
|   utt_duration <= 1554.5625: resp (4.0/2.0)
|   utt_duration > 1554.5625: dir-accion (4.0/2.0)
first_3 = MDL: x (3.0/1.0)
first_3 = MTD
|   last_2 = DB
|   |   utt_duration <= 4113.5625: resp (11.0/3.0)
|   |   utt_duration > 4113.5625: dir-accion (2.0)
|   last_2 = LH: dir-accion (1.0)
|   last_2 = UT: resp (0.0)
|   last_2 = LT: resp (0.0)
|   last_2 = BS: x (7.0/3.0)
|   last_2 = TS: resp (0.0)
|   last_2 = LS: resp (0.0)
|   last_2 = ST: resp (0.0)
|   last_2 = DS: resp (0.0)
|   last_2 = BT: resp (0.0)
|   last_2 = TL: resp (0.0)
|   last_2 = US: resp (1.0)
|   last_2 = HL: sol-inf (2.0/1.0)
|   last_2 = BH: dir-accion,_resp (1.0)
|   last_2 = MB: resp (0.0)
|   last_2 = SH: sol-inf (1.0)
|   last_2 = HS: dir-accion,_resp (1.0)
|   last_2 = TB: resp (0.0)
|   last_2 = SB: sol-inf (1.0)
|   last_2 = UH: resp (0.0)
|   last_2 = DL: resp (0.0)
|   last_2 = BU: resp (0.0)
|   last_2 = HD: resp (0.0)
|   last_2 = LU: resp (0.0)
|   last_2 = HB: sol-inf (1.0)
|   last_2 = UM: resp (0.0)
|   last_2 = DD: resp (0.0)
|   last_2 = SL: resp (0.0)
|   last_2 = DM: resp (0.0)
|   last_2 = DU: resp (0.0)
```

```

| last_2 = SU: resp (0.0)
first_3 = BUT
| utt_duration <= 1856.6875: x (3.0/1.0)
| utt_duration > 1856.6875: sol-inf (3.0/2.0)
first_3 = BTS: x (3.0)
first_3 = BHS: resp (3.0/2.0)
first_3 = BTL
| utt_duration <= 348.1875: resp (2.0)
| utt_duration > 348.1875: x (5.0/1.0)
first_3 = MDD
| utt_duration <= 3835: x (2.0/1.0)
| utt_duration > 3835: sol-inf (2.0/1.0)
first_3 = BHD: resp (1.0)
first_3 = MUS: resp (4.0/2.0)
first_3 = BUS: sol-inf (3.0/2.0)
first_3 = MUU: sol-inf (4.0/2.0)
first_3 = BUH: dir-accion (4.0/2.0)
first_3 = MSM: resp (1.0)
first_3 = MBS: dir-accion (5.0/3.0)
first_3 = MDU: sol-inf (2.0)
first_3 = MMM: dir-accion (1.0)

```

Number of Leaves : 194

Size of the tree : 221

Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	193	37.4031 %
Incorrectly Classified Instances	323	62.5969 %
Kappa statistic	0.1571	
Mean absolute error	0.1526	
Root mean squared error	0.3087	
Relative absolute error	89.946 %	
Root relative squared error	106.1361 %	
Total Number of Instances	516	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
0.184	0.109	0.226	0.184	0.203	sol-inf
0.588	0.296	0.483	0.588	0.53	resp
0.45	0.24	0.362	0.45	0.401	dir-accion
0.222	0.187	0.277	0.222	0.247	x
0	0.004	0	0	0	dir-accion,_resp
0	0	0	0	0	sol-inf,_resp
0	0.002	0	0	0	oferta
0	0	0	0	0	sol-inf,_oferta
0	0	0	0	0	dir-accion,_oferta

=== Confusion Matrix ===

a b c d e f g h i <-- classified as

```
14 18 33 11 0 0 0 0 0 | a = sol-inf
12 97 23 33 0 0 0 0 0 | b = resp
18 25 54 22 0 0 1 0 0 | c = dir-accion
15 50 31 28 2 0 0 0 0 | d = x
  1  6  5  4  0  0  0  0  0 | e = dir-accion,_resp
  2  2  2  0  0  0  0  0  0 | f = sol-inf,_resp
  0  2  0  3  0  0  0  0  0 | g = oferta
  0  1  0  0  0  0  0  0  0 | h = sol-inf,_oferta
  0  0  1  0  0  0  0  0  0 | i = dir-accion,_oferta
```

Modelo 1.1.B
Target: commgr

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:
univ_data_one_int_firstintsint_commgr_with_optimal_predmood_numoftones_
USER_21dic2006-weka.filters.unsupervised.attribute.Remove-R1-4,8-10,12-
17,19-20
Instances: 516
Attributes: 6
first_1
first_2
first_3
last_2
utt_duration
common_ground
Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree

first_3 = MST: x (40.0/17.0)
first_3 = BST: acepta (8.0/4.0)
first_3 = MBU
| utt_duration <= 1304.3125: acepta (5.0)
| utt_duration > 1304.3125
| | utt_duration <= 2629.5: otra (2.0/1.0)
| | utt_duration > 2629.5: x (3.0/1.0)
first_3 = BSU
| last_2 = DB: acepta (0.0)
| last_2 = LH: acepta (0.0)
| last_2 = UT: acepta (5.0/1.0)
| last_2 = LT: acepta (1.0)
| last_2 = BS: acepta (0.0)
| last_2 = TS: x (2.0/1.0)
| last_2 = LS: acepta (0.0)
| last_2 = ST: acepta (0.0)
| last_2 = DS: afirm (1.0)
| last_2 = BT: acepta (0.0)
| last_2 = TL: rechaza (1.0)
| last_2 = US: x (1.0)
| last_2 = HL: x (1.0)
| last_2 = BH: acepta (0.0)
| last_2 = MB: acepta (0.0)
| last_2 = SH: acepta (0.0)
| last_2 = HS: afirm (1.0)
| last_2 = TB: acepta (0.0)
| last_2 = SB: acepta (0.0)
| last_2 = UH: acepta (0.0)
| last_2 = DL: afirm,_espera (1.0)
| last_2 = BU: acepta (0.0)
| last_2 = HD: acepta (0.0)
| last_2 = LU: acepta (0.0)

```
| last_2 = HB: acepta (0.0)
| last_2 = UM: acepta (0.0)
| last_2 = DD: acepta (0.0)
| last_2 = SL: acepta (0.0)
| last_2 = DM: acepta (0.0)
| last_2 = DU: acepta (0.0)
| last_2 = SU: acepta (0.0)
first_3 = MTB: acepta (54.0/24.0)
first_3 = MTL: x (14.0/7.0)
first_3 = MSH: x (38.0/15.0)
first_3 = MSB
| last_2 = DB: x (0.0)
| last_2 = LH: x (0.0)
| last_2 = UT: x (3.0/1.0)
| last_2 = LT: x (0.0)
| last_2 = BS: acepta (1.0)
| last_2 = TS: x (1.0)
| last_2 = LS: quiza,_afirm (1.0)
| last_2 = ST: acepta (1.0)
| last_2 = DS: afirm (1.0)
| last_2 = BT: acepta (5.0/2.0)
| last_2 = TL: x (0.0)
| last_2 = US: x (1.0)
| last_2 = HL: x (1.0)
| last_2 = BH: x (0.0)
| last_2 = MB: x (0.0)
| last_2 = SH: x (0.0)
| last_2 = HS: x (0.0)
| last_2 = TB: x (0.0)
| last_2 = SB: x (0.0)
| last_2 = UH: x (0.0)
| last_2 = DL: x (0.0)
| last_2 = BU: x (0.0)
| last_2 = HD: x (0.0)
| last_2 = LU: x (0.0)
| last_2 = HB: x (0.0)
| last_2 = UM: x (0.0)
| last_2 = DD: x (0.0)
| last_2 = SL: x (0.0)
| last_2 = DM: x (0.0)
| last_2 = DU: x (0.0)
| last_2 = SU: x (0.0)
first_3 = MHD: x (6.0/1.0)
first_3 = MMS
| utt_duration <= 4183.5: acepta,_afirm (2.0/1.0)
| utt_duration > 4183.5: x (3.0)
first_3 = MBT: acepta (27.0/9.0)
first_3 = MSU: x (13.0/7.0)
first_3 = BT: acepta (6.0/1.0)
first_3 = MUT: x (9.0/3.0)
first_3 = MB: acepta (33.0/7.0)
first_3 = MHS
| utt_duration <= 2201.75: acepta (2.0/1.0)
| utt_duration > 2201.75
| | utt_duration <= 5302.375: x (3.0/1.0)
| | utt_duration > 5302.375: acepta (2.0/1.0)
first_3 = BSH: acepta (3.0/1.0)
```

```

first_3 = MSL: x (21.0/12.0)
first_3 = MTS: x (9.0/6.0)
first_3 = MHL: x (12.0/8.0)
first_3 = MML: x (1.0)
first_3 = MLS
|   utt_duration <= 5275.1875: x (4.0/1.0)
|   utt_duration > 5275.1875: afirm (3.0/1.0)
first_3 = BHL: acepta (7.0/2.0)
first_3 = MMT: x (4.0/1.0)
first_3 = MDS: x (7.0/4.0)
first_3 = MLH
|   utt_duration <= 2241.25: acepta (2.0)
|   utt_duration > 2241.25: x (5.0/3.0)
first_3 = MLU
|   utt_duration <= 3414.8125: afirm (2.0/1.0)
|   utt_duration > 3414.8125: x (2.0)
first_3 = MHB: x (4.0/2.0)
first_3 = MDB: x (8.0/3.0)
first_3 = MBH: x (1.0)
first_3 = MDH: x (2.0/1.0)
first_3 = MLT: acepta (20.0/12.0)
first_3 = BUU: acepta (3.0/2.0)
first_3 = BTD
|   utt_duration <= 1270.1875: acepta (3.0/1.0)
|   utt_duration > 1270.1875: x (2.0/1.0)
first_3 = MUH: x (4.0/2.0)
first_3 = MSD
|   utt_duration <= 1347.6875: acepta (2.0/1.0)
|   utt_duration > 1347.6875
|   |   utt_duration <= 7111.6875: x (4.0/1.0)
|   |   utt_duration > 7111.6875: acepta (2.0)
first_3 = MDL: x (3.0/2.0)
first_3 = MTD: acepta (29.0/18.0)
first_3 = BUT
|   utt_duration <= 2162.1875: x (4.0/1.0)
|   utt_duration > 2162.1875: acepta (2.0/1.0)
first_3 = BTS: x (3.0/2.0)
first_3 = BHS: acepta (3.0/2.0)
first_3 = BTL
|   utt_duration <= 776.125: acepta (5.0)
|   utt_duration > 776.125: x (2.0)
first_3 = MDD
|   utt_duration <= 3835: afirm (2.0/1.0)
|   utt_duration > 3835: x (2.0)
first_3 = BHD: afirm (1.0)
first_3 = MUS
|   utt_duration <= 2682.6875: x (2.0/1.0)
|   utt_duration > 2682.6875: acepta (2.0/1.0)
first_3 = BUS: x (3.0/2.0)
first_3 = MUU: x (4.0)
first_3 = BUH: x (4.0/2.0)
first_3 = MSM: afirm (1.0)
first_3 = MBS: acepta-parte (5.0/3.0)
first_3 = MDU: x (2.0/1.0)
first_3 = MMM: x (1.0)

```

Number of Leaves : 127

Size of the tree : 145

Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	239	46.3178 %
Incorrectly Classified Instances	277	53.6822 %
Kappa statistic	0.1817	
Mean absolute error	0.0587	
Root mean squared error	0.185	
Relative absolute error	90.9273 %	
Root relative squared error	103.686 %	
Total Number of Instances	516	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
0.677	0.47	0.448	0.677	0.54	x
0.619	0.299	0.528	0.619	0.57	acepta
0.018	0.033	0.063	0.018	0.028	afirm
0	0	0	0	0	afirm,_acepta-
0	0	0	0	0	parte,_exclama
0	0	0	0	0	quiza,_afirm
0	0	0	0	0	repet-refr
0	0	0	0	0	exclama,_acepta,_afirm
0	0.004	0	0	0	ack
0	0	0	0	0	perform
0	0	0	0	0	acepta-parte
0	0.004	0	0	0	acepta,_afirm
0	0	0	0	0	reafirm
0	0	0	0	0	afirm,_rechaza
0	0.006	0	0	0	rechaza
0	0	0	0	0	rechaza-parte
0	0	0	0	0	afirm,_acepta
0	0	0	0	0	espera
0	0	0	0	0	espera,_repet-refr
0	0	0	0	0	otra
0	0	0	0	0	reafirm,_complem
0	0	0	0	0	conv-abre
0	0	0	0	0	afirm,_espera
0	0	0	0	0	conv-cierra

=== Confusion Matrix ===

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q
r	126	47	9	0	0	0	0	1	0	0	2	0	0	1	0	0	0
s	0	0	0	0	0	0	0	1	0	0	0	0	0	2	0	0	0
t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
u	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
v	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
w	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

<-- classified as

a = x

b = acepta

c = afirm

Modelo 1.2.A
Target: obligations

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: universal_data_one_int_x_23nov2006-
weka.filters.unsupervised.attribute.Remove-R1-8,11,13
Instances: 1043
Attributes: 3
utterance_mood
utt_duration
obligations
Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree

