

TESIS DE DOCTORADO

**PROGRAMA DE DESARROLLO DE LAS CIENCIAS BÁSICAS
ÉCOLE DOCTORALE CONNAISSANCE, LANGAGE, MODÉLISATION**

IDENTIFICACIÓN DE OPINIONES DE DIFERENTES FUENTES EN TEXTOS EN ESPAÑOL

Autor
Aiala Rosá

Orientadores
**Jean-Luc Minel
Dina Wonsever**

Setiembre de 2011

TRIBUNAL

Dra. Laura Alonso Alemany (revisora)

Facultad de Matemática, Astronomía y Física,
Universidad Nacional de Córdoba
Argentina

Dr. Héctor Cancela

Profesor Titular,
Instituto de Computación,
Facultad de Ingeniería,
Universidad de la República,
Uruguay

Dr. Javier Couto

Responsable R&D, Syllabs, Ingénierie des langues pour les NTIC
Chercheur Associé, Laboratoire MoDyCO, Université Paris Ouest Nanterre La Défense
Francia

Dra. Ivette Mathieu

Directeure de Recherche, CNRS,
Francia

Dr. Jean-Luc Minel

Professeur des Universités Paris Ouest Nanterre La Défense,
Francia

Dr. Horacio Saggion (revisor)

Profesor Investigador,
Departamento de tecnologías de la información y las comunicaciones,
Universitat Pompeu Fabra
España

Dra. Dina Wonsever

Profesora Titular,
Instituto de Computación,
Facultad de Ingeniería,
Universidad de la República,
Uruguay

Resumen

Este documento presenta un estudio de las expresiones que transmiten opiniones de diferentes fuentes en textos en español. El trabajo incluye la definición de un modelo para los predicados de opinión y sus argumentos (la fuente, el asunto y el mensaje), la creación de un léxico de predicados de opinión que tienen asociada información proveniente del modelo y la realización de tres sistemas informáticos.

Desarrollamos un primer sistema, basado en reglas contextuales, que obtiene valores de medida F parcial (incluyendo entre los elementos correctos los elementos reconocidos en forma parcial) satisfactorios: 92 % para el predicado, 81 % para la fuente, 75 % para el asunto, 89 % para el mensaje y 85 % para la opinión completa. En particular, para el reconocimiento de la fuente se obtuvo un 79 % de medida F exacta (sin incluir elementos reconocidos en forma parcial).

El segundo sistema desarrollado se basa en el modelo *Conditionnal Random Fields* (CRF) y se realizó solo para el reconocimiento de las fuentes. El sistema alcanza un valor de medida F exacta de 76 %.

Un tercer sistema, que combina las dos técnicas anteriores incorporando la salida del sistema de reglas para el reconocimiento de fuentes como un nuevo atributo del sistema basado en CRF, mejora sensiblemente los resultados obtenidos por los dos sistemas anteriores: 83 % de medida F exacta.

En cuanto al reconocimiento de las fuentes de las opiniones, nuestro sistema obtiene resultados muy satisfactorios (83 % de medida F exacta), si tomamos como referencia trabajos realizados para otros idiomas que pueden considerarse similares al nuestro, si bien presentan varias diferencias en su enfoque y su alcance. Estos trabajos alcanzan valores de medida F (exacta o parcial) que se sitúan entre 63 % y 89,5 %.

Por otro lado, durante el desarrollo de esta tesis generamos diversos recursos de utilidad para el procesamiento automático del español: un léxico de predicados de opinión, un corpus de 13.000 palabras anotado con las opiniones y sus elementos y un corpus de 40.000 palabras anotado con los predicados de opinión y sus fuentes.

ÍNDICE DE CONTENIDO

Agradecimientos.....	11
<u>PRIMERA PARTE: INTRODUCCIÓN.....</u>	<u>13</u>
1. Introducción.....	15
1.1 Motivación.....	15
1.2 Conceptos importantes.....	17
1.3 Objetivos del trabajo.....	21
1.4 Descripción general del trabajo.....	22
1.5 Principales contribuciones.....	25
1.6 Organización del documento.....	27
<u>SEGUNDA PARTE: ESTUDIOS PREVIOS.....</u>	<u>29</u>
2. Trabajos relacionados.....	31
2.1 Trabajos sobre identificación de las opiniones y sus componentes en textos de prensa	32
2.2 Minería de opiniones - Análisis de sentimientos.....	38
2.3 Trabajos para el español.....	40
2.4 Síntesis.....	41
3. Marco teórico lingüístico.....	43
3.1 Los predicados de opinión verbales.....	43
3.1.1 Verbos introductores de cita.....	44
3.1.2 Clasificación verbal basada en alternancias de diátesis.....	47
3.1.3 Clasificación verbal basada en frames.....	52
3.1.4 Clasificación verbal basada en tipos de procesos.....	53
3.1.5 Clases verbales y recurso léxico para el español.....	56
3.2 Los predicados de opinión nominales.....	60

3.3 Los predicados de opinión preposicionales.....	61
3.4 Síntesis.....	61
<u>TERCERA PARTE: NUESTRA PROPUESTA.....</u>	<u>63</u>
4. Metodología de trabajo y análisis de corpus.....	65
4.1 Metodología.....	65
4.2 Las opiniones en textos de prensa del español.....	66
4.2.1 Definición de opinión.....	66
4.2.2 Corpus de trabajo.....	67
4.2.3 Conjunto de ejemplos seleccionados.....	68
4.3 Síntesis.....	71
5. Un modelo para los predicados de opinión y sus argumentos.....	73
5.1 Breve introducción a las estructuras de rasgos tipificadas.....	74
5.2 Representación de los predicados de opinión y sus argumentos.....	75
5.2.1 Los argumentos de la opinión.....	77
5.2.2 El predicado de opinión.....	77
5.2.2.1 Los verbos de opinión.....	79
5.2.2.2 Los nombres de opinión.....	86
5.2.2.3 Los predicados de opinión preposicionales.....	88
5.3 Utilización del modelo en el desarrollo de los sistemas informáticos.....	89
5.4 Síntesis.....	90
<u>CUARTA PARTE: SISTEMAS INFORMÁTICOS.....</u>	<u>91</u>
6. Repertorio de predicados de opinión.....	93
6.1 Representación interna de las piezas léxicas.....	94
6.2 Síntesis.....	95
7. Sistema basado en reglas contextuales.....	97
7.1 El formalismo de reglas contextuales.....	99

7.2 Los módulos de reglas.....	103
7.2.1 Reglas para el reconocimiento del predicado.....	103
7.2.2 Reglas para el reconocimiento de la fuente.....	103
7.2.3 Reglas para el reconocimiento del asunto.....	105
7.2.4 Reglas para el reconocimiento del mensaje.....	107
7.2.5 Reglas para armar la opinión completa.....	107
7.3 La orientación semántica.....	108
7.4 Aplicación del sistema de reglas. Ejemplos.....	110
7.5 Síntesis.....	115
8. Evaluación del sistema basado en reglas contextuales.....	117
8.1 Resultados en el reconocimiento del predicado.....	117
8.2 Resultados en el reconocimiento del mensaje.....	118
8.3 Resultados en el reconocimiento de la fuente.....	119
8.4 Resultados en el reconocimiento del asunto.....	119
8.5 Resultados de la determinación de la orientación semántica.....	119
8.6 Síntesis.....	120
9. Sistema basado en aprendizaje automático.....	121
9.1 Introducción al modelo Conditional Random Fields.....	121
9.2 Descripción de los experimentos.....	122
9.2.1 Atributos de entrada.....	123
9.2.2 Tamaño del corpus.....	126
9.2.3 Combinación de atributos de entrada.....	126
9.3 Resultados de los experimentos.....	127
9.3.1 Análisis de los resultados de los experimentos realizados sobre el conjunto básico de atributos.....	128
9.3.2 Análisis de los resultados de los experimentos realizados sobre variantes del conjunto de atributos.....	129
9.4 Síntesis.....	130
10. Comparación y combinación de los dos sistemas.....	133

10.1 Comparación de los resultados de los dos sistemas.....	133
10.2 Combinación de los dos sistemas.....	134
10.3 Resultados del sistema combinado.....	134
10.4 Síntesis.....	136
<u>QUINTA PARTE: CONCLUSIONES.....</u>	<u>137</u>
11. Conclusiones.....	139
12. Extensiones y trabajos futuros.....	143
<u>BIBLIOGRAFÍA.....</u>	<u>145</u>
<u>ANEXOS.....</u>	<u>155</u>
Anexo A: Ejemplos seleccionados de corpus.....	157
Anexo B: Repertorio de predicados de opinión.....	159
B.1 Verbos de opinión.....	159
B.1.1 Verbos de discurso referido.....	159
B.1.2 Verbos de discurso reproducido y creencia.....	160
B.1.3 Verbos positivos con sujeto fuente.....	162
B.1.4 Verbos negativos con sujeto fuente.....	162
B.1.5 Verbos positivos con sujeto asunto.....	163
B.1.6 Verbos negativos con sujeto asunto.....	163
B.2 Nombres de opinión.....	163
B.2.1 Nombres neutros.....	163
B.2.2 Nombres positivos.....	165
B.2.3 Nombres negativos.....	165
Anexo C: Reglas Contextuales.....	167

Agradecimientos

En primer lugar, quiero agradecer el apoyo permanente de mis directores de tesis, Dina y Jean-Luc, quienes durante el desarrollo de este trabajo estuvieron siempre disponibles, tanto personalmente como a distancia, para mis consultas y mis propuestas. Sus aportes y su guía fueron indispensables para la realización de esta tesis. En particular, les agradezco enormemente su comprensión en los momentos en los cuales mis obligaciones familiares o laborales alteraron la agenda de trabajo prevista.

También agradezco mucho a los miembros del tribunal, particularmente a los revisores, por sus correcciones y comentarios.

Muchos colegas y compañeros de trabajo me brindaron su apoyo durante estos años. Quiero agradecer muy especialmente a Marisa Malcuori, quien me asesoró sobre diversas cuestiones lingüísticas, y a Guillermo Moncecchi, quien me sugirió métodos y herramientas para las tareas de Aprendizaje Automático. También agradezco mucho a todos mis compañeros del grupo PLN del Instituto de Computación (InCo) de la Facultad de Ingeniería, por sus aportes y comentarios; y a mis compañeros docentes en diferentes cursos y tareas del InCo, por haberme facilitado la tarea durante el tiempo que dediqué a escribir esta tesis.

Agradezco mucho a mi familia: a mi mamá, mi papá y mis hermanos, Ale, Tati y Nela, por todo su apoyo y por ayudarme con mis hijos; y a Juan, por ocuparse de todo cuando no estoy en casa y por escucharme y alentarme siempre.

Finalmente, quiero dedicarles este trabajo a *Manu* y *Cleo*, mis hijos preciosos.

Apoyo recibido

Esta tesis recibió apoyo económico de la Agencia Nacional de Investigación e Innovación (beca de doctorado) y del *Conseil Régional d'Île de France (Soutien à l'Encadrement de Thèse en Cotutelle Internationale)*.

PRIMERA PARTE: Introducción

1. Introducción

Actualmente es muy común la lectura de prensa electrónica a través de Internet, por lo que resulta de utilidad contar con sistemas para la extracción de diferente tipo de información según las necesidades de los usuarios. Para esto es necesario conocer las características propias de este tipo de texto de modo de lograr un análisis correcto de su contenido.

Un recurso muy frecuente en los textos de prensa es la citación por parte del autor de la noticia a expresiones de otras personas. Resulta particularmente importante, por lo tanto, poder asociar la información expresada en el texto con la fuente apropiada.

Por lo general, las expresiones de diferentes fuentes citadas por el autor de un texto constituyen opiniones, es decir, texto de contenido subjetivo, aunque también es posible que lo que esas fuentes expresan sea información fáctica u objetiva. El hecho de que muchas de las expresiones citadas transmitan opiniones agrega un elemento interesante a tener en cuenta: poder establecer si se trata de una opinión positiva o negativa. Para que esto último tenga sentido, también es necesario decir en relación a qué tema la fuente está expresando su opinión.

En este trabajo nos centramos en la identificación de fragmentos de texto en donde se transmiten expresiones de diferentes fuentes. Estudiamos las diversas formas del discurso reproducido que suelen utilizarse en los textos de prensa, ya que se trata del mecanismo típico de citación. También identificamos otros procedimientos que permiten mencionar puntos de vista de diferentes fuentes, en algunos casos, sin que se citen las palabras emitidas.

Como dijimos, muchas de estas expresiones transmiten información de carácter subjetivo, por lo que utilizamos el término opinión para referirnos a todas ellas, incluyendo a las que transmiten un contenido objetivo.

Una vez identificadas las opiniones de diferentes fuentes, buscamos determinar el tema del que se habla, que llamamos asunto, y establecer un valor de orientación semántica, que podrá ser positivo, negativo o neutro (para las expresiones de contenido objetivo).

1.1 Motivación

En el ejemplo (1.1)¹ se pueden observar menciones a seis fuentes diferentes, resaltadas mediante subrayado, en un mismo fragmento de texto de tan solo cuatro párrafos.

En el primer párrafo, observamos una primera opinión, la de la fuente *el ministro de Economía Danilo Astori* que transmite la información contenida en la subordinada que sigue al verbo *informó*. Pero dentro de esa subordinada se mencionan otras opiniones, introducidas de manera menos clara. Por un lado, se hace referencia a *el fallo de la Justicia*, existe aquí una nueva fuente, *la Justicia*, que se pronunció, a través de un *fallo*. Por otro lado, la utilización del verbo *estimar* sugiere una mención a otra opinión, de la fuente *el gobierno*, pero por tratarse de

1 Todos los ejemplos incluidos en el documento pertenecen a nuestro corpus de desarrollo, salvo que se indique explícitamente lo contrario. Las características de este corpus se describen en la sección 4.2.2. Este ejemplo particular es un fragmento de la nota "Nuevo fallo de la SCJ" publicado el 01/04/2008 en www.montevideo.com.uy.

un evento orientado hacia el futuro, no se transmite una postura determinada, ya que esta aún no fue definida.