```
-----  
utterance_mood = int: sol-inf (291.0/63.0)  
utterance_mood = dec  
| utt_duration <= 1313.75  
| | utt_duration <= 865.1875  
| | | utt_duration <= 474.625  
| | | | utt_duration <= 470.875  
| | | | | utt_duration <= 357.625  
| | | | | | utt_duration <= 310.125  
| | | | | | | utt_duration <= 272.125: resp (6.0/3.0)  
| | | | | | | | utt_duration > 272.125  
| | | | | | | | | utt_duration <= 300.625: compromiso  
(5.0/1.0)  
| | | | | | | | | | utt_duration > 300.625: x (7.0/2.0)  
| | | | | | | | | | utt_duration > 310.125: compromiso (25.0/9.0)  
| | | | | | | | | | utt_duration > 357.625: resp (68.0/42.0)  
| | | | | | | | | | utt_duration > 470.875: compromiso (5.0)  
| | | | | | | | | | utt_duration > 474.625: x (206.0/119.0)  
| | | | | | | | | | utt_duration > 865.1875  
| | | | | | | | | | utt_duration <= 985.5  
| | | | | | | | | | utt_duration <= 898.3125: resp (9.0/3.0)  
| | | | | | | | | | utt_duration > 898.3125: x (15.0/4.0)  
| | | | | | | | | | utt_duration > 985.5: resp (57.0/29.0)  
| | | | | | | | | | utt_duration > 1313.75  
| | | | | | | | | | utt_duration <= 6661.6875  
| | | | | | | | | | utt_duration <= 1333.5: resp (4.0/2.0)  
| | | | | | | | | | utt_duration > 1333.5: x (241.0/165.0)  
| | | | | | | | | | utt_duration > 6661.6875  
| | | | | | | | | | utt_duration <= 10076.75: dir-accion (28.0/9.0)  
| | | | | | | | | | utt_duration > 10076.75  
| | | | | | | | | | utt_duration <= 11408.1875: resp (4.0/1.0)  
| | | | | | | | | | utt_duration > 11408.1875: dir-accion (12.0/6.0)  
utterance_mood = otra: x (46.0/18.0)  
utterance_mood = imp  
| utt_duration <= 2965.375  
| | utt_duration <= 1096.5625: dir-accion (2.0/1.0)  
| | utt_duration > 1096.5625: sol-inf (6.0/1.0)
```

| utt_duration > 2965.375: dir-accion (6.0/2.0)

Number of Leaves : 20

Size of the tree : 37

Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	500	47.9386 %
Incorrectly Classified Instances	543	52.0614 %
Kappa statistic	0.3264	
Mean absolute error	0.1201	
Root mean squared error	0.2529	
Relative absolute error	75.8353 %	
Root relative squared error	89.9409 %	
Total Number of Instances	1043	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
0.846	0.094	0.762	0.846	0.802	sol-inf
0.307	0.199	0.297	0.307	0.302	resp
0.267	0.052	0.4	0.267	0.32	dir-accion
0.295	0.062	0.363	0.295	0.325	compromiso
0.493	0.263	0.401	0.493	0.442	x
0	0	0	0	0	dir-accion,_resp
0	0	0	0	0	sol-inf,_resp
0	0	0	0	0	oferta
0	0	0	0	0	sol-inf,_oferta
0	0	0	0	0	dir-accion,_oferta

=== Confusion Matrix ===

a	b	c	d	e	f	g	h	i	j	<-- classified as
231	19	7	0	16	0	0	0	0	0	a = sol-inf
6	69	22	32	96	0	0	0	0	0	b = resp
25	30	32	0	33	0	0	0	0	0	c = dir-accion
0	30	0	33	49	0	0	0	0	0	d = compromiso
25	74	16	24	135	0	0	0	0	0	e = x
3	5	2	0	6	0	0	0	0	0	f = dir-accion,_resp
13	2	0	0	1	0	0	0	0	0	g = sol-inf,_resp
0	3	0	1	1	0	0	0	0	0	h = oferta
0	0	0	1	0	0	0	0	0	0	i = sol-inf,_oferta
0	0	1	0	0	0	0	0	0	0	j = dir-accion,_oferta

0	0	0	0	0	espera_en_admon.-
tarea					
0	0	0	0	0	perform
0.05	0.003	0.25	0.05	0.083	acepta-parte
0	0.001	0	0	0	reafirm
0	0	0	0	0	afirm,_rechaza
0	0	0	0	0	espera
0	0.001	0	0	0	oferta,_conv-abre
0	0	0	0	0	rechaza
0	0	0	0	0	quiza
0	0	0	0	0	reafirm,_espera
0	0	0	0	0	opcion-
abierta,_acepta					
0	0	0	0	0	rechaza-parte
0	0	0	0	0	oferta
0	0	0	0	0	otra
0	0	0	0	0	reafirm,_complem
0	0	0	0	0	conv-abre
0	0	0	0	0	afirm,_correccion
0	0	0	0	0	afirm,_espera
0	0	0	0	0	opcion-
abierta,_rechaza					
0	0	0	0	0	
afirm,_perform,_conv-cierra					
0	0	0	0	0	espera,_SNE
0	0	0	0	0	afirm,_conv-cierra

=== Confusion Matrix ===

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	
r	s	t	u	v	w	x	y	z	aa	ab	ac	ad	ae	af	ag	ah	ai	
aj	ak	al	am	an	ao	<-- classified as												
	249	0	26	15	17	0	0	1	0	5	2	0	0	0	0	0	0	
	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	
	0	0	0	0	0	0	a = x										0	0
	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	0	0	0	0	0	0	b = acepta,_espera,_repet-refr										0	0
	63	0	311	2	6	0	0	0	0	0	3	0	0	0	0	0	0	
	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
	0	0	0	0	0	0	c = acepta										0	0
	38	0	4	11	0	0	0	1	0	0	0	0	0	0	0	0	0	
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	0	0	0	0	0	0	d = espera,_repet-refr										0	0
	42	0	18	0	10	0	0	0	0	1	5	0	0	0	0	0	0	
	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	
	0	0	0	0	0	0	e = afirm										0	0
	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	0	0	0	0	0	0	f = afirm,_acepta-parte,_exclama										0	0
	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	0	0	0	0	0	0	g = quiza,_afirm										0	0
	5	0	0	1	1	0	0	4	0	0	0	0	0	0	0	0	0	
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	0	0	0	0	0	0	h = SNE										0	0


```

0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | ab = quiza
0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | ac = reafirm,_espera
1 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | ad = opcion-abierta,_acepta
1 0 3 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | ae = rechaza-parte
1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | af = oferta
1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | ag = otra
0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | ah = reafirm,_complem
0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | ai = conv-abre
1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | aj = afirm,_correccion
0 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | ak = afirm,_espera
1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | al = opcion-abierta,_rechaza
0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | am = afirm,_perform,_conv-cierra
1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | an = espera,_SNE
0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | ao = afirm,_conv-cierra

```

Modelo 1.3

Target: *predicted_mood*

=== Run information ===

```
Scheme:          weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:        universal_data_one_int_firstintsint_agreem_und_09dic2006-
weka.filters.unsupervised.attribute.Remove-R1-3,9,11-13,16-19-
weka.filters.unsupervised.attribute.Remove-R5-
weka.filters.unsupervised.attribute.Remove-R4
Instances:       1043
Attributes:      6
                 first_1
                 first_2
                 first_3
                 last_2
                 utterance_mood
                 utt_duration
Test mode:       10-fold cross-validation
```

=== Classifier model (full training set) ===

J48 pruned tree

```
last_2 = DB: dec (107.0/17.0)
last_2 = LH: dec (8.0/3.0)
last_2 = TS: int (94.0/37.0)
last_2 = UT: int (98.0/34.0)
last_2 = LT: int (22.0/5.0)
last_2 = BS: dec (136.0/16.0)
last_2 = LS: dec (31.0/9.0)
last_2 = MB: dec (87.0/2.0)
last_2 = ST: dec (20.0/9.0)
last_2 = DS: dec (20.0/2.0)
last_2 = TB: dec (90.0/17.0)
last_2 = BT
|   first_2 = MS
|   |   utt_duration <= 737.75: dec (11.0)
|   |   utt_duration > 737.75
|   |   |   utt_duration <= 786.25: int (2.0)
|   |   |   utt_duration > 786.25: dec (9.0/2.0)
|   first_2 = BS: dec (0.0)
|   first_2 = BU: dec (0.0)
|   first_2 = MB: dec (32.0/12.0)
|   first_2 = MT: int (1.0)
|   first_2 = MH: int (9.0)
|   first_2 = MM: dec (0.0)
|   first_2 = BT: dec (9.0/3.0)
|   first_2 = MU: int (1.0)
|   first_2 = ML: otra (3.0/1.0)
|   first_2 = BH: dec (0.0)
|   first_2 = MD: int (11.0/5.0)
last_2 = TL: dec (35.0/14.0)
last_2 = US
|   utt_duration <= 1089.625: int (21.0/1.0)
```

```

|   utt_duration > 1089.625: dec (43.0/10.0)
last_2 = HL: dec (10.0/4.0)
last_2 = BH: dec (36.0/16.0)
last_2 = SH: dec (7.0/3.0)
last_2 = HB: dec (14.0/2.0)
last_2 = SB: dec (28.0/7.0)
last_2 = HS: dec (17.0/6.0)
last_2 = UH: dec (3.0/1.0)
last_2 = DL: dec (10.0)
last_2 = SL: dec (6.0/2.0)
last_2 = BU: imp (1.0)
last_2 = HD: int (2.0/1.0)
last_2 = LU: int (1.0)
last_2 = UM: dec (1.0)
last_2 = DD: dec (4.0/1.0)
last_2 = DM: int (1.0)
last_2 = DU: dec (1.0)
last_2 = SU: dec (1.0)

```

Number of Leaves : 45

Size of the tree : 50

Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	775	74.3049 %
Incorrectly Classified Instances	268	25.6951 %
Kappa statistic	0.4164	
Mean absolute error	0.1809	
Root mean squared error	0.3146	
Relative absolute error	75.2207 %	
Root relative squared error	90.8313 %	
Total Number of Instances	1043	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
0.598	0.134	0.633	0.598	0.615	int
0.868	0.464	0.787	0.868	0.826	dec
0	0.003	0	0	0	otra
0	0.001	0	0	0	imp

=== Confusion Matrix ===

a	b	c	d	<-- classified as
174	115	2	0	a = int
89	601	1	1	b = dec
12	34	0	0	c = otra
0	14	0	0	d = imp

Correctly Classified Instances	407	39.0221 %
Incorrectly Classified Instances	636	60.9779 %
Kappa statistic	0.2179	
Mean absolute error	0.1396	
Root mean squared error	0.2752	
Relative absolute error	88.1499 %	
Root relative squared error	97.8737 %	
Total Number of Instances	1043	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
0.601	0.164	0.566	0.601	0.583	sol-inf
0.187	0.134	0.276	0.187	0.223	resp
0.475	0.143	0.302	0.475	0.369	dir-accion
0.196	0.032	0.423	0.196	0.268	compromiso
0.445	0.306	0.342	0.445	0.387	x
0	0.003	0	0	0	dir-accion,_resp
0	0	0	0	0	sol-inf,_resp
0	0	0	0	0	oferta
0	0	0	0	0	sol-inf,_oferta
0	0	0	0	0	dir-accion,_oferta

=== Confusion Matrix ===

a	b	c	d	e	f	g	h	i	j	<-- classified as
164	21	37	3	47	1	0	0	0	0	a = sol-inf
37	42	43	15	87	1	0	0	0	0	b = resp
21	11	57	1	30	0	0	0	0	0	c = dir-accion
10	25	1	22	54	0	0	0	0	0	d = compromiso
46	48	46	11	122	1	0	0	0	0	e = x
2	1	3	0	10	0	0	0	0	0	f = dir-accion,_resp
10	1	1	0	4	0	0	0	0	0	g = sol-inf,_resp
0	2	0	0	3	0	0	0	0	0	h = oferta
0	1	0	0	0	0	0	0	0	0	i = sol-inf,_oferta
0	0	1	0	0	0	0	0	0	0	j = dir-accion,_oferta

0	0.002	0	0	0	acepta-parte
0	0.001	0	0	0	reafirm
0	0	0	0	0	afirm,_rechaza
0	0.002	0	0	0	espera
0	0	0	0	0	oferta,_conv-abre
0	0	0	0	0	rechaza
0	0	0	0	0	quiza
0	0	0	0	0	reafirm,_espera
0	0	0	0	0	opcion-
abierta,_acepta					
0	0	0	0	0	rechaza-parte
0	0	0	0	0	oferta
0	0	0	0	0	otra
0	0	0	0	0	reafirm,_complem
0	0	0	0	0	conv-abre
0	0	0	0	0	afirm,_correccion
0	0	0	0	0	afirm,_espera
0	0	0	0	0	opcion-
abierta,_rechaza					
0	0	0	0	0	
afirm,_perform,_conv-cierra					
0	0	0	0	0	espera,_SNE
0	0	0	0	0	afirm,_conv-cierra

=== Confusion Matrix ===

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q
r	s	t	u	v	w	x	y	z	aa	ab	ac	ad	ae	af	ag	ah	ai
aj	ak	al	am	an	ao	<-- classified as											
	236	1	54	11	13	0	0	0	0	1	0	0	0	0	0	0	0
	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	a = x										
	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	b = acepta,_espera,_repet-refr										
	92	0	280	4	7	0	0	0	1	0	1	0	0	0	0	0	0
	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	c = acepta										
	37	0	8	4	3	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	d = espera,_repet-refr										
	54	0	13	1	8	0	0	0	0	0	1	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	e = afirm										
	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	f = afirm,_acepta-parte,_exclama										
	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	g = quiza,_afirm										
	2	0	8	0	1	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	h = SNE										
	3	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	i = repet-refr										


```

1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | ac = reaffirm,_espera
4 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | ad = opcion-abierta,_acepta
2 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | ae = rechaza-parte
0 0 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | af = oferta
1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | ag = otra
0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | ah = reaffirm,_complem
0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | ai = conv-abre
1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | aj = afirm,_correccion
2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | ak = afirm,_espera
1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | al = opcion-abierta,_rechaza
0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | am = afirm,_perform,_conv-cierra
0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | an = espera,_SNE
0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | ao = afirm,_conv-cierra

```

Apéndice XIV. Árboles generados con WEKA para la configuración 2

Modelo 2.1

Target: *optimal_pred_mood*

=== Run information ===

```
Scheme:          weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:
universal_data_one_int_firstintsint_agreem_und_with_predmood_numberofto
nes_09dic2006-weka.filters.unsupervised.attribute.Remove-R1-4,8-10,12-
14,16,19-21
Instances:      1043
Attributes:     7
                first_1
                first_2
                first_3
                last_2
                utterance_mood
                utt_duration
                speaker_role
Test mode:     10-fold cross-validation
```

=== Classifier model (full training set) ===

J48 pruned tree

```
-----
last_2 = DB: dec (107.0/17.0)
last_2 = LH: dec (8.0/3.0)
last_2 = TS
| speaker_role = u: dec (36.0/11.0)
| speaker_role = s: int (58.0/9.0)
last_2 = UT
| speaker_role = u: dec (45.0/18.0)
| speaker_role = s
| | utt_duration <= 659.3125
| | | utt_duration <= 520.5625: int (4.0/1.0)
| | | utt_duration > 520.5625: dec (2.0)
| | utt_duration > 659.3125: int (47.0)
last_2 = LT
| speaker_role = u: dec (7.0/3.0)
| speaker_role = s: int (15.0/1.0)
last_2 = BS: dec (136.0/16.0)
last_2 = LS: dec (31.0/9.0)
last_2 = MB: dec (87.0/2.0)
last_2 = ST
| speaker_role = u: dec (11.0/1.0)
| speaker_role = s: int (9.0/1.0)
last_2 = DS: dec (20.0/2.0)
last_2 = TB: dec (90.0/17.0)
last_2 = BT
| first_2 = MS
| | utt_duration <= 737.75: dec (11.0)
```

```

| | utt_duration > 737.75
| | | utt_duration <= 786.25: int (2.0)
| | | utt_duration > 786.25: dec (9.0/2.0)
| first_2 = BS: dec (0.0)
| first_2 = BU: dec (0.0)
| first_2 = MB: dec (32.0/12.0)
| first_2 = MT: int (1.0)
| first_2 = MH: int (9.0)
| first_2 = MM: dec (0.0)
| first_2 = BT
| | speaker_role = u: dec (6.0/1.0)
| | speaker_role = s: int (3.0/1.0)
| first_2 = MU: int (1.0)
| first_2 = ML: otra (3.0/1.0)
| first_2 = BH: dec (0.0)
| first_2 = MD
| | speaker_role = u: dec (5.0/2.0)
| | speaker_role = s: int (6.0/1.0)
last_2 = TL: dec (35.0/14.0)
last_2 = US
| speaker_role = u: dec (34.0/6.0)
| speaker_role = s
| | utt_duration <= 1178.8125: int (22.0/1.0)
| | utt_duration > 1178.8125: dec (8.0/3.0)
last_2 = HL: dec (10.0/4.0)
last_2 = BH: dec (36.0/16.0)
last_2 = SH: dec (7.0/3.0)
last_2 = HB: dec (14.0/2.0)
last_2 = SB: dec (28.0/7.0)
last_2 = HS
| speaker_role = u: dec (11.0/2.0)
| speaker_role = s: int (6.0/2.0)
last_2 = UH: dec (3.0/1.0)
last_2 = DL: dec (10.0)
last_2 = SL: dec (6.0/2.0)
last_2 = BU: imp (1.0)
last_2 = HD: int (2.0/1.0)
last_2 = LU: int (1.0)
last_2 = UM: dec (1.0)
last_2 = DD: dec (4.0/1.0)
last_2 = DM: int (1.0)
last_2 = DU: dec (1.0)
last_2 = SU: dec (1.0)

```

Number of Leaves : 55

Size of the tree : 70

Time taken to build model: 0.19 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	807	77.373	%
Incorrectly Classified Instances	236	22.627	%
Kappa statistic	0.4531		