(1.1) Tras el Consejo de Ministros de este lunes, el ministro de Economía Danilo Astori informó que el gobierno acatará el fallo de la Justicia, y que, una vez comunicados los fallos restantes, se estimará de qué manera obtener los recursos que se dejen de percibir.

Por su parte, el Partido Colorado presentará este martes en el Senado un proyecto de ley que deroga el tributo, y el Partido Nacional no descarta interpelar al titular de Economía.

El Directorio del Partido Nacional hizo público su rechazo a las manifestaciones del Partido Socialista sobre la ministra de la SCJ, Sara Bossio, a las que calificó de "un ataque absolutamente inusitado a la Justicia".

El diputado Javier García opinó que "es gravísimo. Cuando se pone en juego cierta sospecha sobre el honor y el trabajo de una jueza que tiene 40 años de actuación, se lo hace sin medir consecuencias, por el solo hecho de discrepar con un veredicto, nos parece de una gravedad muy importante".

En el tercer párrafo, se menciona una opinión de la fuente *el Directorio del Partido Nacional* sobre *las manifestaciones del Partido Socialista sobre la ministra de la SCJ, Sara Bossio*, la opinión es negativa. También se menciona otra opinión, con fuente *El Partido Socialista* y asunto *la ministra de la SCJ, Sara Bossio*. En este caso no se transmite el contenido ni la orientación semántica de lo dicho.

Finalmente, en el cuarto párrafo, se cita la opinión del *diputado Javier García* que, al igual que la opinión de *el Directorio del Partido Nacional*, tiene como tema *las manifestaciones del Partido Socialista sobre la ministra de la SCJ, Sara Bossio* y es negativa.

Dentro de la cita a las palabras de esta última fuente se mencionan otras posturas. Por un lado, la expresión *cierta sospecha sobre el honor y el trabajo de una jueza que tiene 40 años de actuación* transmite una opinión que debe atribuirse al *Partido Socialista*, aunque no esté dicho explícitamente. La expresión *discrepar con el veredicto* también transmite la opinión del *Partido Socialista* (su discrepancia) y, además, hace referencia a otra opinión, a través del nombre *veredicto*, aunque no se menciona explícitamente su fuente (que es la jueza mencionada anteriormente) ni su contenido. Por último, la expresión *nos parece de una gravedad importante* transmite una opinión que debemos atribuir a la misma fuente que emite la cita completa, información que se desprende de la utilización del pronombre *nos*.

Además de ser un problema interesante en sí mismo por su gran complejidad, el análisis de las opiniones de diferentes fuentes en textos periodísticos brinda información que puede ser útil a diversas aplicaciones:

- Es información fundamental para sistemas de respuesta automática a preguntas como: *¿Quiénes hablaron sobre el tema X?*, *¿Cuáles son las opiniones sobre el tema X?*, *¿Quiénes están a favor/en contra de X?*, *¿Qué opina X sobre Y?*.
- Si se desea realizar un análisis temporal de un texto, la identificación de las expresiones emitidas por diferentes fuentes es muy importante, ya que el hecho de que existan diferentes participantes en un discurso implica la existencia de diferentes ejes de referencia. Cuando habla el autor del texto, el momento de la enunciación coincide con el momento de escritura del texto, pero cuando se citan las palabras de otras fuentes, es

necesario ubicarse en el eje de referencia de esa nueva situación de enunciación con un foco temporal propio.

- Para la generación de resúmenes automáticos es importante contar con información sobre las opiniones mencionadas en los textos, principalmente, tener resuelta la atribución de cada segmento a sus fuentes correspondientes. De otro modo, el resumen puede resultar completamente inconsistente, mezclando información de diferentes fuentes.
- El análisis de las opiniones permite rastrear en repositorios de textos las diferentes posturas de una persona respecto a algún tema a lo largo del tiempo. También es posible recorrer el texto leyendo las opiniones de diferentes personas sobre un mismo tema.

Un ejemplo concreto de utilización de información sobre opiniones es "Appinions" (Web Appinions)², un sitio web que ofrece diferentes aplicaciones para encontrar opiniones, en diferentes dominios. En particular, una de las aplicaciones que allí se ofrecen, "AppinionsPrism" (Web AppinionsPrism), permite consultar textos de noticias publicados en la web, filtrando los resultados por tema (*topic*), por el autor de la opinión (*opinion holder*) o por una palabra clave cualquiera. Las opiniones recuperadas se muestran indicando para cada una el autor correspondiente y, además, tienen calculado un valor de polaridad, positivo, negativo o neutro.

Una motivación especial para nuestro trabajo es la carencia de desarrollos para el español en el área del Procesamiento de Lenguaje Natural (PLN) en general y, en particular, en lo que tiene que ver con la extracción de opiniones.

Por un lado, la carencia de herramientas para el español es una desventaja, ya que no siempre contamos con los recursos necesarios para llevar adelante desarrollos informáticos. Por ejemplo, no pudimos contar en las primeras etapas de nuestro trabajo con un analizador sintáctico completo, ni podemos contar con un etiquetador de roles semánticos, ambas herramientas utilizadas en varios de los trabajos que estudiamos. Tampoco contamos con recursos léxicos ni corpus anotados que permitan realizar aprendizaje automático.

Por otro lado, esa carencia nos incentiva a colaborar con el desarrollo del PLN para el español, en nuestro caso particular, desarrollando un sistema para el reconocimiento de opiniones. Veremos en el capítulo 2, "Trabajos relacionados", que no hay trabajos sobre identificación de las opiniones y sus componentes para el español, solo hay algunas propuestas relativas a la creación de diccionarios afectivos y clasificación de textos según su orientación semántica.

1.2 Conceptos importantes

El concepto de opinión con el cual trabajamos se inspira en las propuestas de (Bethard et al, 2004, Wiebe et al, 2005, Liu, 2010)³. En base a estos trabajos elaboramos la siguiente definición:

Opinión: Segmento de texto que transmite las expresiones o posturas de alguna fuente sobre algún asunto.

2 La notación (Web XX) indica que se da la referencia web de XX en la sección de referencias del documento.

3 En la sección 4.2.1 describimos las definiciones propuestas en los trabajos estudiados y explicamos los elementos que nuestra definición toma de cada uno de ellos.

La opinión es, por lo tanto, un fragmento de texto que transmite lo que alguien opina o dice sobre algo. Hablamos de expresiones, para abarcar todo lo que alguien puede decir, incluyendo información objetiva, como lo que *el ministro de Economía* informa en el primer párrafo de (1.1). Incluimos también las posturas, ya que en algunos casos no se citan las palabras pronunciadas por alguna persona pero sí se menciona su postura respecto a algo, como la postura negativa del *Directorio del Partido Nacional* respecto a *las manifestaciones del Partido Socialista sobre la Ministra de la SCJ, Sara Bossio*, en el segundo párrafo de (1.1).

Los elementos de la opinión

Llamamos fuerte⁴ al autor de la opinión, que generalmente es una persona, como *el ministro de Economía Danilo Astori* en (1.1). También puede ser una organización, como *el Partido Nacional* o *el Partido Socialista* en (1.1); o un documento, en particular, una publicación como *el semanario* en el ejemplo (1.2) que mostramos a continuación.

- (1.2) Según el semanario, legisladores oficialistas estarían negociando con el entorno del presidente Vázquez y con sectores de la oposición una fórmula alternativa al proyecto, para que no fracasase debido al anunciado veto presidencial.

El asunto es el tema sobre el que se opina. Este elemento puede estar mencionado en forma explícita, como en el caso del segmento *sobre la ministra de la SCJ, Sara Bossio* en el ejemplo (1.1). Pero es frecuente que el asunto no aparezca introducido en forma explícita y deba ser deducido a partir de un análisis del mensaje. Por ejemplo, en el primer párrafo de (1.1), *el ministro de Economía Danilo Astori* habla sobre *el fallo de la Justicia*; obtenemos esta información analizando la subordinada al verbo *informó*.

Además de identificar la fuente y el asunto, para completar la opinión es necesario identificar el segmento que contiene las palabras expresadas. Utilizamos el término mensaje para referirnos a este elemento que podemos definir como el contenido de la opinión. En el ejemplo (1.1), las subordinadas que siguen a los verbos *informó* y *opinó* constituyen el mensaje de las opiniones introducidos por esos verbos.

Para transmitir opiniones de diferentes fuentes, los autores recurren a diferentes recursos lingüísticos, entre los cuales un caso típico es el discurso reproducido, en donde se utiliza un verbo de comunicación (verbo de *decir*) como introductor de una cita directa (1.3) o indirecta (1.4).

- (1.3) Damiani **había confesado** a Observa: "A Matosas lo banco hasta el final".

- (1.4) El investigador de la Politécnica **afirma** que el principal problema de este sistema es conseguir que sea fácil de usar.

En estos casos es raro encontrar un constituyente específico que introduzca el asunto (como grupos introducidos por *sobre*, *respecto a* o expresiones similares). El mensaje es la cita reproducida, que en (1.3) corresponde al texto entre comillas y en (1.4) a la subordinada con *que*. La fuente es el sujeto del verbo introductor de cita: en (1.3) es *Damiani* y en (1.4) *el investigador de la Politécnica*.

4 En la sección 4.2.1 explicamos la procedencia de los términos utilizados para referirnos a los elementos de la opinión.

Además del discurso reproducido, mecanismo principal de cita, abarcamos en nuestro estudio otras expresiones que permiten hacer referencia a opiniones de diferentes fuentes: verbos de diversas clases semánticas de contenido subjetivo, como valoración, evaluación, sensación (por ejemplo, *discrepar* en 1.1); elementos específicos de atribución de autoría como *según* (en el ejemplo 1.2); nombres, que en general son derivados tanto de los verbos de discurso como de los verbos pertenecientes a las otras clases mencionadas (como *manifestaciones* o *rechazo* en 1.1).

De estas últimas observaciones sobre los diferentes mecanismos que identificamos para la expresión de opiniones surge un nuevo elemento que es importante tener en cuenta: el elemento indicador de la presencia de una opinión, que llamamos predicado de opinión. Como ya vimos, se trata generalmente de un verbo de comunicación o de ciertas clases semánticas subjetivas, pero también puede ser un nombre, o una preposición o locución prepositiva (*según, a juicio de, de acuerdo con*). Los elementos que definimos anteriormente para la opinión -fuente, asunto y mensaje- son por lo general argumentos de estos predicados y, como veremos más adelante en este documento, los identificamos luego de que localizamos al predicado que los nuclea.

Recapitulando, en nuestro trabajo consideramos que la opinión contiene tres elementos que la definen: una fuente, un asunto y un mensaje; y un elemento más, el predicado de opinión, que permite identificar en los textos las expresiones que transmiten opiniones. Retomamos el ejemplo (1.1), indicando las diferentes opiniones allí expresadas. Para mostrar cada elemento de la opinión utilizamos la siguiente notación: predicado en negrita, fuente subrayada, asunto en cursiva y mensaje con fondo gris.

- O1 el ministro de Economía Danilo Astori **informó** *que el gobierno acatará el fallo de la Justicia, y que, una vez comunicados los fallos restantes, se estimará de qué manera obtener los recursos que se dejen de percibir.*
- O2 el fallo de la Justicia
- O3 *se estimará de qué manera obtener los recursos que se dejen de percibir*
- O4 El Directorio del Partido Nacional hizo público su **rechazo** *a las manifestaciones del Partido Socialista sobre la ministra de la SCJ, Sara Bossio*
- O5 las manifestaciones del Partido Socialista *sobre la ministra de la SCJ, Sara Bossio*
- O6 *a las que* **calificó** de "un ataque absolutamente inusitado a la Justicia".
- O7 El diputado Javier García **opinó** *que "es gravísimo. Cuando se pone en juego cierta sospecha sobre el honor y el trabajo de una jueza que tiene 40 años de actuación, se lo hace sin medir consecuencias, por el solo hecho de discrepar con un veredicto, nos parece de una gravedad muy importante".*
- O8 *cierta* **sospecha** *sobre el honor y el trabajo de una jueza que tiene 40 años de actuación*
- O9 **discrepar** *con un veredicto*
- O10 **veredicto**
- O11 nos parece *de una gravedad muy importante*

Entre las 11 opiniones que identificamos en el fragmento de texto de (1.1), observamos algunos fenómenos lingüísticos que intervienen a la hora de hacer una interpretación correcta y completa de cada opinión:

- La resolución de anáforas y correferencias: En O3, O6 y O11 encontramos elementos anafóricos haciendo referencia a las fuentes (O3 y O11) y el asunto (O6) de las opiniones allí expresadas. Además hay fuentes ausentes, que deben ser recuperadas en el contexto (O6, O8, O9 y O10). Para completar el análisis, es necesario también establecer las cadenas de correferencias de modo de establecer las diferentes menciones a una misma fuente.
- La factividad de los eventos: En O3 el verbo *estimaré* está en tiempo futuro, por lo que se trata de un evento presentado como no factivo (no realizado) en el texto. La postura que la fuente tomará respecto al asunto no se conoce en el momento de la enunciación.

Por un lado, se trata de fenómenos que no son particulares de las expresiones de opinión, sino que están presentes en cualquier análisis que se realice del contenido de los textos. Por otro lado, su incidencia en el contexto del estudio de opiniones tiene ciertas particularidades que es importante tener en cuenta.

En lo que tiene que ver con la resolución de correferencias, tiene sentido asumir que las opiniones consecutivas en un texto suelen corresponder a una misma fuente, por más que entre ellas existan fragmentos de texto extensos (sin ocurrencias de otras opiniones). En el capítulo 11, "Extensiones y trabajos futuros", mencionamos una extensión a nuestro trabajo que aborda esta problemática.

La orientación semántica

Un último concepto que queremos introducir es el de orientación semántica. Es la única propiedad puramente semántica que definimos para la opinión. Se trata de un valor, positivo, negativo o neutro, asociado a cada opinión presente en un texto.