```
Mean absolute error          0.162
Root mean squared error      0.3021
Relative absolute error      67.3734 %
Root relative squared error   87.2125 %
Total Number of Instances    1043
```

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
0.543	0.06	0.778	0.543	0.64	int
0.938	0.533	0.776	0.938	0.849	dec
0	0.003	0	0	0	otra
0	0.001	0	0	0	imp

=== Confusion Matrix ===

	a	b	c	d	<-- classified as
158	131	2	0	0	a = int
41	649	1	1	0	b = dec
4	42	0	0	0	c = otra
0	14	0	0	0	d = imp

Modelo 2.2.A
Target: obligations

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:
univ_data_one_int_firstintsint_commgr_with_optimal_predmood_numoftones_12dic2006-weka.filters.unsupervised.attribute.Remove-R1-16
Instances: 1043
Attributes: 5
optimal_pred_mood
utt_duration
speaker_role
obligations
common_ground
Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree

```
-----  
common_ground = x  
| speaker_role = u  
| | utt_duration <= 1328.6875: x (31.0/11.0)  
| | utt_duration > 1328.6875: dir-accion (155.0/77.0)  
| speaker_role = s  
| | utt_duration <= 509.75: compromiso (4.0/1.0)  
| | utt_duration > 509.75: sol-inf (127.0/13.0)  
common_ground = acepta,_espera,_repet-refr: sol-inf (2.0)  
common_ground = acepta  
| speaker_role = u  
| | utt_duration <= 2218.4375: resp (146.0/33.0)  
| | utt_duration > 2218.4375  
| | | utt_duration <= 3095.3125  
| | | | utt_duration <= 2503.5625: sol-inf (3.0/1.0)  
| | | | utt_duration > 2503.5625  
| | | | utt_duration <= 2769.5: dir-accion (2.0)  
| | | | utt_duration > 2769.5: sol-inf (2.0)  
| | | utt_duration > 3095.3125: dir-accion (28.0/5.0)  
| speaker_role = s  
| | optimal_pred_mood = dec: compromiso (182.0/79.0)  
| | optimal_pred_mood = int  
| | | utt_duration <= 968.125: compromiso (12.0/6.0)  
| | | utt_duration > 968.125: sol-inf (11.0/1.0)  
| | optimal_pred_mood = imp: compromiso (0.0)  
| | optimal_pred_mood = otra: compromiso (0.0)  
common_ground = espera,_repet-refr: sol-inf (54.0/1.0)  
common_ground = afirm  
| speaker_role = u  
| | utt_duration <= 6006.8125: resp (51.0/25.0)  
| | utt_duration > 6006.8125: x (5.0/1.0)  
| speaker_role = s: x (21.0/10.0)  
common_ground = afirm,_acepta-parte,_exclama: resp (1.0)  
common_ground = quiza,_afirm: x (2.0)
```

```

common_ground = SNE: sol-inf (11.0)
common_ground = repet-refr: x (5.0)
common_ground = afirm,_acepta
|   utt_duration <= 1474.5625: x (2.0/1.0)
|   utt_duration > 1474.5625
|   |   utt_duration <= 1823.6875: resp (2.0)
|   |   utt_duration > 1823.6875: x (2.0)
common_ground = opcion-abierta
|   optimal_pred_mood = dec: resp (43.0/12.0)
|   optimal_pred_mood = int: x (4.0/1.0)
|   optimal_pred_mood = imp: resp (0.0)
|   optimal_pred_mood = otra: resp (0.0)
common_ground = oferta,_acepta: x (2.0)
common_ground = exclama,_acepta,_afirm: x (1.0)
common_ground = acepta,_afirm: x (3.0)
common_ground = sne: sol-inf (1.0)
common_ground = ack: x (24.0)
common_ground = conv-cierra: x (5.0)
common_ground = convenc-abre,_oferta: x (1.0)
common_ground = back-channel: x (3.0)
common_ground = espera_en_admon.-tarea: x (1.0)
common_ground = perform: x (4.0)
common_ground = acepta-parte
|   utt_duration <= 1219.3125: dir-accion (4.0/1.0)
|   utt_duration > 1219.3125: dir-accion,_resp (16.0/3.0)
common_ground = reafirm
|   utt_duration <= 1474.5625: x (3.0/1.0)
|   utt_duration > 1474.5625: resp (13.0/3.0)
common_ground = afirm,_rechaza: resp (4.0)
common_ground = espera: sol-inf (13.0/4.0)
common_ground = oferta,_conv-abre: x (4.0)
common_ground = rechaza
|   utt_duration <= 2914.625: x (9.0/4.0)
|   utt_duration > 2914.625: dir-accion (2.0/1.0)
common_ground = quiza: x (1.0)
common_ground = reafirm,_espera: sol-inf (1.0)
common_ground = opcion-abierta,_acepta
|   utt_duration <= 2654.625: resp (2.0)
|   utt_duration > 2654.625: x (2.0)
common_ground = rechaza-parte: resp (4.0)
common_ground = oferta: x (2.0)
common_ground = otra: x (1.0)
common_ground = reafirm,_complem: resp (1.0)
common_ground = conv-abre: x (1.0)
common_ground = afirm,_correccion: x (1.0)
common_ground = afirm,_espera: x (2.0)
common_ground = opcion-abierta,_rechaza: x (1.0)
common_ground = afirm,_perform,_conv-cierra: x (1.0)
common_ground = espera,_SNE: sol-inf (1.0)
common_ground = afirm,_conv-cierra: resp (1.0)

```

Number of Leaves : 64

Size of the tree : 84

Time taken to build model: 0.3 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	691	66.2512 %
Incorrectly Classified Instances	352	33.7488 %
Kappa statistic	0.5812	
Mean absolute error	0.0845	
Root mean squared error	0.2129	
Relative absolute error	53.362 %	
Root relative squared error	75.7288 %	
Total Number of Instances	1043	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
0.725	0.04	0.865	0.725	0.789	sol-inf
0.751	0.097	0.681	0.751	0.715	resp
0.833	0.093	0.538	0.833	0.654	dir-accion
0.893	0.087	0.552	0.893	0.683	compromiso
0.405	0.094	0.607	0.405	0.486	x
0.813	0.003	0.813	0.813	0.813	dir-accion,_resp
0	0	0	0	0	sol-inf,_resp
0	0	0	0	0	oferta
0	0	0	0	0	sol-inf,_oferta
0	0	0	0	0	dir-accion,_oferta

=== Confusion Matrix ===

a	b	c	d	e	f	g	h	i	j	<-- classified as
198	9	50	2	14	0	0	0	0	0	a = sol-inf
5	169	4	9	38	0	0	0	0	0	b = resp
4	7	100	0	8	1	0	0	0	0	c = dir-accion
1	0	0	100	11	0	0	0	0	0	d = compromiso
11	55	28	69	111	0	0	0	0	0	e = x
0	1	2	0	0	13	0	0	0	0	f = dir-accion,_resp
10	1	1	1	1	2	0	0	0	0	g = sol-inf,_resp
0	5	0	0	0	0	0	0	0	0	h = oferta
0	1	0	0	0	0	0	0	0	0	i = sol-inf,_oferta
0	0	1	0	0	0	0	0	0	0	j = dir-accion,_oferta

Modelo 2.2.B
Target: commgr

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:
univ_data_one_int_firstintsint_commgr_with_optimal_predmood_numoftones_12dic2006-weka.filters.unsupervised.attribute.Remove-R1-16
Instances: 1043
Attributes: 5
 optimal_pred_mood
 utt_duration
 speaker_role
 obligations
 common_ground
Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree

```
obligations = sol-inf
|  speaker_role = u
|  |  utt_duration <= 1644.25
|  |  |  utt_duration <= 1588.875
|  |  |  |  utt_duration <= 1323.1875
|  |  |  |  |  utt_duration <= 1159.3125: x (7.0/1.0)
|  |  |  |  |  utt_duration > 1159.3125: acepta (4.0/1.0)
|  |  |  |  |  utt_duration > 1323.1875
|  |  |  |  |  utt_duration <= 1360.9375: espera,_repet-refr
(2.0/1.0)
|  |  |  |  |  utt_duration > 1360.9375: x (7.0/2.0)
|  |  |  |  |  utt_duration > 1588.875: acepta (3.0/1.0)
|  |  |  |  |  utt_duration > 1644.25: x (53.0/11.0)
|  speaker_role = s
|  |  utt_duration <= 1183.1875
|  |  |  utt_duration <= 516.375
|  |  |  |  utt_duration <= 492.25: SNE (9.0/3.0)
|  |  |  |  utt_duration > 492.25: acepta,_espera,_repet-refr
(2.0/1.0)
|  |  |  |  utt_duration > 516.375: x (113.0/22.0)
|  |  |  |  utt_duration > 1183.1875
|  |  |  |  optimal_pred_mood = dec
|  |  |  |  |  utt_duration <= 1766.6875: x (11.0/6.0)
|  |  |  |  |  utt_duration > 1766.6875: espera,_repet-refr (14.0/1.0)
|  |  |  |  |  optimal_pred_mood = int
|  |  |  |  |  |  utt_duration <= 3305.25: espera,_repet-refr (39.0/19.0)
|  |  |  |  |  |  utt_duration > 3305.25: x (7.0)
|  |  |  |  |  |  optimal_pred_mood = imp: espera,_repet-refr (0.0)
|  |  |  |  |  |  optimal_pred_mood = otra: espera,_repet-refr (2.0)
obligations = resp
|  utt_duration <= 1304.3125: acepta (135.0/20.0)
|  utt_duration > 1304.3125
```

```

| | speaker_role = u
| | | utt_duration <= 1828.375
| | | | utt_duration <= 1436.9375: x (3.0/2.0)
| | | | utt_duration > 1436.9375: acepta (13.0/5.0)
| | | | utt_duration > 1828.375: afirm (31.0/13.0)
| | speaker_role = s: opcion-abierta (43.0/13.0)
obligations = dir-accion: x (120.0/37.0)
obligations = compromiso: acepta (112.0/3.0)
obligations = x
| utt_duration <= 921.875
| | speaker_role = u
| | | utt_duration <= 712
| | | | utt_duration <= 448
| | | | | utt_duration <= 324.375: rechaza (3.0/1.0)
| | | | | utt_duration > 324.375: x (4.0/2.0)
| | | | | utt_duration > 448: acepta (21.0/10.0)
| | | | utt_duration > 712
| | | | | utt_duration <= 786.25: x (4.0/2.0)
| | | | | utt_duration > 786.25
| | | | | | utt_duration <= 810.9375: afirm (2.0)
| | | | | | utt_duration > 810.9375: x (8.0/5.0)
| | speaker_role = s: acepta (101.0/30.0)
| utt_duration > 921.875
| | speaker_role = u
| | | utt_duration <= 2069.125
| | | | utt_duration <= 1405.0625: x (18.0/7.0)
| | | | utt_duration > 1405.0625
| | | | | utt_duration <= 1588.875: afirm (7.0/1.0)
| | | | | utt_duration > 1588.875
| | | | | | utt_duration <= 1697.9375: x (3.0)
| | | | | | utt_duration > 1697.9375: afirm (9.0/5.0)
| | | | utt_duration > 2069.125
| | | | | utt_duration <= 6427.0625: x (41.0/19.0)
| | | | | utt_duration > 6427.0625: afirm (6.0/2.0)
| | speaker_role = s
| | | utt_duration <= 1579.125
| | | | optimal_pred_mood = dec: afirm (14.0/8.0)
| | | | optimal_pred_mood = int: opcion-abierta (2.0/1.0)
| | | | optimal_pred_mood = imp: afirm (0.0)
| | | | optimal_pred_mood = otra: afirm (0.0)
| | | | utt_duration >1579.125: opcion-abierta (31.0/18.0)
obligations = dir-accion,_resp: acepta-parte (16.0/3.0)
obligations = sol-inf,_resp
| speaker_role = u
| | utt_duration <= 3226.125: acepta-parte (4.0/2.0)
| | utt_duration > 3226.125: espera (2.0/1.0)
| speaker_role = s: x (10.0)
obligations = oferta: acepta (5.0)
obligations = sol-inf,_oferta: acepta (1.0)
obligations = dir-accion,_oferta: acepta (1.0)

```

Number of Leaves : 47

Size of the tree : 81

Time taken to build model: 0.17 seconds

=== Stratified cross-validation ===
 === Summary ===

Correctly Classified Instances	698	66.9223 %
Incorrectly Classified Instances	345	33.0777 %
Kappa statistic	0.5482	
Mean absolute error	0.0219	
Root mean squared error	0.1112	
Relative absolute error	58.3427 %	
Root relative squared error	81.679 %	
Total Number of Instances	1043	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
0.83	0.182	0.666	0.83	0.739	x
0	0	0	0	0	
acepta,_espera,_repet-refr					
0.811	0.128	0.788	0.811	0.799	acepta
0.593	0.025	0.561	0.593	0.577	espera,_repet-refr
0.403	0.045	0.419	0.403	0.411	afirm
0	0	0	0	0	afirm,_acepta-
parte,_exclama					
0	0	0	0	0	quiza,_afirm
0.455	0.004	0.556	0.455	0.5	SNE
0	0	0	0	0	repet-refr
0	0.001	0	0	0	afirm,_acepta
0.872	0.036	0.532	0.872	0.661	opcion-abierta
0	0	0	0	0	oferta,_acepta
0	0	0	0	0	
exclama,_acepta,_afirm					
0	0	0	0	0	acepta,_afirm
0	0	0	0	0	sne
0	0.004	0	0	0	ack
0	0.001	0	0	0	conv-cierra
0	0	0	0	0	convenc-
abre,_oferta					
0	0	0	0	0	back-channel
0	0	0	0	0	espera_en_admon.-
tarea					
0	0.003	0	0	0	perform
0.65	0.006	0.684	0.65	0.667	acepta-parte
0	0	0	0	0	reafirm
0	0	0	0	0	afirm,_rechaza
0	0.001	0	0	0	espera
0	0.003	0	0	0	oferta,_conv-abre
0	0.001	0	0	0	rechaza
0	0	0	0	0	quiza
0	0	0	0	0	reafirm,_espera
0	0	0	0	0	opcion-
abierta,_acepta					
0	0	0	0	0	rechaza-parte
0	0.001	0	0	0	oferta
0	0	0	0	0	otra
0	0	0	0	0	reafirm,_complem
0	0	0	0	0	conv-abre

```

0      0      0      0      0      0      afirm,_correccion
0      0      0      0      0      0      afirm,_espera
0      0      0      0      0      0      opcion-
abierta,_rechaza
0      0      0      0      0
afirm,_perform,_conv-cierra
0      0      0      0      0      0      espera,_SNE
0      0      0      0      0      0      afirm,_conv-cierra

```

=== Confusion Matrix ===

```

  a  b  c  d  e  f  g  h  i  j  k  l  m  n  o  p  q
r  s  t  u  v  w  x  y  z  aa ab ac ad ae af ag ah ai
aj ak al am an ao <-- classified as
263  0 17 16 13  0  0  0  0  0  1  0  0  0  0  1  0
0  0  0  2  3  0  0  1  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0 |  a = x
  0  0  0  1  0  0  0  1  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0 |  b = acepta,_espera,_repet-refr
 54  0 313  5  9  0  0  0  0  0  3  0  0  0  0  1  1
0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0 |  c = acepta
 19  0  0 32  1  0  0  2  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0 |  d = espera,_repet-refr
 14  0 14  0 31  0  0  0  0  0 16  0  0  0  0  0  0
0  0  0  1  1  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0 |  e = afirm
  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0 |  f = afirm,_acepta-parte,_exclama
  2  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0 |  g = quiza,_afirm
  4  0  0  2  0  0  0  5  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0 |  h = SNE
  1  0  3  0  0  0  0  0  0  0  1  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0 |  i = repet-refr
  1  0  1  0  2  0  0  0  0  0  1  0  0  0  0  0  0
0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0 |  j = afirm,_acepta
  1  0  2  0  1  0  0  0  0  0 41  0  0  0  0  0  0
0  0  0  0  0  0  0  0  2  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0 |  k = opcion-abierta
  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  1  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0 |  l = oferta,_acepta
  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0 |  m = exclama,_acepta,_afirm
  2  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0 |  n = acepta,_afirm

```

```

0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | o = sne
2 0 20 0 0 0 0 0 0 1 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0
0 0 0 0 0 0 | p = ack
2 0 3 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | q = conv-cierra
0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | r = convenc-abre,_oferta
0 0 3 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | s = back-channel
0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | t = espera_en_admon.-tarea
2 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | u = perform
4 0 2 0 1 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 13 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | v = acepta-parte
5 0 1 0 9 0 0 0 0 0 1 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | w = reafirm
0 0 3 0 0 0 0 0 0 0 0 1 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | x = afirm,_rechaza
10 0 1 0 2 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | y = espera
0 0 0 0 0 0 0 0 0 0 0 3 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
0 0 0 0 0 0 | z = oferta,_conv-abre
3 0 4 0 1 0 0 0 0 0 0 0 0 0 0 2 0
0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 | aa = rechaza
0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | ab = quiza
0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | ac = reafirm,_espera
0 0 1 0 0 0 0 0 0 0 0 3 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | ad = opcion-abierta,_acepta
0 0 3 0 1 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | ae = rechaza-parte
0 0 0 0 0 0 0 0 0 0 0 2 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | af = oferta
1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | ag = otra

```

```

1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | ah = reaffirm,_complem
1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | ai = conv-abre
0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | aj = afirm,_correccion
1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | ak = afirm,_espera
0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | al = opcion-abierta,_rechaza
0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | am = afirm,_perform,_conv-cierra
1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | an = espera,_SNE
0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 | ao = afirm,_conv-cierra

```

Apéndice XV. Árboles de decisión generados con WEKA para la configuración 3

Modelo 3.1.A

Target: *obligations*

=== Run information ===

```
Scheme:          weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:        sepln2007_universe_editado_con_nombres_cortos_18may2007-
weka.filters.unsupervised.attribute.Remove-R7
Instances:      1043
Attributes:     11
                 number_of_tones
                 first_1
                 first_2
                 first_3
                 last_2
                 optimal_pred_mood
                 speaker_role
                 obligations
                 obligations_minus1
                 commgr
                 commgr_minus1
Test mode:      10-fold cross-validation
```

=== Classifier model (full training set) ===

J48 pruned tree

```
commgr = x
| commgr_minus1 = x
| | first_3 = MST
| | | number_of_tones <= 7: x (5.0/1.0)
| | | number_of_tones > 7: sol-inf (3.0/1.0)
| | first_3 = BST: sol-inf (3.0/1.0)
| | first_3 = BUT: x (3.0/1.0)
| | first_3 = MBU: dir-accion (1.0)
| | first_3 = BSU: dir-accion (1.0)
| | first_3 = MTD: x (1.0)
| | first_3 = MTB
| | | obligations_minus1 = sol-inf: x (0.0)
| | | obligations_minus1 = x: dir-accion (2.0/1.0)
| | | obligations_minus1 = resp: x (0.0)
| | | obligations_minus1 = sol-inf_accion-graf: x (0.0)
| | | obligations_minus1 = dir-accion: x (2.0)
| | | obligations_minus1 = compromiso: x (0.0)
| | | obligations_minus1 = oferta_sol-inf: x (0.0)
| | | obligations_minus1 = dir-accion_resp: x (0.0)
| | | obligations_minus1 = nil: x (0.0)
| | | obligations_minus1 = despliega: x (0.0)
| | | obligations_minus1 = accion-graf: x (0.0)
| | | obligations_minus1 = sol-inf_resp: x (0.0)
| | | obligations_minus1 = sol-inf_accion-graf_resp: x (0.0)
```

```

| | | obligations_minus1 = oferta: x (0.0)
| | | obligations_minus1 = dir-accion_oferta: x (0.0)
| | | obligations_minus1 = despliega_resp: x (0.0)
| | first_3 = MTL: x (1.0)
| | first_3 = MSH: dir-accion (8.0/1.0)
| | first_3 = MSB: x (0.0)
| | first_3 = MB: x (0.0)
| | first_3 = MHD: sol-inf (3.0/1.0)
| | first_3 = MMS: dir-accion (1.0)
| | first_3 = MBT: sol-inf (2.0)
| | first_3 = MSU: dir-accion (2.0)
| | first_3 = BT: x (0.0)
| | first_3 = BTS: x (0.0)
| | first_3 = MUT: dir-accion (2.0/1.0)
| | first_3 = MHS: x (0.0)
| | first_3 = BSH: x (0.0)
| | first_3 = MHL: x (0.0)
| | first_3 = MSL: x (1.0)
| | first_3 = MTS: x (0.0)
| | first_3 = MBS: x (0.0)
| | first_3 = MML: x (1.0)
| | first_3 = MLS: x (1.0)
| | first_3 = MLT: sol-inf (1.0)
| | first_3 = BHL: x (0.0)
| | first_3 = MMT: dir-accion (1.0)
| | first_3 = MDS: x (0.0)
| | first_3 = MLH: x (1.0)
| | first_3 = BUU: x (0.0)
| | first_3 = MLU: x (0.0)
| | first_3 = MHB: x (0.0)
| | first_3 = MDB: x (4.0/1.0)
| | first_3 = MBH: dir-accion (1.0)
| | first_3 = MDH: x (0.0)
| | first_3 = BUH: x (0.0)
| | first_3 = BTB: x (0.0)
| | first_3 = BTL: x (1.0)
| | first_3 = MUH: x (0.0)
| | first_3 = MSD: sol-inf (1.0)
| | first_3 = MDL: x (1.0)
| | first_3 = BHS: x (0.0)
| | first_3 = MDD: x (0.0)
| | first_3 = BHD: x (0.0)
| | first_3 = MUS: x (0.0)
| | first_3 = BUS: x (0.0)
| | first_3 = MUU: sol-inf (3.0/2.0)
| | first_3 = MSM: x (0.0)
| | first_3 = MMD: x (0.0)
| | first_3 = MDU: x (0.0)
| | first_3 = MMM: x (0.0)
| commgr_minus1 = acepta_espera_repet-refr: dir-accion (0.0)
| commgr_minus1 = ack
| | number_of_tones <= 8: x (2.0)
| | number_of_tones > 8: dir-accion (2.0)
| commgr_minus1 = espera_repet-refr: sol-inf (2.0/1.0)
| commgr_minus1 = afirm: dir-accion (3.0/1.0)
| commgr_minus1 = accion-graf
| | optimal_pred_mood = dec: dir-accion (4.0/2.0)

```

```

| | optimal_pred_mood = int: sol-inf (2.0)
| | optimal_pred_mood = imp: sol-inf (0.0)
| | optimal_pred_mood = otra: sol-inf (0.0)
| commgr_minus1 = afirm_quiza: dir-accion (0.0)
| commgr_minus1 = exclama: dir-accion (0.0)
| commgr_minus1 = acepta
| | number_of_tones <= 15: sol-inf (91.0/52.0)
| | number_of_tones > 15: dir-accion (18.0/4.0)
| commgr_minus1 = repet-refr: dir-accion (2.0)
| commgr_minus1 = afirm_acepta: x (1.0)
| commgr_minus1 = opcion-abierta_despliega: sol-inf (2.0/1.0)
| commgr_minus1 = SNE: dir-accion_resp (1.0)
| commgr_minus1 = nil: dir-accion (0.0)
| commgr_minus1 = back-channel: sol-inf (2.0/1.0)
| commgr_minus1 = rechaza: dir-accion (10.0/3.0)
| commgr_minus1 = perform: sol-inf (1.0)
| commgr_minus1 = acepta-parte: dir-accion (2.0/1.0)
| commgr_minus1 = reafirm: sol-inf (2.0/1.0)
| commgr_minus1 = afirm_rechaza: dir-accion (0.0)
| commgr_minus1 = espera: dir-accion (0.0)
| commgr_minus1 = oferta_acepta: dir-accion (0.0)
| commgr_minus1 = afirm_acepta_conv-cierra: dir-accion (0.0)
| commgr_minus1 = accion-graf_acepta: x (1.0)
| commgr_minus1 = oferta_conv-abre: dir-accion (0.0)
| commgr_minus1 = quiza: dir-accion (0.0)
| commgr_minus1 = reafirm_espera: dir-accion (0.0)
| commgr_minus1 = afirm_despliega: dir-accion (0.0)
| commgr_minus1 = opcion-abierta: x (1.0)
| commgr_minus1 = conv-cierra: dir-accion (0.0)
| commgr_minus1 = rechaza-parte: dir-accion (1.0)
| commgr_minus1 = afirm_accion-graf: x (1.0)
| commgr_minus1 = opcion-abierta_despliega_acepta: sol-inf (1.0)
| commgr_minus1 = otra: dir-accion (0.0)
| commgr_minus1 = reafirm_complem: dir-accion (0.0)
| commgr_minus1 = afirm_correccion: dir-accion (0.0)
| commgr_minus1 = afirm_espera: sol-inf (1.0)
| commgr_minus1 = oferta: dir-accion (0.0)
| commgr_minus1 = afirm_conv-cierra: dir-accion (0.0)
| commgr_minus1 = espera_correccion: dir-accion (0.0)
| commgr_minus1 = espera_SNE: dir-accion (0.0)
commgr = acepta_espera_repet-refr: sol-inf (2.0)
commgr = acepta
| speaker_role = u
| | commgr_minus1 = x: resp (10.0/4.0)
| | commgr_minus1 = acepta_espera_repet-refr: resp (2.0)
| | commgr_minus1 = ack: dir-accion (4.0/2.0)
| | commgr_minus1 = espera_repet-refr: resp (43.0/3.0)
| | commgr_minus1 = afirm
| | | number_of_tones <= 12: x (3.0/1.0)
| | | number_of_tones > 12: dir-accion (3.0)
| | commgr_minus1 = accion-graf: resp (51.0)
| | commgr_minus1 = afirm_quiza: resp (0.0)
| | commgr_minus1 = exclama: resp (0.0)
| | commgr_minus1 = acepta
| | | obligations_minus1 = sol-inf: sol-inf (3.0/1.0)
| | | obligations_minus1 = x: dir-accion (4.0/1.0)
| | | obligations_minus1 = resp: resp (12.0)

```

```

| | | obligations_minus1 = sol-inf_accion-graf: resp (0.0)
| | | obligations_minus1 = dir-accion: dir-accion (7.0)
| | | obligations_minus1 = compromiso: resp (0.0)
| | | obligations_minus1 = oferta_sol-inf: resp (0.0)
| | | obligations_minus1 = dir-accion_resp: resp (0.0)
| | | obligations_minus1 = nil: resp (0.0)
| | | obligations_minus1 = despliega: resp (0.0)
| | | obligations_minus1 = accion-graf: resp (0.0)
| | | obligations_minus1 = sol-inf_resp: resp (0.0)
| | | obligations_minus1 = sol-inf_accion-graf_resp: resp (0.0)
| | | obligations_minus1 = oferta: resp (0.0)
| | | obligations_minus1 = dir-accion_oferta: resp (0.0)
| | | obligations_minus1 = despliega_resp: resp (0.0)
| | commgr_minus1 = repet-refr: resp (0.0)
| | commgr_minus1 = afirm_acepta: resp (0.0)
| | commgr_minus1 = opcion-abierta_despliega: dir-accion (18.0/8.0)
| | commgr_minus1 = SNE: resp (0.0)
| | commgr_minus1 = nil: resp (0.0)
| | commgr_minus1 = back-channel: dir-accion (1.0)
| | commgr_minus1 = rechaza: resp (0.0)
| | commgr_minus1 = perform: resp (0.0)
| | commgr_minus1 = acepta-parte: x (2.0)
| | commgr_minus1 = reafirm: x (1.0)
| | commgr_minus1 = afirm_rechaza: resp (0.0)
| | commgr_minus1 = espera: resp (0.0)
| | commgr_minus1 = oferta_acepta: sol-inf (1.0)
| | commgr_minus1 = afirm_acepta_conv-cierra: resp (0.0)
| | commgr_minus1 = accion-graf_acepta: resp (1.0)
| | commgr_minus1 = oferta_conv-abre: oferta (5.0/1.0)
| | commgr_minus1 = quiza: resp (0.0)
| | commgr_minus1 = reafirm_espera: resp (0.0)
| | commgr_minus1 = afirm_despliega: resp (0.0)
| | commgr_minus1 = opcion-abierta
| | | number_of_tones <= 5: x (2.0)
| | | number_of_tones > 5: sol-inf (2.0/1.0)
| | commgr_minus1 = conv-cierra: resp (0.0)
| | commgr_minus1 = rechaza-parte: resp (0.0)
| | commgr_minus1 = afirm_accion-graf: x (1.0)
| | commgr_minus1 = opcion-abierta_despliega_acepta: resp (0.0)
| | commgr_minus1 = otra: dir-accion (1.0)
| | commgr_minus1 = reafirm_complem: resp (0.0)
| | commgr_minus1 = afirm_correccion: sol-inf (1.0)
| | commgr_minus1 = afirm_espera: resp (0.0)
| | commgr_minus1 = oferta: oferta_sol-inf (2.0/1.0)
| | commgr_minus1 = afirm_conv-cierra: resp (0.0)
| | commgr_minus1 = espera_correccion: resp (1.0)
| | commgr_minus1 = espera_SNE: resp (0.0)
| speaker_role = s
| | obligations_minus1 = sol-inf: x (38.0/11.0)
| | obligations_minus1 = x: x (34.0/15.0)
| | obligations_minus1 = resp
| | | optimal_pred_mood = dec: compromiso (43.0/7.0)
| | | optimal_pred_mood = int: sol-inf (4.0/1.0)
| | | optimal_pred_mood = imp: compromiso (0.0)
| | | optimal_pred_mood = otra: compromiso (0.0)
| | obligations_minus1 = sol-inf_accion-graf: compromiso (0.0)
| | obligations_minus1 = dir-accion: compromiso (71.0/19.0)

```