En los diferentes ejemplos que mostramos, observamos que el valor de orientación semántica puede obtenerse a partir del análisis de los cuatro elementos que definimos para la opinión:

- a) El valor de orientación semántica puede venir expresado en el predicado de opinión, cuando este es un verbo o una nominalización perteneciente a una clase semántica subjetiva. Por ejemplo, en la opinión O4, la utilización del nombre *rechazo* nos indica que se trata de una opinión negativa de la fuente (*El Directorio del Partido Nacional*) respecto al asunto (*las manifestaciones del Partido Socialista sobre la ministra de la SCJ, Sara Bossio*).
- b) La orientación semántica también puede determinarse a partir del introductor de asunto, como en (1.5), citado a continuación, en donde sabemos que se transmite una opinión negativa de la fuente (*la Iglesia Católica*) ya que el asunto (*cualquier tipo de despenalización*) está introducido por la expresión *contraria a*.

(1.5) La Iglesia Católica también hizo pública su **opinión**, *contraria a cualquier tipo de despenalización*.

- c) También la fuente puede contener indicios de la orientación semántica de la opinión, como *los optimistas* en el ejemplo (1.6) a continuación, que nos permite deducir que el autor va a citar una opinión positiva de esta fuente.

(1.6) *Los optimistas sostienen* que la probabilidad de que esto suceda es muy baja pues el comportamiento que han mostrado los embalses hasta el momento es un indicador de que San Pedro será benévolo y los niveles de agua no llegarán a extremos.

- d) Si no se da ninguno de los casos anteriores, la orientación semántica debe deducirse a partir del análisis del mensaje, buscando en él elementos subjetivos y calculando de alguna forma su combinación. Por ejemplo, en (1.4) se utiliza un predicado neutro (*afirma*) pero sabemos que la opinión de la fuente (*el investigador de la Politécnica*) respecto al asunto (*este sistema*) es negativa, ya que se dice del sistema que es un problema y, además, de lo dicho se puede deducir que no es fácil de usar.

En algunos casos, la utilización de ciertos estilos particulares, como la ironía, dificulta la tarea de determinar la orientación semántica de una opinión. En el ejemplo (1.7), el mensaje de la opinión contiene elementos positivos (*suspiro de alivio*), sin embargo, la expresión *con sorna*, modificando el predicado indica que la opinión es negativa. En otros casos solamente el contexto podrá ayudar a detectar la ironía.

(1.7) "Cierro los ojos y me parece ver algunos rostros exhalando un profundo suspiro de alivio", *comentó con sorna*.

La negación, aplicada sobre predicados de opinión subjetivos, como *apoyar* o *rechazar*, funciona como un inversor del valor de orientación semántica. Por ejemplo, en (1.8) *no rechazaba* es un predicado positivo, mientras que en (1.9), *rechazaba* es negativo⁵. Esta interpretación tiene más sentido que interpretar la negación como la expresión de un evento no realizado.

(1.8) ... *su país no rechazaba la idea de crear un parque binacional en la zona en disputa*.

(1.9) ... *su país rechazaba la idea de crear un parque binacional en la zona en disputa*.

Por otro lado, para predicados de opinión neutros, como *decir* u *opinar*, la negación sí afecta la factividad: *X no opinó sobre Y* expresa un evento no realizado. Este tema constituye una línea muy interesante a seguir en investigaciones futuras.

1.3 Objetivos del trabajo

Como se explicó, el trabajo que presentamos en este documento aborda el problema de la identificación de opiniones en textos periodísticos en español, centrándonos en la identificación de las expresiones en las cuales el autor del texto transmite palabras o posturas de otras fuentes.

⁵ Es posible interpretar la negación *no rechazar* como una expresión positiva más débil que la afirmación *aceptar*, por lo que la negación actuaría simultáneamente como inversor y debilitador del valor de orientación semántica.

Buscamos identificar los predicados de opinión presentes en el texto y, a partir de ellos, los restantes elementos de la opinión -fuente, asunto y mensaje- y su orientación semántica.

Nos propusimos los siguientes objetivos:

- Elaborar una definición de opinión y establecer los elementos que la componen.
- Definir un modelo teórico que contemple las propiedades sintácticas y semánticas de las expresiones que transmiten opiniones.
- Desarrollar un sistema informático para el reconocimiento automático de las opiniones. Dentro de este objetivo, incluimos como sub-objetivo la investigación de la interacción entre métodos simbólicos y métodos estadísticos.

1.4 Descripción general del trabajo

A partir del análisis de un corpus de prensa del español, teniendo en cuenta los elementos que normalmente son considerados importantes para la definición de la opinión en el área Minería de Opiniones y basándonos en un estudio lingüístico de la expresión de opiniones en español, determinamos nuestra propia definición de opinión (que ya adelantamos en la sección 1.2 de este capítulo), estableciendo los elementos que la componen y sus propiedades.

El estudio lingüístico de la expresión de opiniones dio lugar a la creación de un modelo de estas expresiones. El modelo se centra principalmente en los elementos predicativos que introducen opiniones en los textos, que llamamos predicados de opinión. También se estudia la forma en que se expresan algunos de sus argumentos, los que corresponden a los elementos definidos para la opinión, que llamamos argumentos de opinión. Se trata de un modelo basado en estructuras de rasgos tipificadas (*Typed Feature Structures*, TFS), que intenta modelar las diferentes estructuras sintácticas en las que participan los predicados de opinión y sus argumentos, mostrando los roles semánticos que toman los argumentos según la estructura. Los predicados son en su mayoría verbos pero incluyen también nombres y ciertas locuciones preposicionales. Si bien existen sistemas informáticos que permiten la definición de gramáticas basadas en TFS y su aplicación para el análisis sintáctico de textos, nosotros las utilizamos solo como formalismo para modelar las principales propiedades sintácticas y semánticas de los predicados y argumentos de opinión. No utilizamos, por lo tanto, el modelo basado en TFS como parte de la implementación informática.

Para poder reconocer las opiniones en textos, es necesario contar, como recurso fundamental, con un repertorio de predicados de opinión, por ser estos elementos los que indican la presencia de una opinión en un texto. Creamos nuestro propio repertorio de predicados, incorporando, para cada entrada, información proveniente del modelo teórico definido.

Desarrollamos tres sistemas informáticos para el reconocimiento de las opiniones, los tres utilizan el repertorio de predicados de opinión. El primer sistema desarrollado se basa en la escritura de reglas contextuales, el segundo está basado en métodos estadísticos y el tercero combina la salida del sistema de reglas con métodos estadísticos.

El sistema de reglas contextuales utiliza un intérprete de reglas escrito en Prolog. El sistema produce como salida el texto con etiquetas xml que muestran las opiniones, sus elementos y su orientación semántica, calculada a partir de la orientación semántica de cada elemento.

De todos los fenómenos involucrados con la identificación de opiniones, el sistema de reglas aborda los siguientes aspectos:

- Resuelve la identificación de predicados de opinión verbales, nominales y preposicionales.
- Resuelve la identificación de la fuente de cada opinión, siempre que esta se mencione explícitamente dentro de la cláusula correspondiente a un predicado identificado.
- Resuelve la identificación del asunto en los casos en los cuales este es mencionado en un constituyente específico: grupos preposicionales introducidos por *sobre*, *respecto a*, *en lo referente a* y expresiones similares; grupos nominales que son objeto directo de verbos como *apoyar*, *rechazar*, etc.; complementos del nombre para casos de predicados nominales, como *el rechazo hacia Y*.
- Resuelve la identificación del mensaje en sus diferentes formas (cita directa entre comillas, cita indirecta, expresiones mixtas con fragmentos entre comillas dentro de cita indirecta, entre otros).
- Resuelve la determinación de la orientación semántica del predicado y del asunto, ya que se trata de un valor asociado a las piezas léxicas (por ejemplo, el predicado *apoyar* es positivo, el introductor de asunto *en contra de* es negativo). También se resuelve para algunos casos la determinación de orientación semántica de la fuente.
- Establece qué predicado, fuente, asunto y mensaje corresponden a una misma opinión.
- Calcula la orientación semántica de la opinión completa, a partir de los valores de orientación semántica del predicado, el asunto y la fuente, siempre que alguno de estos tenga un valor no neutro.

El sistema da como salida el mismo texto de entrada al que se le agregan anotaciones para la fuente, el asunto, el mensaje, el predicado y la opinión completa. Cada elemento tiene asociado un atributo, 'os', que indica su valor de orientación semántica. Los posibles valores son: 'pos' (positiva), 'neg' (negativa), 'neu' (neutra) y 'nd' (no determinada). Si alguno de los elementos tiene para este atributo un valor positivo o negativo, la opinión completa también tiene ese valor para el atributo 'os'. El atributo 'os' del mensaje siempre lleva el valor no determinado, ya que en este trabajo no abordamos el estudio de la orientación semántica de este elemento

Ilustramos lo anterior a través de algunos ejemplos. Para las siguientes oraciones:

Mujica respaldó importante inversión minera.

"Si no hay una posible ubicación en la costa atlántica tendríamos que salir por la Laguna Merín y esto impactaría en los costos logísticos, porque el 75% del producto en destino es transporte", señaló Puntigliano.

el sistema genera las salidas:

```
<opinion os="pos">
  <fuente os="neu">Mujica</fuente>
  <predicado os="pos">respaldó</predicado>
  <asunto os="neu">importante inversión minera</asunto>
</opinion>

<opinion os="nd">
  <mensaje os="nd">"Si no hay una posible ubicación en la costa atlántica
  tendríamos que salir por la Laguna Merín y esto impactaría en los costos
  logísticos, porque el 75% del producto en destino es transporte"</mensaje>,
  <predicado os="neu">señaló</predicado>
  <fuente os="neu">Puntigliano</fuente>
</opinion>
```

La evaluación de este sistema arroja los siguientes valores de medida F (media armónica entre *precision* y *recall*), calculados considerando como resultados correctos los solapamientos entre elementos recuperados y elementos a recuperar: para el predicado 92%, para la fuente 81%, para el asunto 75%, para el mensaje 89% y para la opinión completa 85%.

Para la fuente realizamos una segunda evaluación, asumiendo que los predicados de opinión están todos correctamente identificados, obteniendo un 79% de medida F, calculada en forma exacta (teniendo en cuenta solamente elementos exactos correctos, sin incluir solapamientos). Esta segunda evaluación se realizó con el objetivo de poder comparar este resultado con el que arroja nuestro segundo sistema informático.

Por otra parte, en cuanto a la determinación de la orientación semántica de las opiniones recuperadas, no se obtuvieron resultados muy relevantes. Esto se debió a que los elementos tenidos en cuenta para determinar este valor (predicados no neutros como *apoyar*, *rechazar* o introductores de asunto no neutros como *a favor de*, *en contra de*) resultaron tener baja frecuencia en el corpus de testeo; además no se encontraron ocurrencias de fuentes no neutras. Encontramos 25 predicados no neutros en un total de 281 predicados y 3 introductores de asunto no neutros en un total de 74 menciones explícitas al asunto. De todos modos, en el capítulo 8, que describe la evaluación del sistema, hacemos algunos comentarios sobre estos casos de orientación no neutra.

El segundo desarrollo informático, basado en aprendizaje automático, consiste en diferentes experimentos realizados sobre un corpus de entrenamiento, aplicando el modelo Campos Aleatorios Condicionales (traducción del inglés *Conditional Random Fields*, CRF) para el reconocimiento de las fuentes de las opiniones. En el corpus de entrenamiento se anotaron en forma manual las fuentes, como valor de salida del aprendizaje. Además, el corpus contiene varios atributos de entrada, siendo uno de los más relevantes la pertenencia a nuestro repertorio de predicados de opinión. Entre los diferentes experimentos llevados a cabo, el mejor resultado alcanzado para el reconocimiento de fuentes es 76% de medida F (evaluación exacta, sin solapamientos).

Cabe aclarar que, para los experimentos con CRF, elegimos enfocarnos en el reconocimiento de la fuente ya que nuestra principal motivación para este trabajo es resolver la atribución de las expresiones presentes en un texto a sus fuentes correspondientes. Además, consideramos que los resultados obtenidos por el sistema de reglas para la fuente, si bien son buenos, debían ser mejorados.

Descartamos aplicar aprendizaje automático para reconocer el asunto porque encontramos muy poco material en los corpus con los que trabajamos. En cuanto al predicado, los resultados del sistema de reglas son muy buenos y creemos que la mejor forma de elevarlos aún más es incrementando el repertorio de predicados de opinión. Por último, en lo que tiene que ver con el mensaje, creemos que para abordarlo desde un enfoque basado en aprendizaje automático se requeriría la incorporación de atributos sintácticos más sofisticados, ya que se trata de un segmento de texto mucho más extenso que la fuente, con estructura oracional.

Por último, el tercer sistema informático, enfocado al reconocimiento de fuentes, combina la salida del sistema de reglas con las técnicas de aprendizaje automático. Se obtuvieron resultados que superan los valores arrojados por cualquiera de las dos técnicas por separado, alcanzando un 83% de medida F (considerando segmentos exactos). La combinación de las dos técnicas se logra incluyendo un nuevo atributo de entrada en el corpus de entrenamiento que indica para cada palabra si ésta forma parte de una fuente reconocida por el sistema de reglas o no.

1.5 Principales contribuciones

Se propuso un modelo completo para los predicados de opinión del español que creemos puede ser de gran utilidad si se incorpora a alguna gramática de rasgos basada en TFS para este idioma.

Se creó un repertorio de 155 predicados de opinión que tienen asociada la información definida en el modelo teórico. Este repertorio, además de ser de gran utilidad en sistemas informáticos para identificación de opiniones, puede incorporarse al componente léxico de una gramática como la mencionada en el párrafo anterior.

Se desarrolló un sistema de reglas que permite identificar opiniones en textos de prensa en español, dando para cada opinión la fuente, el asunto y el mensaje. El sistema toma como entrada texto pre-procesado por un analizador morfo-sintáctico y genera una salida con formato xml. Este formato de salida facilita posibles procesamientos posteriores por parte de módulos que utilicen la información relativa a las opiniones.

Se mejoró particularmente el resultado para la identificación de la fuente, combinando la salida del sistema de reglas con un clasificador basado en CRF, siendo este proceso fácilmente aplicable a partir de la salida xml del sistema de reglas.

En cuanto a la identificación de fuentes, en particular, los resultados obtenidos por el sistema combinado (83% de medida F exacta) son muy buenos, tomando como referencia los valores de medida F reportados en trabajos aplicados a otros idiomas, como se muestra en la tabla 1.1.