```

| | obligations_minus1 = compromiso: compromiso (0.0)
| | obligations_minus1 = oferta_sol-inf: compromiso (0.0)
| | obligations_minus1 = dir-accion_resp: compromiso (9.0/1.0)
| | obligations_minus1 = nil: compromiso (0.0)
| | obligations_minus1 = despliega: compromiso (0.0)
| | obligations_minus1 = accion-graf: compromiso (0.0)
| | obligations_minus1 = sol-inf_resp: compromiso (2.0/1.0)
| | obligations_minus1 = sol-inf_accion-graf_resp: compromiso (0.0)
| | obligations_minus1 = oferta: compromiso (0.0)
| | obligations_minus1 = dir-accion_oferta: compromiso (1.0)
| | obligations_minus1 = despliega_resp: compromiso (0.0)
commgr = espera_repet-refr: sol-inf (54.0/1.0)
commgr = afirm
| obligations_minus1 = sol-inf: resp (30.0/4.0)
| obligations_minus1 = x: x (19.0/5.0)
| obligations_minus1 = resp: x (16.0/3.0)
| obligations_minus1 = sol-inf_accion-graf: x (0.0)
| obligations_minus1 = dir-accion: x (0.0)
| obligations_minus1 = compromiso: x (7.0/1.0)
| obligations_minus1 = oferta_sol-inf: x (0.0)
| obligations_minus1 = dir-accion_resp: x (0.0)
| obligations_minus1 = nil: x (0.0)
| obligations_minus1 = despliega: x (0.0)
| obligations_minus1 = accion-graf: x (0.0)
| obligations_minus1 = sol-inf_resp: x (1.0)
| obligations_minus1 = sol-inf_accion-graf_resp: x (0.0)
| obligations_minus1 = oferta: x (0.0)
| obligations_minus1 = dir-accion_oferta: x (0.0)
| obligations_minus1 = despliega_resp: x (0.0)
commgr = afirm_acepta-parte_exclama: resp (1.0)
commgr = afirm_quiza: x (2.0)
commgr = SNE: sol-inf (12.0)
commgr = accion-graf
| obligations_minus1 = sol-inf: sol-inf_accion-graf (0.0)
| obligations_minus1 = x
| | commgr_minus1 = x: sol-inf_accion-graf (11.0)
| | commgr_minus1 = acepta_espera_repet-refr: sol-inf_accion-graf
(0.0)
| | commgr_minus1 = ack: sol-inf_accion-graf (6.0)
| | commgr_minus1 = espera_repet-refr: sol-inf_accion-graf (0.0)
| | commgr_minus1 = afirm: sol-inf_accion-graf (0.0)
| | commgr_minus1 = accion-graf: sol-inf_accion-graf (0.0)
| | commgr_minus1 = afirm_quiza: sol-inf_accion-graf (0.0)
| | commgr_minus1 = exclama: sol-inf_accion-graf (0.0)
| | commgr_minus1 = acepta: sol-inf_accion-graf_resp (8.0/1.0)
| | commgr_minus1 = repet-refr: sol-inf_accion-graf (0.0)
| | commgr_minus1 = afirm_acepta: sol-inf_accion-graf (0.0)
| | commgr_minus1 = opcion-abierta_despliega: sol-inf_accion-graf
(0.0)
| | commgr_minus1 = SNE: sol-inf_accion-graf (0.0)
| | commgr_minus1 = nil: sol-inf_accion-graf (0.0)
| | commgr_minus1 = back-channel: sol-inf_accion-graf (0.0)
| | commgr_minus1 = rechaza: sol-inf_accion-graf (0.0)
| | commgr_minus1 = perform: sol-inf_accion-graf (0.0)
| | commgr_minus1 = acepta-parte: sol-inf_accion-graf (0.0)
| | commgr_minus1 = reafirm: sol-inf_accion-graf (0.0)
| | commgr_minus1 = afirm_rechaza: sol-inf_accion-graf (0.0)

```

```

| | commgr_minus1 = espera: sol-inf_accion-graf (0.0)
| | commgr_minus1 = oferta_acepta: sol-inf_accion-graf (0.0)
| | commgr_minus1 = afirm_acepta_conv-cierra: sol-inf_accion-graf
(0.0)
| | commgr_minus1 = accion-graf_acepta: sol-inf_accion-graf (0.0)
| | commgr_minus1 = oferta_conv-abre: sol-inf_accion-graf (0.0)
| | commgr_minus1 = quiza: sol-inf_accion-graf (0.0)
| | commgr_minus1 = reafirm_espera: sol-inf_accion-graf (0.0)
| | commgr_minus1 = afirm_despliega: sol-inf_accion-graf (0.0)
| | commgr_minus1 = opcion-abierta: sol-inf_accion-graf (0.0)
| | commgr_minus1 = conv-cierra: sol-inf_accion-graf (0.0)
| | commgr_minus1 = rechaza-parte: sol-inf_accion-graf (0.0)
| | commgr_minus1 = afirm_accion-graf: sol-inf_accion-graf (0.0)
| | commgr_minus1 = opcion-abierta_despliega_acepta: sol-
inf_accion-graf (0.0)
| | commgr_minus1 = otra: sol-inf_accion-graf (0.0)
| | commgr_minus1 = reafirm_complem: sol-inf_accion-graf (0.0)
| | commgr_minus1 = afirm_correccion: sol-inf_accion-graf (0.0)
| | commgr_minus1 = afirm_espera: sol-inf_accion-graf (0.0)
| | commgr_minus1 = oferta: sol-inf_accion-graf (0.0)
| | commgr_minus1 = afirm_conv-cierra: sol-inf_accion-graf (0.0)
| | commgr_minus1 = espera_correccion: sol-inf_accion-graf (0.0)
| | commgr_minus1 = espera_SNE: sol-inf_accion-graf (0.0)
| obligations_minus1 = resp: sol-inf_accion-graf (2.0)
| obligations_minus1 = sol-inf_accion-graf: sol-inf_accion-graf (0.0)
| obligations_minus1 = dir-accion: sol-inf_accion-graf (5.0)
| obligations_minus1 = compromiso: sol-inf_accion-graf (72.0/1.0)
| obligations_minus1 = oferta_sol-inf: sol-inf_accion-graf (0.0)
| obligations_minus1 = dir-accion_resp: sol-inf_accion-graf (0.0)
| obligations_minus1 = nil: sol-inf_accion-graf (0.0)
| obligations_minus1 = despliega: sol-inf_accion-graf (0.0)
| obligations_minus1 = accion-graf: sol-inf_accion-graf (0.0)
| obligations_minus1 = sol-inf_resp: sol-inf_accion-graf_resp (1.0)
| obligations_minus1 = sol-inf_accion-graf_resp: sol-inf_accion-graf
(0.0)
| obligations_minus1 = oferta: sol-inf_accion-graf (0.0)
| obligations_minus1 = dir-accion_oferta: sol-inf_accion-graf (0.0)
| obligations_minus1 = despliega_resp: sol-inf_accion-graf (0.0)
commgr = repet-refr: x (5.0)
commgr = afirm_acepta
| obligations_minus1 = sol-inf: resp (2.0)
| obligations_minus1 = x: x (3.0)
| obligations_minus1 = resp: x (3.0)
| obligations_minus1 = sol-inf_accion-graf: dir-accion_resp (1.0)
| obligations_minus1 = dir-accion: x (0.0)
| obligations_minus1 = compromiso: x (0.0)
| obligations_minus1 = oferta_sol-inf: x (0.0)
| obligations_minus1 = dir-accion_resp: x (0.0)
| obligations_minus1 = nil: x (0.0)
| obligations_minus1 = despliega: x (0.0)
| obligations_minus1 = accion-graf: x (0.0)
| obligations_minus1 = sol-inf_resp: x (0.0)
| obligations_minus1 = sol-inf_accion-graf_resp: x (0.0)
| obligations_minus1 = oferta: x (0.0)
| obligations_minus1 = dir-accion_oferta: x (0.0)
| obligations_minus1 = despliega_resp: x (0.0)
commgr = opcion-abierta_despliega: resp (41.0/9.0)

```

commgr = oferta_acepta: x (2.0)
commgr = afirm_acepta_exclama: x (1.0)
commgr = ack: x (24.0)
commgr = conv-cierra: x (5.0)
commgr = oferta_conv-abre: x (5.0)
commgr = back-channel: x (3.0)
commgr = espera_en_admon.-tarea: x (1.0)
commgr = perform: x (4.0)
commgr = acepta-parte
| obligations_minus1 = sol-inf: dir-accion_resp (1.0)
| obligations_minus1 = x: dir-accion_resp (0.0)
| obligations_minus1 = resp: dir-accion_resp (0.0)
| obligations_minus1 = sol-inf_accion-graf: dir-accion_resp
(13.0/1.0)
| obligations_minus1 = dir-accion: dir-accion (1.0)
| obligations_minus1 = compromiso: dir-accion (2.0)
| obligations_minus1 = oferta_sol-inf: dir-accion_resp (0.0)
| obligations_minus1 = dir-accion_resp: dir-accion (1.0)
| obligations_minus1 = nil: dir-accion_resp (0.0)
| obligations_minus1 = despliega: dir-accion_resp (0.0)
| obligations_minus1 = accion-graf: dir-accion_resp (0.0)
| obligations_minus1 = sol-inf_resp: dir-accion_resp (0.0)
| obligations_minus1 = sol-inf_accion-graf_resp: sol-inf_resp (2.0)
| obligations_minus1 = oferta: dir-accion_resp (0.0)
| obligations_minus1 = dir-accion_oferta: dir-accion_resp (0.0)
| obligations_minus1 = despliega_resp: dir-accion_resp (0.0)
commgr = reafirm
| obligations_minus1 = sol-inf: resp (10.0)
| obligations_minus1 = x: x (3.0/1.0)
| obligations_minus1 = resp: x (2.0)
| obligations_minus1 = sol-inf_accion-graf: resp (0.0)
| obligations_minus1 = dir-accion: x (1.0)
| obligations_minus1 = compromiso: resp (0.0)
| obligations_minus1 = oferta_sol-inf: resp (0.0)
| obligations_minus1 = dir-accion_resp: resp (0.0)
| obligations_minus1 = nil: resp (0.0)
| obligations_minus1 = despliega: resp (0.0)
| obligations_minus1 = accion-graf: resp (0.0)
| obligations_minus1 = sol-inf_resp: resp (0.0)
| obligations_minus1 = sol-inf_accion-graf_resp: resp (0.0)
| obligations_minus1 = oferta: resp (0.0)
| obligations_minus1 = dir-accion_oferta: resp (0.0)
| obligations_minus1 = despliega_resp: resp (0.0)
commgr = afirm_rechaza: resp (4.0)
commgr = espera: sol-inf (13.0/4.0)
commgr = accion-graf_acepta: sol-inf_accion-graf (3.0/1.0)
commgr = rechaza: x (11.0/6.0)
commgr = quiza: x (1.0)
commgr = reafirm_espera: sol-inf (1.0)
commgr = afirm_despliega: resp (3.0)
commgr = opcion-abierta: x (6.0)
commgr = opcion-abierta_despliega_acepta: x (3.0/1.0)
commgr = despliega: resp (2.0/1.0)
commgr = rechaza-parte: resp (4.0)
commgr = opcion-abierta_acepta: resp (1.0)
commgr = oferta: x (2.0)
commgr = otra: x (1.0)

2	0	0	0	0	95	0	0	0	0	1	0	0		f = sol-
inf_accion-graf														
0	0	2	0	1	0	13	0	0	0	0	0	0		g = dir-
accion_resp														
1	3	0	0	2	0	1	1	0	0	0	0	0		h = sol-
inf_resp														
0	1	0	0	0	0	0	0	4	0	0	0	0		i = oferta
0	0	0	0	1	0	0	0	0	0	0	0	0		j = accion-
graf														
0	0	0	0	0	1	0	0	0	0	7	0	0		k = sol-
inf_accion-graf_resp														
0	0	0	0	0	0	0	0	0	0	0	0	1		l =
oferta_sol-inf														
0	0	0	0	0	0	0	0	0	0	0	1	0		m = dir-
accion_oferta														

Modelo 3.1.B
Target: commgr

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: sepln2007_universe_editado_con_nombres_cortos_18may2007
Instances: 1043
Attributes: 12
 number_of_tones
 first_1
 first_2
 first_3
 last_2
 optimal_pred_mood
 utt_duration
 speaker_role
 obligations
 obligations_minus1
 commgr
 commgr_minus1
Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree

obligations = sol-inf
| speaker_role = u
| | commgr_minus1 = x: x (12.0/1.0)
| | commgr_minus1 = acepta_espera_repet-refr: x (0.0)
| | commgr_minus1 = ack: espera (3.0/1.0)
| | commgr_minus1 = espera_repet-refr: x (1.0)
| | commgr_minus1 = afirm: x (3.0/2.0)
| | commgr_minus1 = accion-graf: espera (4.0/1.0)
| | commgr_minus1 = afirm_quiza: x (0.0)
| | commgr_minus1 = exclama: x (0.0)
| | commgr_minus1 = acepta: x (34.0/2.0)
| | commgr_minus1 = repet-refr: x (0.0)
| | commgr_minus1 = afirm_acepta: x (0.0)
| | commgr_minus1 = opcion_abierta_despliega: acepta (8.0/2.0)
| | commgr_minus1 = SNE: x (0.0)
| | commgr_minus1 = nil: x (0.0)
| | commgr_minus1 = back-channel: x (0.0)
| | commgr_minus1 = rechaza: x (3.0)
| | commgr_minus1 = perform: x (1.0)
| | commgr_minus1 = acepta-parte: x (0.0)
| | commgr_minus1 = reafirm: x (1.0)
| | commgr_minus1 = afirm_rechaza: espera_repet-refr (1.0)
| | commgr_minus1 = espera: espera (1.0)
| | commgr_minus1 = oferta_acepta: acepta (1.0)
| | commgr_minus1 = afirm_acepta_conv-cierra: x (0.0)
| | commgr_minus1 = accion-graf_acepta: x (0.0)
| | commgr_minus1 = oferta_conv-abre: x (0.0)
| | commgr_minus1 = quiza: x (0.0)
| | commgr_minus1 = reafirm_espera: x (0.0)

```

| | commgr_minus1 = afirm_despliega: x (0.0)
| | commgr_minus1 = opcion-abierta: acepta (1.0)
| | commgr_minus1 = conv-cierra: x (0.0)
| | commgr_minus1 = rechaza-parte: x (0.0)
| | commgr_minus1 = afirm_accion-graf: x (0.0)
| | commgr_minus1 = opcion-abierta_despliega_acepta: x (0.0)
| | commgr_minus1 = otra: x (0.0)
| | commgr_minus1 = reafirm_complem: x (0.0)
| | commgr_minus1 = afirm_correccion: acepta (1.0)
| | commgr_minus1 = afirm_espera: x (1.0)
| | commgr_minus1 = oferta: x (0.0)
| | commgr_minus1 = afirm_conv-cierra: x (0.0)
| | commgr_minus1 = espera_correccion: x (0.0)
| | commgr_minus1 = espera_SNE: x (0.0)
| speaker_role = s: espera_repet-refr (99.0/47.0)
obligations = resp
| commgr_minus1 = x
| | number_of_tones <= 5: acepta (17.0/4.0)
| | number_of_tones > 5: afirm (18.0/8.0)
| commgr_minus1 = acepta_espera_repet-refr: acepta (2.0)
| commgr_minus1 = ack: acepta (1.0)
| commgr_minus1 = espera_repet-refr: acepta (48.0/7.0)
| commgr_minus1 = afirm
| | speaker_role = u: afirm (2.0)
| | speaker_role = s: acepta (3.0/2.0)
| commgr_minus1 = accion-graf: acepta (56.0/5.0)
| commgr_minus1 = afirm_quiza: acepta (0.0)
| commgr_minus1 = exclama: acepta (0.0)
| commgr_minus1 = acepta
| | speaker_role = u
| | | obligations_minus1 = sol-inf: afirm (8.0/1.0)
| | | obligations_minus1 = x: afirm (2.0)
| | | obligations_minus1 = resp: acepta (13.0/1.0)
| | | obligations_minus1 = sol-inf_accion-graf: acepta (0.0)
| | | obligations_minus1 = dir-accion: acepta (0.0)
| | | obligations_minus1 = compromiso: acepta (0.0)
| | | obligations_minus1 = oferta_sol-inf: acepta (0.0)
| | | obligations_minus1 = dir-accion_resp: acepta (0.0)
| | | obligations_minus1 = nil: acepta (0.0)
| | | obligations_minus1 = despliega: acepta (0.0)
| | | obligations_minus1 = accion-graf: acepta (0.0)
| | | obligations_minus1 = sol-inf_resp: acepta (0.0)
| | | obligations_minus1 = sol-inf_accion-graf_resp: acepta (0.0)
| | | obligations_minus1 = oferta: acepta (0.0)
| | | obligations_minus1 = dir-accion_oferta: acepta (0.0)
| | | obligations_minus1 = despliega_resp: acepta (0.0)
| | speaker_role = s: opcion-abierta_despliega (33.0/7.0)
| commgr_minus1 = repet-refr: acepta (0.0)
| commgr_minus1 = afirm_acepta: acepta (0.0)
| commgr_minus1 = opcion-abierta_despliega: acepta (0.0)
| commgr_minus1 = SNE: reafirm (9.0)
| commgr_minus1 = nil: acepta (0.0)
| commgr_minus1 = back-channel: acepta (0.0)
| commgr_minus1 = rechaza: acepta (0.0)
| commgr_minus1 = perform: acepta (0.0)
| commgr_minus1 = acepta-parte: acepta (0.0)
| commgr_minus1 = reafirm: afirm (1.0)

```

```

| commgr_minus1 = afirm_rechaza: acepta (0.0)
| commgr_minus1 = espera: afirm (7.0/2.0)
| commgr_minus1 = oferta_acepta: acepta (0.0)
| commgr_minus1 = afirm_acepta_conv-cierra: acepta (0.0)
| commgr_minus1 = accion-graf_acepta: acepta (1.0)
| commgr_minus1 = oferta_conv-abre: acepta (0.0)
| commgr_minus1 = quiza: acepta (0.0)
| commgr_minus1 = reafirm_espera: rechaza (1.0)
| commgr_minus1 = afirm_despliega: acepta (0.0)
| commgr_minus1 = opcion-abierta: acepta (0.0)
| commgr_minus1 = conv-cierra: acepta (0.0)
| commgr_minus1 = rechaza-parte: rechaza-parte (1.0)
| commgr_minus1 = afirm_accion-graf: acepta (0.0)
| commgr_minus1 = opcion-abierta_despliega_acepta: acepta (0.0)
| commgr_minus1 = otra: acepta (0.0)
| commgr_minus1 = reafirm_complem: acepta (0.0)
| commgr_minus1 = afirm_correccion: acepta (0.0)
| commgr_minus1 = afirm_espera: acepta (0.0)
| commgr_minus1 = oferta: acepta (0.0)
| commgr_minus1 = afirm_conv-cierra: acepta (0.0)
| commgr_minus1 = espera_correccion: acepta (1.0)
| commgr_minus1 = espera_SNE: reafirm (1.0)
obligations = dir-accion
| commgr_minus1 = x: x (22.0/2.0)
| commgr_minus1 = acepta_espera_repet-refr: x (0.0)
| commgr_minus1 = ack
| | utt_duration <= 5006: x (2.0/1.0)
| | utt_duration > 5006: acepta (3.0/1.0)
| commgr_minus1 = espera_repet-refr: x (0.0)
| commgr_minus1 = afirm
| | number_of_tones <= 17: acepta (3.0)
| | number_of_tones > 17: x (2.0)
| commgr_minus1 = accion-graf: x (2.0)
| commgr_minus1 = afirm_quiza: x (0.0)
| commgr_minus1 = exclama: x (0.0)
| commgr_minus1 = acepta
| | obligations_minus1 = sol-inf: x (0.0)
| | obligations_minus1 = x: x (10.0/3.0)
| | obligations_minus1 = resp: x (34.0)
| | obligations_minus1 = sol-inf_accion-graf: x (0.0)
| | obligations_minus1 = dir-accion: acepta (8.0/1.0)
| | obligations_minus1 = compromiso
| | | number_of_tones <= 5: afirm (2.0/1.0)
| | | number_of_tones > 5: x (4.0/1.0)
| | obligations_minus1 = oferta_sol-inf: x (0.0)
| | obligations_minus1 = dir-accion_resp: x (0.0)
| | obligations_minus1 = nil: x (0.0)
| | obligations_minus1 = despliega: x (0.0)
| | obligations_minus1 = accion-graf: x (0.0)
| | obligations_minus1 = sol-inf_resp: x (0.0)
| | obligations_minus1 = sol-inf_accion-graf_resp: x (0.0)
| | obligations_minus1 = oferta: x (1.0)
| | obligations_minus1 = dir-accion_oferta: x (0.0)
| | obligations_minus1 = despliega_resp: x (0.0)
| commgr_minus1 = repet-refr: x (2.0)
| commgr_minus1 = afirm_acepta: x (0.0)
| commgr_minus1 = opcion-abierta_despliega: acepta (10.0)

```

```

| commgr_minus1 = SNE: x (0.0)
| commgr_minus1 = nil: x (0.0)
| commgr_minus1 = back-channel: acepta (1.0)
| commgr_minus1 = rechaza: x (7.0)
| commgr_minus1 = perform: x (0.0)
| commgr_minus1 = acepta-parte: acepta-parte (3.0/1.0)
| commgr_minus1 = reafirm: x (0.0)
| commgr_minus1 = afirm_rechaza: x (0.0)
| commgr_minus1 = espera: x (0.0)
| commgr_minus1 = oferta_acepta: x (0.0)
| commgr_minus1 = afirm_acepta_conv-cierra: x (0.0)
| commgr_minus1 = accion-graf_acepta: x (0.0)
| commgr_minus1 = oferta_conv-abre: acepta (1.0)
| commgr_minus1 = quiza: x (0.0)
| commgr_minus1 = reafirm_espera: x (0.0)
| commgr_minus1 = afirm_despliega: x (0.0)
| commgr_minus1 = opcion-abierta: acepta (1.0)
| commgr_minus1 = conv-cierra: x (0.0)
| commgr_minus1 = rechaza-parte: x (1.0)
| commgr_minus1 = afirm_accion-graf: x (0.0)
| commgr_minus1 = opcion-abierta_despliega_acepta: x (0.0)
| commgr_minus1 = otra: acepta (1.0)
| commgr_minus1 = reafirm_complem: x (0.0)
| commgr_minus1 = afirm_correccion: x (0.0)
| commgr_minus1 = afirm_espera: x (0.0)
| commgr_minus1 = oferta: x (0.0)
| commgr_minus1 = afirm_conv-cierra: x (0.0)
| commgr_minus1 = espera_correccion: x (0.0)
| commgr_minus1 = espera_SNE: x (0.0)
obligations = compromiso: acepta (112.0/3.0)
obligations = x
| commgr_minus1 = x
| | utt_duration <= 811
| | | speaker_role = u: ack (4.0/2.0)
| | | speaker_role = s: acepta (34.0/3.0)
| | utt_duration > 811: x (40.0/18.0)
| commgr_minus1 = acepta_espera_repet-refr: acepta (0.0)
| commgr_minus1 = ack
| | first_2 = MS: x (2.0/1.0)
| | first_2 = BS: afirm (0.0)
| | first_2 = BU: afirm (1.0)
| | first_2 = MB: x (1.0)
| | first_2 = MT: conv-cierra (2.0)
| | first_2 = MH: afirm (1.0)
| | first_2 = MM: afirm (0.0)
| | first_2 = BT: afirm (0.0)
| | first_2 = MU: afirm (0.0)
| | first_2 = ML: afirm (0.0)
| | first_2 = BH: ack (1.0)
| | first_2 = MD: afirm (0.0)
| commgr_minus1 = espera_repet-refr: acepta (2.0)
| commgr_minus1 = afirm: acepta (30.0/20.0)
| commgr_minus1 = accion-graf: x (3.0/2.0)
| commgr_minus1 = afirm_quiza: afirm_quiza (2.0)
| commgr_minus1 = exclama: acepta (0.0)
| commgr_minus1 = acepta
| | first_3 = MST

```

```
| | | utt_duration <= 1356: ack (2.0/1.0)
| | | utt_duration > 1356: afirm (4.0/1.0)
| | first_3 = BST: perform (1.0)
| | first_3 = BUT: afirm (0.0)
| | first_3 = MBU: otra (1.0)
| | first_3 = BSU: x (1.0)
| | first_3 = MTD
| | | obligations_minus1 = sol-inf: opcion-abierta_despliega
(0.0)
| | | obligations_minus1 = x: opcion-abierta_despliega (5.0)
| | | obligations_minus1 = resp
| | | | last_2 = DB: ack (2.0/1.0)
| | | | last_2 = LH: acepta (0.0)
| | | | last_2 = TS: acepta (0.0)
| | | | last_2 = UT: acepta (0.0)
| | | | last_2 = LT: acepta (0.0)
| | | | last_2 = BS: acepta (2.0)
| | | | last_2 = LS: acepta (0.0)
| | | | last_2 = MB: acepta (0.0)
| | | | last_2 = ST: acepta (0.0)
| | | | last_2 = DS: acepta (0.0)
| | | | last_2 = TB: acepta (0.0)
| | | | last_2 = BT: acepta (0.0)
| | | | last_2 = TL: acepta (0.0)
| | | | last_2 = US: acepta (0.0)
| | | | last_2 = HL: acepta (0.0)
| | | | last_2 = BH: afirm (2.0)
| | | | last_2 = SH: acepta (0.0)
| | | | last_2 = HB: acepta (0.0)
| | | | last_2 = SB: acepta (0.0)
| | | | last_2 = HS: acepta (0.0)
| | | | last_2 = UH: acepta (0.0)
| | | | last_2 = DL: acepta (0.0)
| | | | last_2 = SL: acepta (0.0)
| | | | last_2 = BU: acepta (0.0)
| | | | last_2 = HD: acepta (0.0)
| | | | last_2 = LU: acepta (0.0)
| | | | last_2 = UM: acepta (0.0)
| | | | last_2 = DD: acepta (0.0)
| | | | last_2 = DM: acepta (0.0)
| | | | last_2 = DU: acepta (0.0)
| | | | last_2 = SU: acepta (0.0)
| | | obligations_minus1 = sol-inf_accion-graf: opcion-
abierta_despliega (0.0)
| | | obligations_minus1 = dir-accion: opcion-abierta_despliega
(0.0)
| | | obligations_minus1 = compromiso: afirm (2.0)
| | | obligations_minus1 = oferta_sol-inf: opcion-
abierta_despliega (0.0)
| | | obligations_minus1 = dir-accion_resp: opcion-
abierta_despliega (0.0)
| | | obligations_minus1 = nil: opcion-abierta_despliega (0.0)
| | | obligations_minus1 = despliega: opcion-abierta_despliega
(0.0)
| | | obligations_minus1 = accion-graf: opcion-abierta_despliega
(0.0)
```

| | | obligations_minus1 = sol-inf_resp: opcion-abierta_despliega (0.0)
| | | obligations_minus1 = sol-inf_accion-graf_resp: opcion-abierta_despliega (0.0)
| | | obligations_minus1 = oferta: opcion-abierta_despliega (0.0)
| | | obligations_minus1 = dir-accion_oferta: opcion-abierta_despliega (0.0)
| | | obligations_minus1 = despliega_resp: opcion-abierta_despliega (0.0)
| | first_3 = MTB
| | | speaker_role = u: afirm (3.0/1.0)
| | | speaker_role = s: acepta (8.0/3.0)
| | first_3 = MTL: x (2.0)
| | first_3 = MSH: x (2.0/1.0)
| | first_3 = MSB: ack (1.0)
| | first_3 = MB
| | | speaker_role = u: x (3.0/2.0)
| | | speaker_role = s
| | | | utt_duration <= 401: ack (2.0)
| | | | utt_duration > 401: acepta (3.0)
| | first_3 = MHD: afirm (0.0)
| | first_3 = MMS: afirm (0.0)
| | first_3 = MBT: x (3.0/2.0)
| | first_3 = MSU: afirm (0.0)
| | first_3 = BT: acepta (1.0)
| | first_3 = BTS: x (2.0/1.0)
| | first_3 = MUT: acepta (1.0)
| | first_3 = MHS: afirm (2.0)
| | first_3 = BSH: afirm (0.0)
| | first_3 = MHL: x (3.0/2.0)
| | first_3 = MSL: x (2.0/1.0)
| | first_3 = MTS: x (1.0)
| | first_3 = MBS: x (1.0)
| | first_3 = MML: afirm (0.0)
| | first_3 = MLS: afirm (1.0)
| | first_3 = MLT: x (3.0/2.0)
| | first_3 = BHL: afirm (0.0)
| | first_3 = MMT: afirm (0.0)
| | first_3 = MDS: x (3.0/1.0)
| | first_3 = MLH: afirm (2.0)
| | first_3 = BUU: afirm (0.0)
| | first_3 = MLU: afirm (1.0)
| | first_3 = MHB: afirm (0.0)
| | first_3 = MDB: afirm (0.0)
| | first_3 = MBH: afirm (0.0)
| | first_3 = MDH: afirm (0.0)
| | first_3 = BUH: afirm (0.0)
| | first_3 = BTD: x (2.0/1.0)
| | first_3 = BTL: acepta (1.0)
| | first_3 = MUH: afirm (0.0)
| | first_3 = MSD: x (2.0/1.0)
| | first_3 = MDL: afirm (1.0)
| | first_3 = BHS: afirm (0.0)
| | first_3 = MDD: afirm (1.0)
| | first_3 = BHD: afirm (0.0)
| | first_3 = MUS: afirm (0.0)
| | first_3 = BUS: afirm (0.0)