Referencia	Descripción	medida F	Tipo de medida
Bethard et al., 2004	Inglés, opiniones proposicionales (<i>X dijo que ..</i>)	89,50%	no especificado
Choi et al., 2005	Inglés	69,00%	con solapamientos
Choi et al., 2006	Inglés	82,00%	con solapamientos
Wiegand et al., 2010	Inglés	62,60%	exacta
Lu, 2010	Chino	78,00%	con solapamientos
Nuestra propuesta	Español	83,00%	exacta

Tabla 1.1: Resultados obtenidos por diferentes sistemas para la identificación de fuentes.

El único trabajo que logra un resultado mejor que el nuestro es el de Bethard et al. (2004), pero el tipo de expresiones que allí se consideran es más limitado, ya que solamente se trabaja con opiniones proposicionales (opiniones expresadas dentro una subordinada a un verbo de opinión, como *X dijo que ...*). Choi et al (2006) logran un resultado del mismo orden que el nuestro, pero se trata de una medida con solapamientos.

La combinación de la salida del sistema de reglas con un método de aprendizaje basado en CRF fue muy exitosa, ya que se mejoró en 4 puntos el resultado del sistema de reglas y en 7 puntos el del sistema basado en CRF. Una experiencia similar, reportada en (Rei et al, 2010), orientado al reconocimiento de marcas de modalidad (*hedge cues*) y su alcance, logró una mejora de 2 puntos al combinar un sistema de reglas manuales con un clasificador basado en CRF, respecto a los resultados del sistema de reglas. Consideramos que esta forma de combinar los resultados de la aplicación de reglas manuales con un método posterior de aprendizaje automático puede ser muy conveniente para diferentes tareas de PLN.

Se propuso un procedimiento para la determinación de la orientación semántica de la opinión basado en el análisis de la orientación semántica de cada uno de sus elementos: predicado, asunto, fuente y mensaje. Se obtuvieron algunos resultados primarios en cuanto a la orientación semántica del predicado y el asunto que indican que el procedimiento es adecuado.

Durante el desarrollo de nuestro trabajo generamos varios recursos para el español: un repertorio de predicados de opinión con información sintáctico-semántica asociada, un corpus de 13.000 palabras anotado con las opiniones y sus elementos, un corpus de 40.000 *tokens* anotado con predicados de opinión y fuentes.

Queremos destacar que una de las contribuciones principales de nuestro trabajo es el avance en el estudio teórico y el desarrollo de herramientas en el área del PLN para el español. Como veremos en el estudio de trabajos relacionados (capítulo 2), se ha trabajado mucho en identificación de opiniones para otros idiomas, sobre todo para el inglés, pero encontramos muy pocos trabajos para el español, en general centrados principalmente en la determinación de la orientación semántica y emociones de los textos. Nuestra propuesta significa un aporte importante, principalmente en lo que refiere a la atribución de expresiones a sus fuentes.

Algunos avances de este trabajo ya fueron publicados: en *Identification automatique de marques d'opinion dans des textes* (Rosá, 2008) presentamos una descripción del problema y de las primeras reglas para reconocimiento de predicados y fuentes; en *Opinion Identification in Spanish Texts* (Rosá et al, 2010) presentamos el sistema completo de reglas y su evaluación; y en *Comparación de dos métodos para la extracción de opiniones en textos en español* (Rosá et al, 2010) presentamos los dos sistemas desarrollados, el sistema basado en reglas y el sistema basado en CRF, y sus resultados.

Además, el trabajo fue expuesto en diversas jornadas de divulgación: *Tercer Seminario de Cooperación Franco-Sudamericana en Ciencias y Tecnologías de la Información y las Comunicaciones* (Montevideo, 2007), *Jornadas Argentinas de Lingüística Informática: Modelización e Ingeniería* (XI Congreso de la Sociedad Argentina de Lingüística, Santa Fe, 2008) y *Jornadas de Informática e Investigación Operativa* (Instituto de Computación, Facultad de Ingeniería, UDELAR, Montevideo, 2008 y 2010).

1.6 Organización del documento

El documento se organiza en cinco partes, la primera contiene esta introducción, las cuatro partes restantes tienen el siguiente contenido:

Segunda parte: Estudios previos

En esta parte se presentan las investigaciones que realizamos previamente a la elaboración de nuestra propuesta. Se estructura en dos capítulos: el capítulo 2, "Trabajos relacionados", y el capítulo 3, "Marco teórico lingüístico". En el capítulo 2 presentamos los principales trabajos en cuanto a identificación automática de opiniones, centrándonos en los que más se acercan a nuestros objetivos. En el capítulo 3 presentamos un estudio de los mecanismos lingüísticos que permiten la expresión de opiniones en textos, uno de ellos es el discurso reproducido. Se presentan, además, diferentes propuestas sobre clasificación de verbos, analizando en detalle las clases verbales relacionadas con la expresión de opiniones.

Tercera parte: Nuestra propuesta

Aquí describimos nuestra propuesta. Por un lado, en el capítulo 4, "Metodología de trabajo y análisis de corpus", describimos la metodología de trabajo que seguimos, explicamos nuestra definición de opinión y analizamos a fondo las expresiones de opinión en base a ejemplos de corpus. Por otro lado, en el capítulo 5, "Un modelo para los predicados de opinión y sus argumentos", mostramos el modelo completo para los predicados de opinión, basado en TFS.

Cuarta parte: Implementación informática

En esta parte se describen los desarrollos informáticos y las evaluaciones de los mismos. Contiene 5 capítulos. El capítulo 6, "Repertorio de predicados de opinión", describe el repertorio de verbos y nombres de opinión que creamos para este trabajo, el cual incluye información relacionada con el modelo propuesto en el capítulo 5. Este repertorio de predicados es utilizado por los dos sistemas informáticos implementados. El capítulo 7, "Sistema basado en reglas contextuales", describe el sistema de reglas para el reconocimiento de las opiniones y sus elementos. El capítulo 8, "Evaluación del sistema de reglas contextuales", presenta la evaluación de los resultados del sistema. El capítulo 9, "Sistema basado en aprendizaje automático", muestra los experimentos realizados con la técnica CRF y los resultados obtenidos en cada experimento. El capítulo 10, "Comparación y combinación de los dos sistemas", muestra el experimento que incluye la salida del sistema de reglas como un dato más para el sistema basado en CRF.

Quinta parte: Conclusiones

Finalmente, en esta parte se presentan las conclusiones, en el capítulo 11, y las extensiones y los trabajos futuros, en el capítulo 12.

SEGUNDA PARTE: Estudios previos

2. Trabajos relacionados

Estudiamos varias propuestas sobre identificación de expresiones de diferentes fuentes en los textos.

En general, las propuestas de identificación de fuentes se enmarcan en el área Minería de Opiniones (*Opinion Mining*) o Análisis de Sentimientos (*Sentiment Analysis*), por lo que se centran principalmente en la detección de expresiones subjetivas, para luego buscar sus fuentes. En algunos de estos trabajos se busca también reconocer el asunto de las opiniones.

Otras propuestas abordan más específicamente la identificación del discurso reproducido, para el cual se determinan las fuentes y las palabras citadas, sin importar si se trata de expresiones subjetivas u objetivas.

En cuanto a la orientación semántica, existen numerosos trabajos dedicados a la clasificación de textos según su subjetividad: por un lado se clasifican los textos en objetivos o subjetivos; por otro lado, los textos subjetivos se clasifican en positivos o negativos. Otras propuestas estudian la subjetividad a nivel del párrafo, la oración o la cláusula.

En lo que sigue, presentamos, en primer lugar, un trabajo de anotación de opiniones y emociones en textos periodísticos. Se trata de una propuesta que tuvo gran repercusión en los trabajos posteriores de identificación de opiniones a nivel de la oración. Por un lado, se establece un esquema completo para anotar las opiniones, determinando de este modo los elementos a tener en cuenta en su identificación. Estos conceptos fueron retomados por varias propuestas, incluyendo la nuestra. Por otro lado, el corpus anotado que se generó fue utilizado para el desarrollo de diversos sistemas informáticos relacionados con el reconocimiento de las opiniones y sus elementos.

Luego, hacemos una revisión de diferentes trabajos, presentando en primer lugar los más cercanos a nuestra propuesta, es decir, los que buscan identificar opiniones de diferentes fuentes. Abarcamos tanto trabajos centrados en las opiniones subjetivas como los que abordan el discurso reproducido en general, sin prestar atención al tema de la subjetividad. Nos interesa principalmente estudiar la gama de expresiones de opinión abarcada en cada propuesta, comparándola con el alcance de nuestro trabajo. También comentamos los métodos aplicados y los resultados obtenidos en cada caso. Todos los trabajos comentados son para textos periodísticos.

También hacemos un breve repaso de los trabajos orientados al estudio de la subjetividad.

Finalmente, hacemos algunas consideraciones sobre los trabajos existentes para el español.

2.1 Trabajos sobre identificación de las opiniones y sus componentes en textos de prensa

Comentamos algunos trabajos que abordan la misma problemática que el nuestro y constituyen antecedentes importantes de nuestra propuesta.

Un esquema de anotación para opiniones

El primer trabajo que queremos destacar es el de Wiebe, Wilson y Cardie (2005), por ser uno de los principales referentes del área. Las autoras proponen un esquema de anotación para opiniones y emociones, llamadas en ese marco *private states* (estados internos), y llevan adelante un proceso de anotación de un corpus de 535 documentos de prensa, el corpus MPQA (*Multi-Perspective Question Answering*).

Las expresiones que estas autoras incluyen en su esquema son: menciones explícitas a estados internos (*EEUU teme un derrame*), discurso reproducido subjetivo (*"El reporte es absurdo", dijo el presidente*), elementos expresivos subjetivos (*El reporte es absurdo*) y discurso reproducido objetivo (*El sargento dijo que el evento tuvo lugar en la noche*).

El esquema de anotación que se propone incluye los elementos siguientes:

- indicador en el texto (*text anchor*): porción de texto que indica la ocurrencia de un estado interno
- fuente (*source*): persona o entidad que expresa su estado interno
- asunto (*target*): objeto del estado interno
- propiedades:
 - intensidad del estado interno: intensidad dada por las palabras utilizadas para expresar el estado interno, los valores posibles son: *bajo, medio, alto, extremo*. Se utiliza sólo si hay elementos subjetivos como *absurdo, fraude*, etc.
 - intensidad de la expresión: intensidad del elemento utilizado para introducir el estado interno, los valores posibles son: *neutro, bajo, medio, alto, extremo*. Por ejemplo, *decir* es neutro, *exhortar* es alto.
 - insustancial: toma un valor verdadero si el estado interno no es sustancial en el discurso, es decir, si no es real (es hipotético o presentado como no real) o no es significativo.
 - tipo de actitud: polaridad del estado interno, los valores posibles son: *positivo, negativo, otro, no corresponde*.

Para discurso reproducido subjetivo y menciones explícitas a estados internos, se utilizan todos los elementos del esquema. Para elementos expresivos subjetivos se utilizan todos, salvo tema, intensidad de la expresión e insustancial. Para discurso reproducido objetivo se usan solamente texto, fuente y tema.

Como se puede ver, los casos de discurso reproducido, tanto subjetivos como objetivos, y las menciones explícitas a estados internos son atribuibles siempre a fuentes citadas por el autor del texto. Todos estos casos están incluidos entre las opiniones que abarca nuestro trabajo. No queda claro en la propuesta de Wiebe, Wilson y Cardie (2005) cómo se analizan las expresiones introducidas por *según* o similares; las autoras no incluyen en su trabajo ejemplos de este tipo de expresión.

Las expresiones que las autoras llaman elementos expresivos subjetivos no necesariamente deben atribuirse a diferentes fuentes. Cuando se trata de expresiones del autor del texto no son abarcadas por nuestro trabajo. Cuando estos elementos expresivos son emitidos por alguna fuente citada por el autor del texto, se encuentran dentro del elemento que nosotros llamamos mensaje, como la palabra *absurdo* en "*El reporte es absurdo*", dijo el presidente. Por lo tanto, desde nuestro punto de vista, nos encontramos ante una opinión citada por el autor del texto, en cuyo mensaje existen elementos subjetivos que le darán un valor de orientación semántica no neutra.

En la propuesta de Wiebe, Wilson y Cardie, se realiza un análisis diferente para estos casos. Se considera que se trata de dos opiniones (o estados internos): una opinión de tipo discurso reproducido subjetivo cuyo elemento indicador es el verbo de reporte (*dijo*), y otra de tipo elemento expresivo subjetivo cuyo elemento indicador es la palabra subjetiva *absurdo*. Las dos opiniones tendrían la misma fuente (*el presidente*).

Por otra parte, cabe mencionar que, si bien el trabajo de Wiebe, Wilson y Cardie (2005) se centra en el estudio de la subjetividad, se incluyen en su esquema de anotación los casos de discurso reproducido de contenido objetivo. Esta decisión es justificada por el hecho de que se trata de expresiones de fuentes diferentes del autor del texto. En este punto nuestra propuesta se asemeja a la de estas autoras, ya que incluimos entre las opiniones a todas las expresiones de otras fuentes, sean subjetivas o no. En particular, las de contenido objetivo serán para nosotros opiniones neutras.

En relación a la distinción entre opiniones objetivas y subjetivas, creemos que la decisión no siempre es del todo clara. En el ejemplo "*EEUU teme un derrame*", dijo Xirao-Nima⁶, incluido en (Wiebe et al, 2005), las autoras marcan dos opiniones: una de tipo discurso reproducido objetivo, con fuente *Xirao-Nima*; y otra de tipo mención explícita a un estado interno, con fuente *EEUU*. No nos parece tan claro que se trate de información objetiva la expresión entrecomillada atribuida a *Xirao-Nima*. La mención a un estado interno subjetivo de otra fuente es de por sí una consideración subjetiva, salvo que la otra fuente exprese verbalmente su estado interno, pero en este ejemplo no parece ser así.

El corpus MPQA, anotado según el esquema descripto, ha sido ampliamente utilizado en diversas aplicaciones: estudio de la utilidad de las anotaciones del MPQA para sistemas Q&A⁷ con preguntas sobre opiniones (*¿Qué opina X sobre Y?*, *¿Quién está a favor de Y?*) (Stoyanov et al, 2005); identificación del grado de intensidad de las opiniones (Wilson et al, 2006); determinación de la polaridad (orientación semántica) de opiniones (Choi et al, 2008); identificación de las fuentes de opiniones (Choi et al, 2005, Kim et al, 2005, Wiegand et al, 2010); entre otras. Los detalles de la versión actual del corpus pueden leerse en (WebMPQA).

Sistemas para la identificación automática de la opinión y sus elementos

Extracción de opiniones proposicionales y sus fuentes (2004, 2005)

En (Bethard et al, 2004, 2005) se presenta un sistema para la extracción de opiniones proposicionales y sus fuentes (*opinion holders*). Llamamos opiniones proposicionales a las proposiciones subordinadas a verbos como *creer*, *declarar*, etc. que transmiten un contenido

6 El ejemplo original es: "*The US fears a spill-over*", said Xirao-Nima.

7 Q&A: Question Answering (Respuestas a preguntas)

subjetivo. Se trata de un elemento similar a lo que en nuestro trabajo llamamos mensaje, salvo que se excluyen las proposiciones de contenido objetivo.

En base a un corpus integrado por oraciones con argumentos proposicionales, tomadas de FrameNet (Baker et al, 1998) y PropBank (Palmer et al, 2005), con anotaciones agregadas que indican si las proposiciones contienen opiniones o no y cuáles son sus fuentes, se entrenaron dos sistemas. El primer sistema, basado en SVM (*Support Vector Machine*), determina si las proposiciones contienen opiniones y además identifica sus fuentes. El segundo sistema combina una primera etapa basada en SVM, solo para segmentar las proposiciones subordinadas, con una segunda etapa basada en Naive Bayes para determinar cuáles proposiciones contienen opiniones. Entre los atributos utilizados para el aprendizaje, destacamos la utilización de un conjunto de palabras subjetivas, creado a partir de un conjunto existente que fue extendido para este trabajo.

Los resultados obtenidos para la identificación de opiniones proposicionales son 51% de *recall* y 58% de *precision* para el primer sistema y 43% de *recall* y 68% de *precision* para el segundo sistema. El reconocimiento de la fuente, para las opiniones proposicionales ya marcadas, solo se evaluó con el primer sistema, dando muy buenos resultados: 89% de *recall* y 90% de *precision*. Suponemos que se trata de medidas exactas, sin considerar solapamientos.

En algunos aspectos nuestro trabajo se asemeja al de (Bethard, 2004, 2005), ya que se basa en la presencia de un verbo de opinión para identificar luego su fuente y la proposición que contiene la opinión (mensaje). A diferencia de ellos, nosotros tenemos en cuenta no solo verbos que introducen opiniones sino también nombres y preposiciones como *según*. Pero, por otro lado, ellos diferencian proposiciones en donde hay un juicio subjetivo de aquellas en las que se transmiten contenidos objetivos. Nosotros no hacemos esta diferencia, postergando esta distinción para una etapa posterior de análisis de la orientación semántica del mensaje.

Identificación de fuente (2005)

Choi et al. (2005) aplicaron una combinación de dos métodos de aprendizaje automático, CRF y patrones de extracción, sobre el corpus MPQA para identificar las fuentes (*source*) de opiniones. Los atributos utilizados para la aplicación de CRF son: palabras en mayúscula, categoría gramatical, pertenencia a un conjunto de palabras que indican opinión, información sintáctica generada por la aplicación de un *parser* y clase semántica (humano, medio de prensa, organización, etc.). Además se agregan como atributo adicional los patrones generados por un método automático que genera patrones de extracción.

El sistema fue evaluado según tres criterios, damos para cada uno la medida F obtenida: identificación exacta (*exact match*), 62%; identificación del núcleo (*head match*), 68%; identificación solapada (*overlap match*), 69.4%.

Como el trabajo se apoya en las anotaciones del MPQA, podemos suponer que se tienen en cuenta todos los tipos de opiniones que allí se anotaron. Por lo tanto, se abarca una variedad mayor de expresiones de opinión respecto a lo que reconocen (Bethard et al, 2005), que solo buscan las fuentes de opiniones proposicionales. También se abarcan más tipos de expresiones que en nuestra propuesta. La diferencia principal, como ya dijimos al comentar el esquema de (Wiebe et al, 2005), está en los elementos expresivos subjetivos.

De todos modos, quedan algunas dudas sobre cuáles son las expresiones que efectivamente abarca el trabajo descripto. Posiblemente no se incluyan los casos de discurso reproducido objetivo, ya que se hace mención permanentemente a expresiones de opinión. Por otro lado, no

se explica cómo son considerados los casos de elementos expresivos subjetivos, que no necesariamente tienen una fuente distinta del autor del texto. Este último no puede ser anotado como fuente con el formato de anotaciones B-I-O⁸, ya que no es mencionado explícitamente en el texto.

Identificación de expresiones de opinión, fuentes y relaciones entre ellas (2006)

En el trabajo de (Choi et al, 2006), además de identificar expresiones de opinión y fuentes, se determina la relación entre las expresiones y las fuentes identificadas. Trabajan en 3 fases: primero se entrenan dos clasificadores independientes, basados en CRF, para el reconocimiento de las expresiones de opinión, por un lado, y de las fuentes, por otro. Luego se entrena un clasificador binario basado en el modelo de Máxima Entropía que determina los pares (expresión-de-opinión, fuente) que están relacionados, es decir, que cumplen que la opinión transmitida por la expresión-de-opinión debe atribuirse a la fuente.

Los atributos utilizados para los clasificadores basados en CRF son similares los que se utilizaron en (Choi et al, 2005), comentado anteriormente, excluyendo los patrones de extracción y agregando información proveniente de WordNet.

Para controlar la formación de pares consistentes y seleccionar los mejores pares (expresión-de-opinión, fuente) se implementó un procedimiento basado en programación lineal entera que permite establecer restricciones como: cada opinión se atribuye a una única fuente, una fuente puede estar asociada a más de una opinión, la cantidad total de pares es igual a la cantidad de expresiones de opinión, entre otras.

Entre diferentes pruebas que se realizaron, los mejores resultados reportados para medida F, considerando solapamiento, son los siguientes: 68.9% para la extracción de pares (expresión-de-opinión, fuente); 82.0% para las expresiones de opinión; y 77.3% para las fuentes.

Identificación de fuente y asunto (2006)

Otro trabajo muy cercano al nuestro es el de Kim y Hovy (2006), quienes buscan identificar fuentes (*opinion-holder*) y asunto (*topic*) en artículos de prensa de Internet. En primer lugar se detectan palabras indicadoras de opinión para lo cual se crea un conjunto de verbos (69 positivos y 151 negativos) y adjetivos (199 positivos y 304 negativos). Luego se etiquetan con roles semánticos las oraciones que contienen ocurrencias de esas palabras. Para esto se identifica el *frame* correspondiente a la palabra indicadora de opinión y se buscan los elementos (roles) que el *frame* tiene asociados en FrameNet (Baker et al, 1998). Finalmente se establece cuál elemento de cada *frame* corresponde a la fuente y cuál al asunto.

El etiquetado con los elementos de cada *frame* es la parte central del trabajo y se resuelve entrenando un clasificador, basado en el modelo de Máxima Entropía, sobre cerca de 20.000 oraciones tomadas del corpus de FrameNet, previamente procesadas por el parser de Charniak. Entre los atributos utilizados para el aprendizaje se incluye: cuál es la palabra indicadora de opinión, el *frame* correspondiente a esa palabra y varios atributos sintácticos (categoría y núcleo de cada constituyente, posición respecto al indicador de opinión dentro del árbol sintáctico, entre otras).

8 Las etiquetas B-I-O se explican en la sección 9.2 de este documento.

El sistema fue evaluado en primera instancia sobre el 10% del corpus armado a partir de FrameNet. Se realizó una segunda evaluación sobre un conjunto de 100 oraciones extraídas de textos de prensa de Internet, anotados por dos humanos, con un 82% de coincidencia entre anotadores. Los valores de medida F obtenidos en la segunda evaluación (suponemos que se trata de medidas exactas) son 61.7% para la detección de opiniones (en base a la presencia de palabras indicadoras de opinión); 39.8% para la fuente; 31.5% para el asunto. Los resultados no alcanzan altos porcentajes de acierto pero superan las medidas base consideradas por los autores.

El trabajo abarca expresiones similares a los elementos expresivos subjetivos de (Wiebe et al, 2005), ya que entre los indicadores de opinión, además de verbos, se incluyen adjetivos. La forma en que se marca la fuente, cuando esta coincide con el autor del texto, para las opiniones indicadas por adjetivos no se explica en el artículo. Una vez más destacamos que este tipo de expresión, nucleada por adjetivos subjetivos, no es abarcado por nuestra propuesta.

Identificación del asunto (2008)

Stoyanov et al (2008) extienden el corpus MPQA anotando el asunto (*topic span*) de las opiniones para lo cual redefinen el concepto de *target* utilizado en el esquema de anotación aplicado a dicho corpus (Wiebe et al, 2005). Estos autores definen *topic span* como el mínimo segmento de texto que hace referencia al objeto, evento o entidad sobre el cual trata la opinión, en el contexto particular en el cual esta es emitida; y *target span* como el componente sintáctico que cubre el contenido de la opinión. Este último concepto abarca tanto lo que nosotros llamamos asunto como lo que llamamos mensaje.

Los autores proponen un algoritmo para identificar opiniones con el mismo tópico (*topic-coreference*). El algoritmo, que determina para cada opinión el *cluster* al cual pertenece (constituido por opiniones con igual tópico), está basado en un algoritmo anterior para correferencias de grupos nominales que aplica aprendizaje automático.

Identificación de fuente (2010)

Otra propuesta de identificación de fuentes de opiniones (*opinion holder*) es la de Wiegand y Klakow (2010) que utilizan *Convolution Kernels*. El aprendizaje se realiza en base a los árboles sintácticos de las oraciones y diferentes atributos léxicos y sintácticos: entidades con nombre, roles semánticos, expresiones de opinión y palabras de comunicación (*decir, informar, etc.*). Algunos de los atributos se toman de (Choi et al, 2005), comentado anteriormente.

Se utilizó como corpus de entrenamiento el MPQA sobre el cual se aplicó un analizador sintáctico y un etiquetador de roles semánticos.

De las diferentes pruebas realizadas aplicando diferentes *kernels*, que utilizan de manera diferente la información de los atributos del corpus de entrenamiento, los mejores resultados reportados alcanzan un 62.6% de Medida_F (medida exacta, sin solapamientos).

Identificación de fuente y asunto (2010)

Lu (2010) también trabaja en la identificación de la fuente y el asunto de las opiniones, pero para el idioma chino. A diferencia de la gran mayoría de los trabajos estudiados, en este caso se presenta un sistema basado en reglas. Las reglas se aplican sobre los textos previamente analizados con un *parser* de dependencias.

Para la identificación de la fuente y el asunto de opiniones se buscan ocurrencias de verbos de reporte (*decir, pensar*) y palabras subjetivas (*bueno, malo*). Para esto, se utilizan diferentes recursos léxicos existentes: por un lado, se cuenta con un repertorio de 68 verbos de reporte, extendido a 308 mediante la incorporación de sinónimos (utilizando WordNet y recursos similares); por otro lado, se utiliza un lexicón de 5046 palabras positivas y otro de 3499 palabras negativas del chino.

Para la escritura de las reglas se asume que la fuente es gobernada por los verbos de reporte y el asunto por las palabras subjetivas. Se generan expresiones candidatas a ser fuente (sujeto de verbo de reporte, nombre propio precediendo dos puntos, candidato a fuente de una opinión anterior, entre otros). Una vez identificadas las fuentes, se identifica el asunto de cada opinión, siguiendo reglas como por ejemplo, el asunto es el sujeto de una proposición subordinada a un verbo de reporte, o el asunto es una palabra que depende de una palabra subjetiva.

Se evalúa el sistema en base a identificación exacta, identificación del núcleo e identificación solapada, medidas propuestas por (Choi et al, 2005). Los valores de medida F obtenidos son: 68.5% (exacta), 73% (núcleo) y 78.4% (solapada) para la fuente y 28.9% (exacta), 38.2% (núcleo) y 59% (solapada) para el asunto.

El sistema descrito participó en el evento NTCIR-8 MOAT (Web NTCIR-8), en donde tuvo un buen desempeño para el reconocimiento de opiniones en textos en chino: obtuvo el primer lugar en el reconocimiento de oraciones que expresan una opinión; el segundo lugar en el reconocimiento de fuente y asunto; y una posición intermedia en cuanto a la determinación de la orientación semántica de las opiniones (Lu et al, 2010b).

Es importante destacar que la identificación del asunto apunta a identificar el asunto dentro de lo que nosotros llamamos mensaje. En nuestra propuesta nos limitamos a identificar menciones explícitas al asunto, es decir, expresiones introducidas por *sobre, en lo referente a*, etc.

En lo que tiene que ver con la identificación de la fuente, las expresiones que se incluyen en esta propuesta son abarcadas por nuestro trabajo, ya que son las nucleadas por verbos de reporte, que es uno de los posibles predicados de opinión que tenemos en cuenta.

Sistemas para la identificación de discurso reproducido

Identificación de fuente, verbo de reporte y cita reproducida (2008)

En (Krestel et al, 2008) se presenta un sistema informático para el inglés que identifica, para cada caso de discurso reproducido (discurso directo y discurso indirecto), la fuente, el verbo de reporte y la cláusula reportada. Para esto se implementaron dos componentes de la plataforma GATE (WebGATE).

El primer componente es un marcador de verbos de reporte que reconoce los verbos de reporte más frecuentes, como *say, report, tell, believe, criticize*, en base a una lista de 54 elementos. El segundo componente aplica transductores para el reconocimiento de patrones como fuente+verbo+cláusula, verbo+fuente+cláusula, entre otros.