```

| | first_3 = MUU: afirm (0.0)
| | first_3 = MSM: afirm (0.0)
| | first_3 = MMD: afirm (0.0)
| | first_3 = MDU: afirm (0.0)
| | first_3 = MMM: afirm (0.0)
| commgr_minus1 = repet-refr
| | speaker_role = u: repet-refr (2.0/1.0)
| | speaker_role = s: acepta (2.0)
| commgr_minus1 = afirm_acepta
| | number_of_tones <= 3: x (2.0/1.0)
| | number_of_tones > 3: acepta (3.0)
| commgr_minus1 = opcion-abierta_despliega
| | first_2 = MS: opcion-abierta_despliega (2.0/1.0)
| | first_2 = BS: afirm_acepta (0.0)
| | first_2 = BU: afirm_acepta (0.0)
| | first_2 = MB: rechaza (1.0)
| | first_2 = MT: x (3.0/2.0)
| | first_2 = MH: afirm_acepta (0.0)
| | first_2 = MM: afirm_acepta (0.0)
| | first_2 = BT: acepta (1.0)
| | first_2 = MU: afirm_acepta (0.0)
| | first_2 = ML: afirm_acepta (1.0)
| | first_2 = BH: afirm_acepta (0.0)
| | first_2 = MD: afirm_acepta (2.0)
| commgr_minus1 = SNE: acepta (0.0)
| commgr_minus1 = nil: oferta_conv-abre (7.0/2.0)
| commgr_minus1 = back-channel: acepta (3.0/1.0)
| commgr_minus1 = rechaza
| | first_3 = MST: afirm (1.0)
| | first_3 = BST: afirm (0.0)
| | first_3 = BUT: afirm (0.0)
| | first_3 = MBU: afirm (0.0)
| | first_3 = BSU: afirm (0.0)
| | first_3 = MTD: afirm (0.0)
| | first_3 = MTB: acepta (2.0/1.0)
| | first_3 = MTL: rechaza (1.0)
| | first_3 = MSH: afirm (0.0)
| | first_3 = MSB: afirm (0.0)
| | first_3 = MB: afirm (0.0)
| | first_3 = MHD: afirm (0.0)
| | first_3 = MMS: afirm (0.0)
| | first_3 = MBT: afirm (0.0)
| | first_3 = MSU: afirm (2.0)
| | first_3 = BT: afirm (0.0)
| | first_3 = BTS: afirm (0.0)
| | first_3 = MUT: afirm (0.0)
| | first_3 = MHS: afirm (0.0)
| | first_3 = BSH: afirm (0.0)
| | first_3 = MHL: afirm (0.0)
| | first_3 = MSL: afirm (0.0)
| | first_3 = MTS: afirm (0.0)
| | first_3 = MBS: afirm (0.0)
| | first_3 = MML: afirm (0.0)
| | first_3 = MLS: afirm (0.0)
| | first_3 = MLT: afirm (0.0)
| | first_3 = BHL: afirm (0.0)
| | first_3 = MMT: afirm (0.0)

```

```

| | first_3 = MDS: afirm (0.0)
| | first_3 = MLH: afirm (0.0)
| | first_3 = BUU: afirm (0.0)
| | first_3 = MLU: afirm (0.0)
| | first_3 = MHB: afirm (0.0)
| | first_3 = MDB: afirm (0.0)
| | first_3 = MBH: afirm (0.0)
| | first_3 = MDH: afirm (0.0)
| | first_3 = BUH: afirm (0.0)
| | first_3 = BTD: afirm (0.0)
| | first_3 = BTL: afirm (0.0)
| | first_3 = MUH: afirm (0.0)
| | first_3 = MSD: acepta (1.0)
| | first_3 = MDL: afirm (0.0)
| | first_3 = BHS: afirm (0.0)
| | first_3 = MDD: afirm (0.0)
| | first_3 = BHD: afirm (0.0)
| | first_3 = MUS: afirm (0.0)
| | first_3 = BUS: afirm (0.0)
| | first_3 = MUU: afirm (0.0)
| | first_3 = MSM: afirm (0.0)
| | first_3 = MMD: afirm (0.0)
| | first_3 = MDU: afirm (0.0)
| | first_3 = MMM: afirm (0.0)
| commgr_minus1 = perform: acepta (0.0)
| commgr_minus1 = acepta-parte: acepta (5.0/2.0)
| commgr_minus1 = reafirm
| | number_of_tones <= 5: x (2.0/1.0)
| | number_of_tones > 5: reafirm (2.0/1.0)
| commgr_minus1 = afirm_rechaza: acepta (2.0/1.0)
| commgr_minus1 = espera: acepta (3.0)
| commgr_minus1 = oferta_acepta: acepta (0.0)
| commgr_minus1 = afirm_acepta_conv-cierra: acepta (0.0)
| commgr_minus1 = accion-graf_acepta: x (2.0/1.0)
| commgr_minus1 = oferta_conv-abre: rechaza (1.0)
| commgr_minus1 = quiza: acepta (0.0)
| commgr_minus1 = reafirm_espera: acepta (0.0)
| commgr_minus1 = afirm_despliega: opcion-abierta (3.0)
| commgr_minus1 = opcion-abierta: acepta (3.0/1.0)
| commgr_minus1 = conv-cierra
| | number_of_tones <= 4: acepta (3.0/1.0)
| | number_of_tones > 4: conv-cierra (2.0)
| commgr_minus1 = rechaza-parte: afirm (1.0)
| commgr_minus1 = afirm_accion-graf: x (2.0/1.0)
| commgr_minus1 = opcion-abierta_despliega_acepta: afirm_acepta (1.0)
| commgr_minus1 = otra: acepta (0.0)
| commgr_minus1 = reafirm_complem: acepta (0.0)
| commgr_minus1 = afirm_correccion: acepta (0.0)
| commgr_minus1 = afirm_espera: afirm (1.0)
| commgr_minus1 = oferta: acepta (0.0)
| commgr_minus1 = afirm_conv-cierra: acepta (2.0/1.0)
| commgr_minus1 = espera_correccion: acepta (0.0)
| commgr_minus1 = espera_SNE: acepta (0.0)
obligations = sol-inf_accion-graf: accion-graf (98.0/2.0)
obligations = dir_accion_resp: acepta-parte (16.0/3.0)
obligations = sol-inf_resp
| obligations_minus1 = sol-inf: acepta (3.0/2.0)

```

```

| obligations_minus1 = x: x (1.0)
| obligations_minus1 = resp: acepta-parte (0.0)
| obligations_minus1 = sol-inf_accion-graf: espera (2.0/1.0)
| obligations_minus1 = dir-accion: acepta-parte (0.0)
| obligations_minus1 = compromiso: acepta-parte (0.0)
| obligations_minus1 = oferta_sol-inf: acepta-parte (0.0)
| obligations_minus1 = dir-accion_resp: acepta-parte (0.0)
| obligations_minus1 = nil: acepta-parte (0.0)
| obligations_minus1 = despliega: acepta-parte (0.0)
| obligations_minus1 = accion-graf: acepta-parte (0.0)
| obligations_minus1 = sol-inf_resp: acepta-parte (0.0)
| obligations_minus1 = sol-inf_accion-graf_resp: acepta-parte (2.0)
| obligations_minus1 = oferta: acepta-parte (0.0)
| obligations_minus1 = dir-accion_oferta: acepta-parte (0.0)
| obligations_minus1 = despliega_resp: acepta-parte (0.0)
obligations = oferta: acepta (5.0)
obligations = accion-graf: afirm_accion-graf (1.0)
obligations = sol-inf_accion-graf_resp: accion-graf (8.0)
obligations = oferta_sol-inf: acepta (1.0)
obligations = dir-accion_oferta: acepta (1.0)

```

Number of Leaves : 406

Size of the tree : 437

Time taken to build model: 0.09 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	720	69.0316 %
Incorrectly Classified Instances	323	30.9684 %
Kappa statistic	0.6054	
Mean absolute error	0.017	
Root mean squared error	0.1016	
Relative absolute error	47.2531 %	
Root relative squared error	76.0237 %	
Total Number of Instances	1043	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
0.71	0.102	0.637	0.71	0.671	x
0	0	0	0	0	
accepta_espera_repet-refr					
0.828	0.156	0.755	0.828	0.79	accepta
0.926	0.046	0.526	0.926	0.671	espera_repet-refr
0.425	0.036	0.47	0.425	0.446	afirm
0	0	0	0	0	afirm_acepta-
parte_exclama					
1	0	1	1	1	afirm_quiza
0.083	0	1	0.083	0.154	SNE
0.99	0.002	0.981	0.99	0.986	accion-graf
0	0.003	0	0	0	repet-refr
0.111	0.003	0.25	0.111	0.154	afirm_acepta

0.756	0.018	0.633	0.756	0.689	opcion-
abierta_despliega					
0	0	0	0	0	oferta_acepta
0	0	0	0	0	
afirm_acepta_exclama					
0	0.009	0	0	0	ack
0.2	0.002	0.333	0.2	0.25	conv-cierra
1	0.002	0.714	1	0.833	oferta_conv-abre
0	0	0	0	0	back-channel
0	0	0	0	0	espera_en_admon.-
tarea					
0	0.001	0	0	0	perform
0.7	0.007	0.667	0.7	0.683	acepta-parte
0.563	0.001	0.9	0.563	0.692	reafirm
0	0	0	0	0	afirm_rechaza
0.154	0.002	0.5	0.154	0.235	espera
0	0	0	0	0	accion-graf_acepta
0	0.002	0	0	0	rechaza
0	0	0	0	0	quiza
0	0	0	0	0	reafirm_espera
0	0	0	0	0	afirm_despliega
0.5	0.001	0.75	0.5	0.6	opcion-abierta
0	0	0	0	0	opcion-
abierta_despliega_acepta					
0	0	0	0	0	despliega
0	0	0	0	0	rechaza-parte
0	0	0	0	0	opcion-
abierta_acepta					
0	0	0	0	0	oferta
0	0	0	0	0	otra
0	0	0	0	0	reafirm_complem
0	0	0	0	0	conv-abre
0	0	0	0	0	afirm_accion-graf
0	0	0	0	0	afirm_correccion
0	0.002	0	0	0	afirm_espera
0	0	0	0	0	opcion-
abierta_rechaza					
0	0	0	0	0	afirm_perform_conv-
cierra					
0	0	0	0	0	espera_SNE
0	0	0	0	0	afirm_conv-cierra

=== Confusion Matrix ===

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q
r	s	t	u	v	w	x	y	z	aa	ab	ac	ad	ae	af	ag	ah	ai
aj	ak	al	am	an	ao	ap	aq	ar	as	<-- classified as							
149	0	21	17	10	0	0	0	0	0	0	0	2	0	0	3	0	0
0	0	1	5	0	0	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0	0	0	a = x							
0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	b = acepta_espera_repet-refr							
35	0	317	9	9	0	0	0	0	3	1	3	0	0	2	2	0	0
0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	c = acepta							

```

    3  0  1  50  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0  0 | d = espera_repet-refr
    13  0  21  0  31  0  0  0  0  0  0  0  4  0  0  2  0  0
0  0  0  1  1  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0  0 | e = afirm
    0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0  0 | f = afirm_acepta-
parte_exclama
    0  0  0  0  0  0  2  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0  0 | g = afirm_quiza
    0  0  0  11  0  0  0  1  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0 | h = SNE
    0  0  0  1  0  0  0  0  104  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0 | i = accion-graf
    0  0  5  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0 | j = repet-refr
    0  0  4  0  1  0  0  0  0  0  0  1  1  0  0  1  0  0
0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0 | k = afirm_acepta
    2  0  2  0  5  0  0  0  0  0  0  0  31  0  0  1  0  0
0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0 | l = opcion-abierta_despliega
    1  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0 | m = oferta_acepta
    0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0 | n = afirm_acepta_exclama
    4  0  15  0  2  0  0  0  0  0  0  1  1  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0  0  1  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0 | o = ack
    1  0  2  0  1  0  0  0  0  0  0  0  0  0  0  0  1  0
0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0 | p = conv-cierra
    0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  5
0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0 | q = oferta_conv-abre
    0  0  3  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0 | r = back-channel
    0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0 | s = espera_en_admon.-tarea
    1  0  0  0  1  0  0  0  0  0  0  0  2  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0 | t = perform
    5  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  14  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0 | u = acepta-parte

```

5	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0		v = reafirm					
0	0	3	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0		w = afirm_rechaza					
6	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	2	0	1	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0	0	0	0		x = espera					
0	0	0	0	1	0	0	0	0	2	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0		y = accion-graf_acepta					
4	0	4	0	2	0	0	0	0	0	0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0		z = rechaza					
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0		aa = quiza					
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0		ab = reafirm_espera					
0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0		ac = afirm_despliega					
0	0	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0		ad = opcion-abierta					
0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0		ae = opcion-					
abierta_despliega_acepta																	
0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0		af = despliega					
0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0		ag = rechaza-parte					
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0		ah = opcion-abierta_acepta					
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0		ai = oferta					
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0		aj = otra					
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0		ak = reafirm_complem					
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0		al = conv-abre					
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0		am = afirm_accion-graf					