Se presenta una evaluación realizada sobre un conjunto de artículos de prensa de unas 6100 palabras. Se consideran correctos los elementos marcados completamente y también algunos elementos marcados en forma parcial, siempre que se mantenga el sentido del contenido reportado. Se presentan valores de *precision, recall* y medida F por cada artículo. Para el

reconocimiento de la cláusula reportada, se obtienen valores entre 96% y 100% de *precision* y valores entre 50% y 100% de *recall*, con una medida F promedio de 84%. Para el reconocimiento del verbo de reporte y la fuente, se obtiene un 100% de *precision* en todos los artículos y valores entre 58% y 100% de *recall*, con una medida F promedio de 87%.

Identificación de discurso directo para diferentes idiomas (2007)

En (Pouliquen et al, 2007) se describe el trabajo realizado para implementar el módulo que muestra las citas de personas frecuentemente mencionados en noticias dentro del sitio web News Explorer (WebNewsExplorer). Se extraen citas para once idiomas diferentes, entre los cuales se encuentra el español. Al momento de la publicación (2007), el sistema extraía en promedio unas 2600 citas por día.

Se reconocen solamente citas directas, en base a un conjunto de reglas que buscan diferentes patrones, que incluyen en todos los casos un verbo de discurso, una fuente (persona pública perteneciente a una base de datos que se actualiza a diario) e indicadores de cita (variantes tipográficas de la comilla). Se resuelven correferencias entre nombres propios, por ejemplo, una mención a una persona por nombre y apellido y otra mención a la misma persona usando solo el apellido. No se resuelven anáforas ni sujeto omitido.

El sistema fue evaluado sobre una pequeña cantidad de citas (55) extraídas para el inglés. El *recall* dentro de un mismo texto es de 54%, pero, como el sistema es capaz de procesar una gran cantidad de artículos diarios (20.000 en total), las citas que no son encontradas en algún artículo, suelen ser encontradas en otros. La *precision* es de 87.5%. También se evaluaron los resultados del sistema para otros idiomas, pero en el artículo no se presentan los valores obtenidos.

2.2 Minería de opiniones - Análisis de sentimientos

El área Minería de Opiniones o Análisis de Sentimientos⁹ es presentada detalladamente en el trabajo de Pang y Lee (2008) y en el capítulo *Sentiment Analysis and Subjectivity*, del libro *Handbook of Natural Language Processing* (Indurkha y Damerau, 2010), escrito por Liu (2010). Presentamos una descripción breve del área en base a estos autores.

Existen diferentes abordajes al problema de la identificación de opiniones, centrados en diferentes tareas, todas relacionadas entre sí:

- Determinar si un documento es subjetivo u objetivo.
- Determinar porciones subjetivas dentro de un documento.
- Determinar la orientación semántica, positiva o negativa, global de un documento.
- Determinar la orientación semántica a nivel del párrafo, de la oración, o incluso de la proposición.
- Extraer opiniones identificando la fuente y el asunto de cada una.

Algunas de las tareas anteriores corresponden al campo de la clasificación de textos, dado que se clasifican los textos según su subjetividad o según su orientación semántica. Otras tareas

⁹ Pang y Lee (2008) discuten las diferentes denominaciones que ha recibido esta área (*Opinion Mining*, *Sentiment Analysis*, *Subjectivity Analysis*, *Review Mining*, *Appraisal Extraction* y *Affective Computing*) indicando el origen de cada una. Ellos resuelven utilizar los términos *Opinion Mining* o *Sentiment Analysis*, en forma indistinta.

corresponden más bien al campo de la extracción de información, ya que se extraen de los textos las opiniones, con sus elementos correspondientes. En este sentido podemos hablar de extracción de opiniones (Kobayashi, 2005).

El dominio sobre el cual se trabaja determina el tipo de tarea que tiene relevancia. Existen muchas propuestas, tanto para clasificación de documentos como para extracción de opiniones, que trabajan sobre comentarios en páginas de Internet, del estilo de blogs o foros. En estos casos, parte de la información que interesa suele estar preestablecida, por ejemplo, es normal que el foro esté destinado a comentar sobre algún tema particular. También es posible encontrar parte de la información registrada en forma estructurada, por ejemplo, es común que el autor de la opinión ingrese su identificación en algún campo específico. Normalmente no es relevante identificar al autor de cada comentario por tratarse de usuarios de algún producto o clientes de alguna empresa que no interesa individualizar.

En las propuestas que trabajan sobre textos periodísticos, en cambio, toma mayor relevancia la identificación del autor de cada opinión, dado que se trata por lo general de personajes públicos. Además, en un texto periodístico, el autor del texto suele citar o hacer referencia a opiniones de diversas personas, lo cual agrega complejidad a esta tarea.

El estudio de la subjetividad está relacionado con la Teoría de la Evaluación (*Appraisal Theory*) (Martin et al, 2005), desarrollada en el marco de la Lingüística Sistémico-Funcional (Halliday, 1985). Esta teoría analiza la forma en la cual los usuarios del lenguaje comunican evaluaciones. Esto incluye la expresión de actitudes (emociones, juicios sobre personas, apreciación de objetos), compromiso (valoración de opiniones de otras personas) y gradación, tanto de las actitudes como del compromiso (Read et al, 2007).

En este marco, se propone una taxonomía en la cual se incluyen en un primer nivel los tres sub-sistemas mencionados (actitud, compromiso y gradación), especificándose cada uno en hasta cuatro niveles más. Para la actitud, por ejemplo, se define un primer nivel con tres elementos: afecto, juicio y apreciación. Luego, para afecto se definen los sub-niveles inclinación, felicidad, seguridad, satisfacción; para juicio se definen los sub-niveles estima y sanción; y para apreciación se definen reacción, composición y valoración. Además, se define un sub-nivel más para algunos de los sub-niveles anteriores.

En el marco del análisis de opiniones, se trabaja normalmente en el estudio de las actitudes y su gradación, buscando clasificar expresiones de emociones, sentimientos, juicios, etc., en positivas o negativas. El análisis completo de textos, de modo de reconocer todos los fenómenos que implica la teoría de la evaluación, incluyendo la determinación de los diferentes niveles que componen la taxonomía, es muy difícil de llevar a cabo con procesamientos automáticos. Con el objetivo de empezar a avanzar por este camino, (Read et al., 2007) trabajaron en la creación de un corpus anotado con esta información. Los valores de concordancia entre anotadores obtenidos revelan la dificultad de la tarea: 0.698 de medida F promedio para los niveles de mayor abstracción de la taxonomía y 0.395 para los niveles más concretos.

En los trabajos centrados en la subjetividad, tanto a nivel del documento como a niveles menores (párrafo, oración, proposición), es de gran importancia la utilización de diccionarios afectivos que permitan identificar palabras de contenido subjetivo en el texto. Por lo tanto, la creación de estos diccionarios constituye una sub-tarea de gran importancia dentro del área. Estos recursos suelen diferenciar palabras positivas de negativas, de modo que son de utilidad también para la determinación de la orientación semántica de los textos subjetivos.

Varios autores trabajaron en la clasificación de palabras según su orientación semántica y en la creación de diccionarios afectivos, principalmente para el inglés. Algunos de los más relevantes son Hatzivassiloglou et al (1997), Turney et al (2003), Riloff et al (2003), Wiebe et al (2004), Esuli (2006).

En base a las palabras positivas y negativas de estos diccionarios afectivos se calcula la orientación semántica de la opinión (Turney, 2002, Pang et al, 2002, Kim et al, 2004), en general en base a la frecuencia de palabras con cada valor subjetivo (positivo o negativo).

Algunos autores estudiaron la incidencia sobre las palabras subjetivas de ciertos elementos que funcionan como modificadores de la valencia (Polanyi et al, 2004, Wilson et al, 2005, Moilanen et al, 2007, Choi et al, 2008).

Por otro lado, Asher et al (2008, 2009) estudiaron la polaridad y la intensidad que resultan de la combinación de diferentes proposiciones que contienen opiniones, según el tipo de relación retórica que se establece entre ellas (continuación, contraste, explicación, etc.). El cálculo de los valores resultantes se realiza mediante la aplicación de reglas como la que sigue: si entre dos proposiciones con polaridades opuestas se establece una relación de contraste, la combinación resultante mantiene la polaridad de la primera proposición, pero el valor de intensidad se debilita por el efecto de la segunda proposición. En el ejemplo (tomado de los autores) [*I agree with you*], [*even if I was shocked*], como la primera proposición es positiva, la segunda proposición es negativa y la relación retórica que se da entre ellas es contraste, la oración resultante es positiva con una intensidad menor a la de la primera proposición. Los autores trabajaron en la anotación manual de un corpus, aplicando estas ideas, con el objetivo de desarrollar en el futuro un procesamiento automático.

2.3 Trabajos para el español

Los trabajos en Minería de Opiniones para el español son escasos. No encontramos trabajos que aborden la identificación de los elementos de la opinión para el español, pero existen algunas propuestas que construyen recursos léxicos afectivos y buscan determinar la orientación semántica o emociones en textos.

Una de las propuestas (García et al, 2008) calcula la emoción (sorpresa, alegría, tristeza, enfado, etc.) global transmitida en un texto. El trabajo es para el español y para el inglés. Se utiliza un recurso léxico para el inglés, ANEW (Bradley et al, 1999), y su adaptación al español, ANSW (Redondo et al, 2007). El recurso cuenta con alrededor de 1000 palabras que tienen asignados valores para diferentes dimensiones emocionales. En base a las palabras de ANEW o ANSW se aplica un algoritmo de desambiguación semántica y luego, solo para la versión en inglés, se crea un conjunto más extenso de palabras incluyendo hiperónimos de WordNet. Finalmente, combinando los valores asociados a las palabras del texto que pertenecen al conjunto creado y aplicando algunos modificadores de polaridad (como negadores), se calcula la emoción global del texto.

En otra propuesta (Moreno et al, 2010) se determina la polaridad de segmentos de texto y una polaridad global para el texto completo en base a las polaridades de ciertas palabras, pertenecientes a una base de datos. La base de datos, creada para este trabajo, contiene 27.000 palabras y locuciones con sus polaridades (valores -2, -1, 1 y 2) determinadas manualmente y un

conjunto de reglas de contexto que permiten la aplicación de modificadores de polaridad (intensificadores, inversores, etc.) a las palabras de la base de datos que ocurren en los textos.

Se ha trabajado en la traducción automática de los recursos léxicos o los corpus anotados del inglés, de modo de aplicarlos para otros idiomas (entre ellos el español) y también en la posibilidad de traducir los textos del español al inglés y luego aplicar sobre ellos los sistemas existentes (Brooke et al, 2010, Banea et al, 2008). Se trata de trabajos orientados a determinar la orientación semántica de los textos. Los resultados obtenidos son interesantes pero igualmente en (Brooke et al, 2010) los autores concluyen que la forma de obtener mejores resultados es trabajar en recursos específicos para cada idioma.

2.4 Síntesis

En lo que tiene que ver con la identificación de opiniones, uno de los principales referentes es el esquema de anotación para opiniones y emociones de Wiebe, Wilson y Cardie (2005). Este esquema establece diferentes tipos de expresiones a tener en cuenta para el estudio de las opiniones. Nosotros hemos considerado en nuestro estudio la mayoría de estas expresiones: discurso reproducido subjetivo, discurso reproducido objetivo y menciones explícitas a estados internos.

En este trabajo se apoya nuestra decisión de identificar todos los casos de discurso reproducido, incluyendo expresiones de carácter objetivo, por más que estas últimas pueden no considerarse estrictamente opiniones. Las autoras mencionadas también incluyen estas expresiones en su trabajo, como un tipo especial de opinión: el discurso reproducido objetivo.

Entre los trabajos que se sitúan dentro del área Minería de Opiniones, estudiamos especialmente los que buscan identificar la fuente (Bethard et al, 2004, 2005; Choi et al, 2005, 2006; Kim et al, 2006; Lu, 2010; Wiegand et al, 2010) y el asunto (Kim et al, 2006; Stoyanov et al, 2008; Lu, 2010) de las opiniones. La gran mayoría de las propuestas utilizan métodos de aprendizaje automático (Bethard et al, 2004, 2005; Choi et al, 2005, 2006; Kim et al, 2006; Stoyanov et al, 2008, Wiegand et al, 2010). Solo uno de los trabajos analizados (Lu, 2010) desarrolla un sistema de reglas, con el cual obtiene resultados un poco mejores que la mayoría de los antes citados. Algunos de estos trabajos (Kim et al, 2006; Wiegand et al, 2010) utilizan un etiquetador de roles semánticos entre sus recursos. Esta es una herramienta que otros autores (Ruppenhofer et al, 2008; Sauri, 2008) consideran de gran utilidad para la identificación de fuente y asunto, aunque no han trabajado en el desarrollo de herramientas para el reconocimiento automático de estos elementos. Para el español no contamos con este recurso.

Por otro lado, también analizamos propuestas para el reconocimiento del discurso reproducido, ya que se trata del mecanismo más habitual para la citación de opiniones de otros. Las dos propuestas analizadas (Krestel et al, 2008; Pouliquen et al, 2007) se basan en la escritura de reglas. En el primer caso se detectan los casos de discurso reproducido, indicándose la fuente, el verbo de reporte y la cláusula reportada, con buenos resultados. En el segundo caso, solo se identifican los casos de discurso directo y se presenta una evaluación muy acotada.

Mencionamos brevemente algunos de los principales trabajos que se han desarrollado para la determinación de la orientación semántica (Turney, 2002; Pang et al, 2002; Kim et al, 2004; Wilson et al, 2005; Choi et al, 2008; Asher et al, 2009) y para la elaboración de diccionarios

afectivos (Hatzivassiloglou et al, 1997; Turney et al, 2003, Rilloy et al, 2003; Wiebe et al, 2004), por tratarse de las líneas más exploradas dentro del área Minería de Opiniones.

Ninguno de los trabajos previos fue desarrollado para el español¹⁰. Analizamos algunos pocos sistemas para este idioma, orientados al estudio de la afectividad. En este campo, una de las principales dificultades que enfrentan las aplicaciones para el español es la carencia de recursos léxicos. García et al (2008) trabajaron con ANSW (Redondo et al, 2007), un recurso léxico que cuenta con alrededor de 1000 palabras con valores emocionales asociados. Otro trabajo (Moreno et al, 2010) utiliza un recurso de tamaño considerablemente mayor (27.000 expresiones anotadas con valores de polaridad) pero no hay información sobre la disponibilidad de dicho recurso.