```
1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 | an = afirm_correccion
2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 | ao = afirm_espera
0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 | ap = opcion-abierta_rechaza
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 | aq = afirm_perform_conv-cierra
0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 | ar = espera_SNE
0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 | as = afirm_conv-cierra
```

Referencias

- Aguilar, 2000.** Aguilar, Lourdes. *La entonación*, en Alcoba, S. (Coord.) *La expresión oral*. Barcelona: Ariel (Ariel Practicum). pp. 115-146.
- Alcina y Brecua, 1975.** Alcina, Juan y Blecua, José Manuel. *Gramática española*, Barcelona, Ariel, 1983⁴.
- Alcoba y Murillo, 1998.** Alcoba, Santiago y Murillo, Julio. (1998) *Intonation in Spanish*, en Hirst, D. y DiCristo, A. (Eds.) *Intonation Systems. A Survey of Twenty Languages*. Cambridge: Cambridge University Press. pp. 152-166.
- Alexandersson et al., 1998.** Alexandersson, J.; Buschbeck-Wolf, B.; Fujinami, T.; Kipp, M.; Koch, S.; Maier, E.; Reithinger, N.; Schmitz, B.; Siegel, M. *Dialogue acts in VERBMOBIL-2, Second Edition, Report 226*. DFKI GmbH. Germany, July, 1998.
- Allen y Core, 1997.** Allen, James F. y Core, Mark. *Draft of DAMSL: Dialogue Act Markup in Several Layers, Technical Report*, The Multiparty Discourse Group. University of Rochester, Rochester, USA.
- Allen et al., 1996.** Allen, James F.; Miller, Bradford W.; Ringger, Eric K. y Sikorski, Teresa. *A Robust System for Natural Spoken Dialogue*, Proceedings of the 1996 Annual Meeting of the Association for Computational Linguistics (ACL'96), pp. 62-70.
- Allen et al., 2000.** Allen, James F.; Byron, Donna K.; Dzikovska, Myroslava, Ferguson, George, Galescu, Lucian y Stent, Amanda. *An Architecture for a Generic Dialogue Shell*. Natural Language Engineering, 6(3-4):213-228.
- Allen et al., 2001.** Allen, James; Byron, Donna; Dzikovska, Myroslava; Ferguson, George; Galescu, Lucian y Stent, Amanda. *Towards Conversational Human-Computer Interaction*. AI Magazine, vol. 22, pp. 27-37.
- Allen et al., 2001b.** Allen, James; Ferguson, George y Stent, Amanda. *An architecture for more realistic conversational systems*, in Proceedings of Intelligent User Interfaces 2001 (IUI-01), pages 1-8, Santa Fe, NM, January 14-17.
- Allwood, 1995.** Allwood, J. *An activity-based approach to pragmatics*. Gothenburg Papers in Theoretical Linguistics 76.
- Austin, 1962.** Austin, J.L. *How to do things with words*. Harvard University Press. Cambridge, MA, USA.
- Ávila, 2003.** Ávila, Sylvia. *La entonación del enunciado interrogativo en el español de la ciudad de México*, en Martín Butragueño, P y Herrera, Z. E. (Eds.) *La tonía. Dimensiones fonéticas y fonológicas*. México: El Colegio de México, Centro de Estudios Lingüísticos y Literarios (Cátedra Jaime Torres Bodet, Estudios de Lingüística 4). pp. 331-356.
- Bakenecker et al., 1994.** Bakenecker, G., Block, U., Batlinger, A., Kompe, R., Nöth, E., y Regel-Brietzmann, P. *Improving Parsing by Incorporating Prosodic Clause Boundaries into a Grammar*. Proc. 3rd ICSLP, pp. 1115-1118, Yokohama, Japan.
- Beckman y Ayers-Elam, 1997.** Beckman, Mary y Ayers-Elam, Gayle. *Guidelines for ToBI Labelling (version 3.0, March 1997)*. The Ohio State University Research Foundation.
- Beckman et al., 2002.** Beckman, Mary E., Díaz-Campos, Manuel; Tevis-McGory, Julia y Morgan, Terrell A. *Intonation across Spanish, in the Tones and Break Indices framework*. Probus 14 (2002), 9-36. Walter de Gruyter.

Breiman et al., 1983. Breiman, Leo; Friedman, Jerome. H.; Olshen, R.A. y Stone., Charles J. *Classification and Regression Trees*. Wadsworth.

Bunt, 1994. Bunt, H. *Context and dialogue control*. THINK Quarterly.

Bunt, 1997. Bunt, Harry. *Dynamic interpretation and dialogue theory*. ITK, Tilburg University, The Netherlands.

Carletta, 1996. Carletta, Jean. *Assessing agreement on classification tasks: the kappa statistic*. Computational Linguistics, 22(2):249–254.

Clark y Schaefer, 1989. Clark, Herbert y Schaefer, Edward. *Contributing to discourse*. Cognitive Science, 13:259–294.

Core y Allen, 1997. Core, Mark y Allen, James, *Coding Dialogs with the DAMSL Annotation Scheme*, AAAI Fall Symposium on Communicative Action in Humans and Machines, pp. 28-35, Boston, MA, U.S.A.

Coria y Pineda, 2005. Coria, Sergio y Pineda Luis. *Predicting obligation dialogue acts from prosodic information*, in Proceedings of the Second Midwest Computational Linguistics Colloquium 2005 (MCLC 2005). Ohio State University. Columbus, OH, U.S.A.

Coria y Pineda, 2005a. Coria, Sergio y Pineda Luis. *Predicting obligation dialogue acts from prosodic and speaker information*, in Research on Computing Science (ISSN 1665-9899), 14th International Congress on Computing, CIC 2005, Centro de Investigacion en Computacion, Instituto Politecnico Nacional. Mexico City, Mexico.

Coria y Pineda, 2006. Coria, Sergio y Pineda Luis. *Predicting dialogue acts from prosodic information*. Seventh International Conference on Intelligent Text Processing and Computational Linguistics, CICLing-2006, Mexico City, Mexico.

Cuétara, 2004. Cuétara, Javier O. *Fonética de la ciudad de México. Aportaciones desde las tecnologías del habla. Tesis para obtener el título de Maestro en Lingüística Hispánica*. Maestría en Lingüística Hispánica, Posgrado en Lingüística, Universidad Nacional Autónoma de México.

Cuétara y Villaseñor, 2004. *Transcribemex: Software para transcripción fonética a partir de transcripción ortográfica en español mexicano*. UNAM-INAOE, México.

Dahlback et al., 1993. Dahlback, Niels; Jonsson, Arne y Ahrenberg, Lars. *Wizard of Oz Studies - Why and How*. Knowledge-based Systems, 6(4), 258-266.

Espesser, 1999. Espesser, Robert. *MES: Motif Environment for Speech* (software). http://www.lpl.univaix.fr/ext/projects/mes_signaix.htm, 1999

Fernández and Picard, 2002. Fernandez, Raul and Picard, Rosalind W. *Dialog Act Classification from Prosodic Features Using Support Vector Machines*. Proceedings of Speech Prosody 2002. Aix-en-Provence, France.

Finke, Lapata et al., 1998. Finke, M.; Lapata, M. et al. *CLARITY: Inferring Discourse Structure from Speech*. In Proceedings AAAI '98 Spring Symposium on Applying Machine Learning to Discourse Processing.

García-Lecumberri, 2003. García-Lecumberri, M.L. *Análisis por Configuraciones: La Escuela Británica*. In Teorías de la Entonación. Pilar Prieto (coord.), Ariel, Barcelona, España, 2003, pp. 35-61.

Garrido, 1991. Garrido, Juan María. *Modelización de patrones melódicos del español para la síntesis y el reconocimiento*. Tesis doctoral. Depto. de Filología Española. Universitat Autònoma de Barcelona, España.

Garrido, 1996. Garrido, Juan María. *Modelling Spanish Intonation for Text-to-Speech Applications*. Doctoral Dissertation.. Department of Spanish Filology, Facultat de Lletres, Universitat Autònoma de Barcelona.

Godfrey et al., 1992. Godfrey, J. ; Holliman, E. y McDaniel, J. *SWITCHBOARD: Telephone Speech Corpus for Research and Development*. In Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing (Vol. 1, pp. 517–520). San Francisco, CA, USA.

Hancher, 1979. Hancher, M. *The Classification of Cooperative Illocutionary Acts*. Language and Society, 8(1):1-14.

Hirschberg y Litman, 1993. Hirschberg, J., and Litman, D. *Empirical Studies on the Disambiguation of Cue Phrases*. Computational Linguistics, Vol. 19, No. 3, 501-530. U.S.A.

Hirst et al., 2000. Hirst, Daniel; DiCristo, Albert y Espesser, Robert. *Levels of representation and levels of analysis for the description of intonation systems*, In M. Horne (ed) *Prosody: Theory and Experiment*, Kluwer-Dordrecht.

Jekat et al., 1995. Jekat, Susanne; Klein, Alexandra; Maier Elisabeth; Maleck, Ilona; Mast, Marion y Quantz, J. Joachim. *Dialogue acts in VERBMOBIL*. VM-Report 65.

Jurafsky, 2000. Jurafsky, D. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*. Prentice Hall. USA, 2000

Jurafsky, 2001. Jurafsky, D. *Pragmatics and Computational Linguistics*. To appear in Laurence R. Horn and Gregory Ward (eds.). *Handbook of Pragmatics*. Oxford: Blackwell

Jurafsky et al., 1997. Jurafsky, Daniel; Shriberg, Elizabeth and Biasca, Debra. *Switchboard SWBD-DAMSL Shallow-Discourse-Function Annotation. Coders Manual. Draft 13*. University of Colorado at Boulder & SRI International. August 1.

Jurafsky y Martin, 2006. Jurafsky, Daniel y Martin, James H. *Speech and Language Processing (2nd Edition)*.

Kompe et al., 1995. Kompe, R., Kießling, A., Niemann, H., Noth, E., Schukat Talamazzini, E., Zottmann, A., and Batliner, A. *Prosodic scoring of word hypotheses graphs*. In J. M. Pardo, E. Enríquez, J. Ortega, J. Ferreiros, J. Macías, & F. J. Valverde (Eds.), *Proceedings of the 4th European Conference on Speech Communication and Technology* (Vol. 2, pp. 1333–1336). Madrid, Spain, 1995.

Kompe et al., 1997. Kompe, R., Kießling, A., Niemann, H., Nöth, E., Batliner, A., Schachtl, S., Ruland, T., and Block, U. *Improving Parsing of Spontaneous Speech with the Help of Prosodic Boundaries*, Proc. 22nd ICASSP, IEEE. Munchen, Germany.

Krippendorff, 1980. Krippendorff, Klaus. *Content Analysis: an Introduction to its Methodology*. Sage Publications.

Léon, 1970. *Systématique des fonctions expressives de l'intonation*, en Léon, Pierre R.; Faure, Georges. y Rigault, André. (eds.). *Analyse des faits prosodiques*, Ottawa, Didier, pp. 57-72.

Levin et al., 1998. Levin, Lori; Thymé-Gobbel, Ann; Lavi, Alon; Ries, Klaus; Zechner, Klaus. *A Discourse Coding Scheme for Conversational Spanish*. International Conference on Spoken Language Processing (ICSLP'98), Sidney, Australia, 30th November-4th December.

- Mast et al., 1996.** Mast, M., Kompe, R., Harbeck, S., Kiealing, A., Niemann, H., y Noth, E.. *Dialog act classification with the help of prosody*. In Int. Conf. on Spoken Language Processing, volume 3, pp. 1728-1731, Philadelphia, USA.
- Moreno y Pineda, 2006.** Moreno, Iván y Pineda, Luis. *Speech Repairs in the DIME Corpus*, Research in Computing Science, Vol. 20, pp. 63 – 74.
- Navarro Tomás, 1948.** Navarro Tomás, Tomás. *Manual de entonación española*. Guadarrama, Madrid, España, 1974⁴.
- Pineda et al., 2002.** Pineda, Luis; Massé, Antonio; Meza, Iván; Salas, Miguel; Schwarz, Erick; Uraga, Esmeralda y Villaseñor, Luis. *The Dime Project*. Proceedings of MICAI 2002, Lectures Notes in Artificial Intelligence, Vol. 2313, Springer-Verlag, pp. 166–175, 2002. ISSN 0302–9743.
- Pineda et al., 2006.** Pineda, Luis; Castellanos, Haydé; Coria, Sergio; Estrada, Varinia; López, Fernanda; López, Isabel; Meza, Iván; Moreno, Iván; Pérez, Patricia y Rodríguez, Carlos. *Balancing transactions in practical dialogues*, in CICLing 2006, Seventh International Conference on Intelligent Text Processing and Computational Linguistics, Centro de Investigación en Computación, Instituto Politécnico Nacional, Mexico City.
- Pineda et al., 2006a.** Pineda, Luis Alberto; Estrada, Varinia Margarita y Coria, Sergio Rafael. *The obligations and common ground structures of task oriented conversations*. Proceedings of the Fourth Workshop in Information and Language Technology TIL-2006, in IBERAMIA 06, Brazil.
- Pineda et al., 2007.** Pineda, Luis Alberto. Department of Computer Science, Institute of Applied Mathematics and Systems. National Autonomous University of Mexico. *The DIME Corpus*. <http://leibniz.iimas.unam.mx/~luis/DIME/CORPUS-DIME.html>
- Power, 1979.** Power, R. *The organization of purposeful dialogs*. Linguistics 17.105–152.
- Prieto, 2003.** Prieto, Pilar, coord. *Teorías de la Entonación*, Ariel, Barcelona, España.
- Quilis, 1981.** Quilis, Antonio. *Fonética acústica de la lengua española*. Gredos (Biblioteca Románica Hispánica, Manuales 49). Madrid, España.
- Rossi et al., 1981.** Rossi, Mario; DiCristo, Albert; Hirst, Daniel; Martin, P.; Nishinuma, Y. *L'intonation, de l'acoustique à la sémantique*, París, Klincksieck.
- Schegloff, 1968.** Schegloff, E. *Sequencing in Conversational Openings*. American Anthropologist 70.1075–1095, U.S.A.
- Schegloff y Sacks, 1973.** Schegloff, E.A., y Sacks, H. *Opening up closings*. Semiotica, 8, 289-327.
- Searle, 1969.** Searle, John R. *Speech Acts. An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.
- Searle, 1969b.** Searle, J. *Actos de habla. Ensayo de Filosofía del Lenguaje. Segunda edición en Español*, Trad. de Luis M. Valdés Villanueva, Colección Teorema, Serie Mayor, Ediciones Cátedra, Madrid, 1986.
- Searle, 1975.** Searle, J. *A taxonomy of illocutionary acts*. Language, Mind and Knowledge, Minnesota Studies in the Philosophy of Science, pp. 344-369, 1975
- Shriberg et al., 1998.** Shriberg, Elizabeth; Bates, Rebecca; Stolcke, Andreas; Taylor, Paul; Jurafsky, Dan; Ries, Klaus; Coccaro, Noah; Martin, Rachel; Meteer, Marie y Van EssDykema, Carol. *Can Prosody Aid the*

Automatic Classification of Dialog Acts in Conversational Speech? Language and Speech 41(3-4), Special Issue on Prosody and Conversation, 439-487, USA.

Siegel y Castellan, 1988. Siegel, Sidney y Castellan, N. John. *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill.

Silverman et al., 1992. Silverman, K., Blaauw, E., Spitz, J., and Pitrelli, J.F. *Towards Using Prosody in Speech Recognition/Understanding Systems: Differences between Read and Spontaneous Speech*. In Proceedings, Fifth DARPA Workshop on Speech and Natural Language.

Sosa, J.M., 1991. *Fonética y fonología de la entonación del español hispanoamericano, tesis doctoral*. Univ. de Massachussets.

Sosa, J.M., 1999. *La entonación del español. Su estructura fónica, variabilidad y dialectología*. Madrid: Cátedra (Lingüística).

Sperber y Wilson, 1986. *Relevance*. Communication and Cognition, Oxford, Blackwell.

Stalnaker, 1978. Stalnaker, R.C. *Assertion*. In P. Cole (Ed.), *Syntax and Semantics: Pragmatics 9* (pp. 315-332). New York: Academic, U.S.A., 1978

'T Hart et al., 1990. 'T Hart, J.; Collier, R. y Cohen, A. *A Perceptual Study of Intonation. An Experimental-Phonetic Approach to Intonation*, Cambridge, Cambridge University Press.

Thorsen, 1979. Thorsen, Nina. *Interpreting Raw Fundamental-Frequency Tracings of Danish*. *Phonetica*, 36, pp. 57-78.

Thorsen, 1980. Thorsen, Nina. *A study of the perception of sentence intonation - Evidence from Danish*, *Journal of the Acoustical Society of America*, 67, 3, pp. 1014-1030.

Traum y Allen, 1994. Traum, David R. y Allen, James F. *Discourse obligations in dialogue processing*. In Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics (ACL-94), pages 1-8.

Wahlster, 1993. Wahlster, W. *VERBMOBIL: Translation of Spontaneous Face-to-Face Dialogs*. In Proceedings of 3rd EUROSPEECH, pp. 29-38, Berlin, Germany.

Wilson y Sperber, 1993. Wilson, Deirdre y Sperber, Dan. *Linguistic form and Relevance*. *Lingua*, 93.

Wilson y Sperber, 1988. Wilson, Deirdre y Sperber, Dan. *Mood and the analysis of non-declarative sentences*. In J. Dancy, J. Moravcsik y C. Taylor (eds) *Human agency: Language, Duty and Value*. Stanford UP, Stanford CA.

Witten y Frank, 2005. Witten, Ian y Frank, Eibe. *Data Mining: Practical Machine Learning Tools and Techniques*. 2nd Edition, Morgan Kaufmann, San Francisco, CA, USA.