¹⁰ Salvo el sistema de (Pouliquen et al, 2007) que es para varios idiomas, entre ellos el español, pero solamente aborda el reconocimiento de citas directas y sus fuentes.

3. Marco teórico lingüístico

Presentamos en este capítulo las características lingüísticas de la opinión. Para comenzar, hacemos una revisión del fenómeno del discurso reproducido y luego estudiamos los diferentes predicados de opinión, que son el elemento central de la opinión ya que, desde un punto de vista sintáctico, nuclean a los componentes restantes que constituyen sus argumentos.

Necesitamos conocer diversas características de los predicados de opinión. En lo referente a los verbos, observamos que estos pertenecen a varias clases semánticas, con diferentes matices de significado pero emparentadas. Los nombres que expresan opiniones son, en general, derivados de los verbos antes mencionados (*opinar/opinión, rechazar/rechazo*), por lo que comparten sus propiedades semánticas y probablemente muchas de sus propiedades sintácticas. Los predicados de opinión preposicionales, *según, de acuerdo con*, etc., constituyen una lista muy acotada y su comportamiento es poco variable.

En cuanto a los elementos que completan la opinión -la fuente, el asunto y el mensaje- en el caso de los verbos y de los nombres, por ser estos deverbales, se trata, en general, de información expresada a través de sus argumentos.

Para el tratamiento de este tipo de información, nos inspiramos en el lexicón de las gramáticas generativas. La concepción de lexicón dentro de la corriente generativista ha evolucionado desde sus orígenes. En un principio se incluía en el lexicón solamente información sobre subcategorización -cantidad y forma sintáctica de los argumentos de los verbos- y sobre restricciones de selección -rasgos de tipo animado, humano, abstracto, etc., de cada argumento- (Chomsky, 1957, 1964). En etapas posteriores (Chomsky, 1981) se incorporó al lexicón información semántica -roles temáticos- concepto que tiene como antecedentes principales a Gruber (1965) y Fillmore (1968).

El estudio que presentamos a continuación se centra principalmente en los verbos. Describimos algunas propuestas que, en mayor o menor medida, incluyen la información que estamos buscando: cuáles son las clases de verbos que permiten expresar opiniones, cuáles son los argumentos de esos verbos y sus propiedades sintáctico-semánticas -forma sintáctica, función sintáctica y roles semánticos- y qué estructuras sintácticas son posibles para cada verbo.

Finalmente, comentamos las características de los predicados de opinión nominales y preposicionales.

3.1 Los predicados de opinión verbales

Desde un punto de vista puramente semántico, nos interesa identificar clases de verbos que expresan opiniones. Describimos brevemente algunas propuestas de clasificación semántica de verbos, originalmente para el inglés pero con adaptaciones posteriores al español. Veremos que en algunos casos, si bien se llega a la determinación de clases semánticas, se parte de criterios de clasificación más bien sintácticos.

En primer lugar analizamos los verbos que permiten la reproducción de citas, en base al estudio para el español de Maldonado (1999).

Luego presentamos diferentes propuestas de clasificación verbal, de modo de poder analizar en detalle las clases de verbos que permiten la expresión de opiniones. Estas propuestas son: las clases de verbos del inglés de Levin (1993), basada en alternancias de diátesis; la clasificación de procesos de Halliday (1985) y la propuesta de clases basadas en *frame features* de Fillmore (1968). Veremos que todos estos trabajos, además de proponer clases de verbos, estudian, en mayor o menor medida, los argumentos de los verbos de cada clase.

3.1.1 Verbos introductores de cita

El discurso reproducido es el mecanismo usual para la reproducción de citas en los textos. Este tipo de expresión lingüística implica la existencia de dos situaciones de enunciación: E_1 y E_2 . En la situación de enunciación E_1 un hablante reproduce la situación de enunciación E_2 , el objeto de E_1 es E_2 . Este proceso puede ser recursivo: el texto citado puede a su vez contener otras citas.

Maldonado (1999) impone como condición para el discurso reproducido el hecho de que en la situación E_1 aparezca siempre el objeto del discurso, es decir, las palabras producidas en la situación de enunciación E_2 , como se ve en el ejemplo (3.1), en cursiva. Una simple mención de un acto lingüístico, como en (3.2), no es 'discurso reproducido' sino 'discurso referido'.

(3.1) Mahuad dijo que *su país no rechazaba la idea de crear un parque binacional en la zona en disputa*.

(3.2) El abogado de Fernando Botero habló sobre el tema con Semana.

Es importante destacar que si bien toda cita representa las palabras de algún enunciador, dichas palabras pueden ser reales o ficticias, exactas o aproximadas, pronunciadas o pensadas.

Procedimientos de cita

Existen variados mecanismos para reproducir un discurso, los más típicos son los llamados 'discurso directo' (DD), ejemplo (3.3), y 'discurso indirecto' (DI), ejemplo (3.4)¹¹.

(3.3) Consultado sobre la lentitud de los procesos judiciales uruguayos Carranza respondió: "Hay una situación de un muy alto número de presos sin condena, hay que agilizar los procesos".

(3.4) El vicepresidente del Instituto Nacional de Carnes de Uruguay, Fernando Pérez Abella, dijo que las conclusiones son muy alentadoras.

Maldonado define al DD como "la reproducción literal de palabras propias o ajenas" y al DI como "la reproducción de esas palabras desde el sistema de referencias deícticas del hablante que reproduce".

En los dos casos existe una expresión introductora, constituida por un 'verbo de *decir*' flexionado (en los ejemplos, los verbos *respondió* y *dijo*), y una cita. En el DD la cita es directa, lo cual se marca por comillas o guiones, y es la reproducción literal de un enunciado. El 'verbo

11 La mayoría de los ejemplos que presentamos en esta sección (3.1.1) pertenecen a nuestro corpus de desarrollo. Para algunas de las características presentadas no se encontraron ejemplos en dicho corpus, por tratarse de estilos poco frecuentes en los textos periodísticos. En esos casos, se toman los ejemplos de Concepción Maldonado y se agrega la indicación [CM] al final del ejemplo.

de *decir*' y la cita están separadas por los dos puntos. En el DI la cita es indirecta, su marca es la conjunción *que* y está subordinada al verbo introductor.

Maldonado describe algunos otros mecanismos de cita: el 'discurso pseudo-directo' (3.12), la 'oratio quasi oblicua' (3.13), el 'discurso indirecto mimético' (3.14), el 'estilo indirecto libre' (3.15) y el 'discurso directo libre' (3.16).

- (3.12) Vázquez dijo que "ustedes han visto la campaña que se ha desatado a través de este fenómeno, que ha sido dramatizado, que ha sido estigmatizado, que parece que este gobierno lo que busca es simplemente perjudicar a los jubilados y a los pensionistas, generalizando una situación que es de unos pocos, que tienen muy buenos salarios, que tienen muy buenas jubilaciones, muy buenas pensiones, y que es justo, desde nuestro punto de vista, que se les cobre un impuesto superior que al que gana menos".
- (3.13) Kirchner le habría pedido en el 2003 que Uruguay no tomara decisiones sobre las pasteras sin cumplir antes con el Tratado del río.
- (3.14) ? Mi sobrina dijo que quería cenar *croquetas de bacalado*. [CM]
- (3.15) Nerviosa se dijo: "¿Dónde habré puesto la carta?". *La había dejado allí encima, estaba segura. Quizá él la había descubierto y había leído lo que allí estaba escrito.* ... [CM]
- (3.16) Estuvimos discutiendo hasta las tantas. Yo: "Me apetece el blanco". Y ella: "Pues a mí, el negro". ... [CM]

Algunos de estos estilos nos interesan ya que pueden aparecer en textos periodísticos, que es el tipo de texto con el que trabajamos. El estilo 'pseudo-directo' y la 'oratio quasi oblicua' son muy frecuentes en la prensa escrita. De hecho, encontramos fácilmente los ejemplos (3.12) y (3.13) en nuestro corpus de prensa.

El 'discurso indirecto mimético', si bien sigue la sintaxis del DI, reproduce elementos agramaticales, por lo que es necesario indicar su literalidad mediante algún recurso formal, como la cursiva del ejemplo (3.14). No parece común la utilización de este estilo en textos de prensa.

El 'estilo indirecto libre' transmite un monólogo interior, mediante una mezcla de DD con DI, de modo que se confunden las palabras del narrador con las del personaje, como sucede con el texto en cursiva del ejemplo (3.15). En el 'discurso directo libre' se transmiten diálogos en donde se suele omitir el verbo introductor de la cita, como en (3.16), aunque generalmente se mantienen otras marcas formales del DD como las comillas o los guiones. Estos dos estilos son característicos del lenguaje literario.

Grupos de verbos que introducen el discurso reproducido

Maldonado identifica dos grupos de verbos que pueden introducir la cita, directa o indirecta, en un discurso reproducido: los verbos de comunicación verbal y los verbos epistémicos. A continuación comentaremos las características de cada clase.

a) Verbos de comunicación verbal

Dentro de esta clase, que suele ser denominada verbos *dicendi*, el término más general es *decir*. Los otros términos que suelen incluirse en esta clase aportan algún tipo de información

adicional sobre el acto verbal que reproducen. Es posible clasificar a los verbos de comunicación verbal en las siguientes subclases

- verbos de opinión: *opinar, considerar, reputar, juzgar, etc.*
- verbos de valoración positiva: *alabar, aplaudir, aprobar, celebrar, felicitar, elogiar, etc.*
- verbos de valoración negativa: *criticar, reprochar, etc.*
- verbos declarativos: *decir, comunicar, mencionar, notificar, manifestar, responder, etc.*
- verbos de manera de decir: *gemir, gritar, susurrar, chillar, balbucear, murmurar, etc.*
- verbos marcadores de la modalidad de enunciación: *exclamar, preguntar, etc.*
- verbos de orden o mandato: *mandar, ordenar, encargar, prohibir, etc.*
- verbos de petición o ruego: *rogar, pedir, suplicar, exigir, solicitar, reclamar, etc.*
- verbos declarativos con valor prospectivo: *anunciar, pronosticar, predecir, prometer, etc.*

Algunos verbos de comunicación, por su contenido léxico, solamente pueden introducir una cita directa (*recitar, declamar, cantar, pronunciar, transcribir, tartamudear, parodiar*).

Observamos que los verbos de valoración, si bien son considerados por la autora como verbos *dicendi*, introductores tanto de DD como DI, no parecen funcionar como introductores de cita, sea esta directa o indirecta. La autora no presenta en este trabajo ejemplos de discurso reproducido con estos verbos. En nuestro corpus de análisis tampoco encontramos ejemplos de verbos de valoración introduciendo una cita, aunque sí los encontramos con algún complemento indicando el objeto que es valorado, como el grupo nominal *el embargo estadounidense* en (3.17), o introduciendo una subordinada que no expresa el contenido de un acto verbal sino una situación que es valorada, como *que Ramos sea consejero departamental* en (3.18). Casos como el de (3.18) no pueden considerarse discurso indirecto ya que la subordinada no contiene las palabras emitidas. Es importante para nuestros objetivos determinar correctamente en qué casos una subordinada es el mensaje (cita indirecta) y en qué casos constituye otro elemento, en particular, en (3.18) la subordinada es el asunto de la opinión.

(3.17) Fidel Castro criticó el embargo estadounidense.

(3.18) Consejo de la UMSA aprobó que Ramos sea consejero departamental.

No se incluye en esta lista el verbo *hablar*, ya que es considerado introductor de discurso referido, como se dijo antes. Sin embargo, encontramos en nuestro corpus de desarrollo casos en donde este verbo introduce una cita, como en (3.19).

(3.19) En una carta escrita por Dalí en Neuilly en abril de 1951, el artista habla sobre su divina inspiración: "Yo quería que el próximo Cristo que pintase".

b) Verbos de percepción, verbos epistémicos y verbos de sentimiento

- Verbos de percepción: *oír, entender, escuchar, etc.*
- Verbos de pensamiento y sentimiento: *creer, pensar, intuir, etc.*

Una diferencia importante entre los verbos del grupo a) y los del grupo b) es que los primeros son en general verbos de tres argumentos, el emisor, el receptor y lo que se comunica, en cambio, los verbos del segundo grupo tienen dos argumentos, el que piensa y lo que se piensa.

En el contexto de este trabajo, como no incluimos al receptor del mensaje como un elemento de la opinión, esta distinción carece de relevancia.

En algunos contextos se observan otros verbos, además de los verbos de estos grupos, que introducen una cita, como en los ejemplos (3.20) y (3.21). La autora excluye explícitamente estos casos de su estudio. En textos de prensa, es muy poco probable que aparezcan casos como estos.

(3.20) "¡Quien fuera joven!", soñó. [CM]

(3.21) Pepa hizo: "¡Fenomenal!". [CM]

3.1.2 Clasificación verbal basada en alternancias de diátesis

Uno de los trabajos más importantes sobre clasificación verbal es el de Levin (1993). La clasificación que allí se propone está motivada por criterios sintáctico-semánticos, se parte de la hipótesis de que los verbos que comparten alternancias de diátesis pertenecen a una misma clase semántica. Se entiende por alternancias de diátesis:

*... alternations in the expressions of arguments, sometimes accompanied by changes of meaning*¹²... (Levin, 1993, p 2)

Explicaremos la relación entre alternancias de diátesis y variación de significado a través del análisis de los verbos *break*, *cut*, *hit* y *touch*, presentado en (Levin, 1993, pp 5 a 11). Los cuatro verbos son transitivos, con dos argumentos, un sujeto y un objeto:

(3.22) Margaret cut the bread.

(3.23) Janet broke the vase.

(3.25) Terry touched the cat.

(3.26) Carla hit the door.

La aplicación de dos alternancias, Construcción Media y Construcción Conativa, a los cuatro ejemplos se muestra en la tabla 3.1.

Construcción media	Construcción conativa
The bread cuts easily.	Margaret cut at the bread.
Crystal vases break easily	* Janet broke at the vase.
* Cats touch easily.	* Terry touched at the cat.
* Door frames hit easily.	Carla hit at the door.

Tabla 3.1: Construcciones media y conativa aplicadas a *cut*, *break*, *touch* y *hit*.

La aceptación de las diferentes alternancias, por parte de cada verbo, se resume en la tabla 3.2.

¹² Traducción: ... alternancias en las expresiones de los argumentos, a veces acompañadas por cambios de significado ...

	<i>hit</i>	<i>cut</i>	<i>break</i>	<i>touch</i>
Construcción media	no	sí	sí	no
Construcción conativa	sí	sí	no	no

Tabla 3.2: Aceptación de las construcciones media y conativa.

De la observación de las alternancias que cada verbo acepta se deducen algunos componentes de significado que algunos de ellos comparten:

- Los verbos *break* y *cut* implican cambio de estado, propiedad relacionada con la alternancia de construcción media.
- Los verbos *hit* y *cut* implican movimiento, esto se relaciona con la construcción conativa.
- El verbo *touch* no implica necesariamente movimiento ni cambio de estado, lo cual tiene relación con la no aceptación de ninguna de las dos alternancias analizadas.

Levin hace un catálogo de las alternancias de diátesis que existen para el inglés. Define cerca de 90 alternancias posibles, agrupadas en grupos y subgrupos. En base a las relaciones entre alternancias de diátesis y componentes de significado, Levin establece 49 clases semánticas de verbos, muchas de las cuales incluyen subclases.

Clases verbales del inglés que contienen verbos de opinión

Para nuestro trabajo, rastreamos entre las clases de Levin los verbos que permiten la expresión de opiniones. Encontramos verbos de opinión dentro de las clases:

- verbos con complementos predicativos
- verbos psicológicos
- verbos de juicio
- verbos de interacción social
- verbos de comunicación

De las clases mencionadas, solo en el caso de 'juicio' y 'comunicación' todos los elementos tienen sentido como verbos de opinión. Las clases restantes incluyen algunas subclases, o simplemente algunos pocos elementos, que creemos pueden expresar opiniones.

En la tabla 3.3, mostramos ejemplos de cada clase, solo para las subclases que incluyen verbos que expresan opiniones.

Las clases de verbos incluidas en la tabla permiten crear una lista a partir de la cual buscar en los textos las opiniones, pero la información que brinda Levin para cada clase no ayuda directamente a la identificación de los argumentos que necesitamos reconocer. No se establece para cada clase el mapeo entre argumentos y roles semánticos, de modo de poder identificar los elementos que buscamos para la opinión (la fuente, el asunto y el mensaje). Por otro lado, sí se estudian las diferentes estructuras sintácticas permitidas para cada clase, lo cual es importante para el trabajo de identificación automática de las opiniones, aunque el estudio no es del todo exhaustivo. Por ejemplo, en el caso de los verbos de *decir*, Levin aclara que estos admiten

diferentes tipos de complementos oracionales, pero este tipo de complemento no se analiza dentro del alcance del libro.

Clase	Subclases que contienen verbos de opinión	Verbos
Verbs with Predicative Complements (8 subclases en total)	Appoint Verbs	<i>acknowledge, consider, imagine, report, want</i>
	Characterize Verbs	<i>accept, appreciate, confirm, reject, remember, repudiate</i>
	Dub Verbs	<i>pronounce</i>
	Declare Verbs	<i>believe, confess, declare, judge, presume, think</i>
	Conjecture Verbs	<i>admit, assert</i>
Psych-Verbs (4 subclases en total)	Admire Verbs	Positive: <i>admire, adore, like</i> Negative: <i>detest, dislike, fear</i>
Judgment Verbs (no tiene subclases)		Positive: <i>acclaim, congratulate</i> Negative: <i>criticize, condemn</i>
Verbs of Social Interaction (3 subclases en total)	Correspond Verbs	<i>agree, argue, disagree, dissent</i>
Verbs of Communication (9 subclases en total)	Verbs of Transfer of a Message	<i>cite, explain, narrate, tell, write</i>
	Tell	<i>tell</i> (único miembro)
	Verbs of Manner of Speaking	<i>murmur, scream</i>
	Verbs of Instrument of Communication	<i>e-mail, fax, signal</i>
	Talk Verbs	<i>speak, talk</i>
	Chitchat Verbs	<i>argue, chat, converse</i>
	Say Verbs	<i>announce, confess, declare, say</i>
	Complain Verbs	<i>complain, object</i>
	Advise Verbs	<i>admonish, advert</i>

Tabla 3.3: Clases de Levin que contienen verbos que pueden transmitir opiniones

Clasificación basada en alternancias de diátesis de los verbos del español

Si bien resulta útil estudiar las clases que Levin propone para el inglés, creemos conveniente analizar propuestas similares para el español. Vázquez, Fernández y Martí (2000) estudian las alternancias de diátesis de los verbos del español. Estas autoras entienden las diátesis como "la expresión sintagmática de diferentes oposiciones semánticas motivadas por distintas estrategias comunicativas".

A diferencia de Levin, en este trabajo se establecen dos grandes grupos de diátesis para el español: 'oposición en la conceptualización de un evento' y 'oposición aspectual'. La primera tiene que ver con diferentes realizaciones sintácticas de los participantes de un evento. La

segunda implica la alternancia entre construcciones con diferente interpretación aspectual, eventiva versus estativa.

Entre las alternancias de diátesis que denotan una oposición en la forma de expresar un evento, se distinguen dos grupos. Por un lado, las 'alternancias de cambio de foco', que implican cambios de posición de los constituyentes de la oración. Dentro de este grupo existen varios subgrupos, uno de ellos es la alternancia entre causatividad y anticausatividad (*Los empleados del aeropuerto han extraviado las maletas / Se han extraviado las maletas*). Por otro lado, las 'alternancias de infraespecificación', en la cual se opone una construcción más específica frente a una más general, por medio de la no especificación de algunos argumentos del verbo (*Pedro comió manzanas / Pedro comió*).

Las posibles alternancias de diátesis que implican una oposición aspectual son tres: 'resultativa' (*El productor ha cortado una película / La película está cortada*), 'media' (*El niño ha esparcido la pintura / Esta pintura se esparce con facilidad*) y 'estativa personal de tiempo no marcado' (*María bailó el tango / María baila el tango*).

Teniendo en cuenta las alternancias aceptadas por cada verbo, las autoras proponen dos grandes clases de verbos: verbos de cambio y verbos de trayectoria. Para cada clase se establecen: los componentes de significado, la estructura eventiva y las alternancias de diátesis. Cada clase se caracteriza por el hecho de que todos sus integrantes admiten un conjunto de alternancias generales. La admisión o no de otras alternancias permite definir algunas subclases, dentro cada una de las dos clases iniciales.

En las dos clases, cambio y trayectoria, encontramos verbos de opinión. Veremos qué características tienen los verbos de opinión que se encuentran en cada clase.

a) Verbos de cambio mental

Este grupo de verbos se encuentra dentro de la clase de verbos de cambio. Son un subgrupo de verbos psicológicos que tienen la particularidad de implicar un cambio en la entidad afectada por el predicado, algunos de estos verbos son *abatir, maravillar, sorprender*. Como el resto de los verbos de esta clase, estos verbos tienen un iniciador causativo. Ese iniciador se expresa sintácticamente a través del sujeto del verbo (3.27)¹³. El objeto del verbo es la entidad afectada o experimentador. En cambio, para otros verbos psicológicos como *amar*, no pertenecientes a la clase de cambio mental, el sujeto se interpreta como un experimentador (3.28) y el objeto es el tema o estímulo.

(3.27) Juan preocupa a sus padres.

(3.28) Romeo amó a Julieta.

Las alternancias principales admitidas por toda la clase de verbos de cambio, y por lo tanto por los verbos de cambio mental, son 'alternancia anticausativa prototípica' (*La calificación obtenida desanimó a Miguel / Miguel se desanimó*) y la 'alternancia resultativa' (*El comentario de Silvia hirió a Ana / Ana está herida*).

Además de esas dos alternancias, los verbos de cambio mental en particular admiten la

13 Los ejemplos de esta sección pertenecen a (Vázquez et al., 2000).

'alternancia media anticausativa' (*María se emociona con facilidad*) y la 'alternancia holística' (*Pedro convenció a todo el mundo con su sencillez / La sencillez de Pedro convenció a todo el mundo*).

Estos verbos no admiten, por otra parte, las alternancias 'pasiva', 'de inversión' (*El sol irradia calor / El calor irradia del sol*) y 'de infraespecificación' (elisión de algunos participantes).

b) Verbos de comunicación

Dentro de los verbos de trayectoria están incluidos los verbos de comunicación, por considerarse que existe una entidad, el mensaje, que es transferida desde un origen, el emisor, hacia un destino, el receptor.

Para todos los verbos de trayectoria los componentes de significado identificados son la entidad, el iniciador y la trayectoria. En el caso de los verbos de comunicación, la entidad puede ser un grupo nominal como *de la historia de Grecia* en (3.29), que coincide con lo que hemos llamado asunto, o la cita directa o indirecta como *que no lo volvería a hacer* en (3.30), que constituye lo que hemos llamado mensaje. Por otro lado, el iniciador es el emisor del mensaje y el destino, subcomponente de la trayectoria, es el receptor del mensaje. También puede expresarse otro subcomponente de la trayectoria, la ruta, como *por el altavoz* en (3.31).

(3.29) El profesor habla de la historia de Grecia a sus alumnos.

(3.30) El niño dijo que no lo volvería a hacer.

(3.31) Les comunicaron la noticia por el altavoz.

La alternancia de diátesis aceptada por todos los verbos de la clase trayectoria es una variante de la alternancia de infraespecificación que consiste en la elisión de algún complemento que exprese un subcomponente de la trayectoria. Esta alternancia es clara para los verbos de comunicación, ya que el receptor o el medio a través del cual se transfiere el mensaje no son obligatorios.

Además de la alternancia de infraespecificación, admitida por toda la clase trayectoria, los verbos de comunicación aceptan la alternancia pasiva (*la noticia les fue comunicada por el altavoz*). Estos verbos también aceptan la alternancia de infraespecificación del componente entidad (*María habló del asunto con sus padres / María habló con sus padres*).

Algunas alternancias son rechazadas por todos los verbos de trayectoria: anticausativa prototípica, anticausativa de proceso y media anticausativa. Los verbos de comunicación, en particular, no admiten la alternancia de inversión (*El mozo cargó los paquetes en el coche / El mozo cargó el coche de paquetes*) ni la resultativa (*La casa ya está vendida*).

En su trabajo, las autoras incluyen un catálogo de verbos del español (también del catalán y del inglés), indicando a cuál de las dos grandes clases iniciales estos pertenecen pero sin indicar cuál es la subclase correspondiente. Para el objetivo de nuestro trabajo el catálogo no aporta demasiada información, ya que entre los 1000 verbos del catálogo muy pocos son de opinión. Además, solo es posible saber si son de cambio o de trayectoria.

Por otra parte, muchos de los verbos que consideramos verbos de opinión, como *rechazar*, *apoyar*, *criticar*, *gustar*, *agradar*, no son mencionados en el trabajo. No pertenecen a ninguna de las dos clases definidas, o al menos no se incluyen ejemplos similares en la lista final de verbos que permitan deducir la clase que les corresponde. Además, no se hace mención a otras posibles clases verbales que puedan contener los verbos no considerados entre los de cambio o trayectoria.

De todos modos, toda la información relativa a las diátesis aceptadas o rechazadas por los verbos de cambio mental y de comunicación constituye un dato importante a la hora de implementar un procedimiento para la identificación automática de los componentes de la opinión, ya que la posición que ocupan los diferentes argumentos del verbo, algunos de los cuales coinciden con los componentes que estamos buscando, se deduce de las diferentes diátesis posibles para estos.

3.1.3 Clasificación verbal basada en *frames*

Fillmore (1968) propone incluir en las entradas léxicas lo que él denomina *frame features*, que consisten en conjuntos de *case frames* en los cuales dicha pieza puede ser insertada. Un *case frame* establece las diferentes combinaciones de 'casos', en el sentido de argumentos con sus roles semánticos asociados, que son aceptados por una pieza léxica. Como ejemplo, tomaremos el estudio que Fillmore hace del verbo *open* (1968, p. 51).

El verbo *open* puede aceptar diferentes combinaciones de argumentos:

(3.32) The door opened.

(3.33) John opened the door.

(3.34) The wind opened the door.

(3.35) John opened the door with a chisel.

Cada estructura posible se especifica mediante un *case frame*, como se muestra a continuación (A indica caso agentivo o rol agente, O indica caso objetivo o rol objeto, I indica caso instrumental o rol instrumento):

[____ O] para el ejemplo 3.32

[____ O + A] para el ejemplo 3.33

[____ O + I] para el ejemplo 3.34

[____ O + I + A] para el ejemplo 3.35

El conjunto de *case frames* posibles para esta pieza léxica se especifica en forma resumida mediante el siguiente *frame feature*:

+ [____ O (I) (A)]

El signo + significa que el *frame feature* representa los entornos posibles para la pieza (un signo - indicaría entornos no permitidos), la línea indica la posición en donde debe ser insertada la pieza léxica que responde al conjunto de *case frames* que describe este *frame feature*, los paréntesis indican elementos opcionales.

Cada *frame feature* determina una clase de piezas léxicas que comparten el conjunto de configuraciones allí especificadas.

En su propuesta, Fillmore no entra en detalles sobre las diferentes clases de verbos que los frames determinan, pero existe un proyecto posterior de clasificación verbal, basada en esta propuesta, que constituye la base de un recurso léxico muy importante, FrameNet.

Un recurso léxico basado en el concepto de frame

Basado en la semántica de *frames* de Fillmore (1968), FrameNet (Baker et al, 1998, Web FrameNet) es un recurso léxico de gran cobertura disponible en la web.

El léxico se clasifica según diferentes *frames* que están organizados en un grafo en donde existen relaciones de herencia y otras relaciones transversales. Para cada *frame* se especifican todos los elementos (argumentos etiquetados con sus roles temáticos) posibles, algunos son elementos centrales (*core elements*) y otros secundarios.

El recurso ofrece una interfaz en línea que permite consultar toda la información allí contenida. Además de buscar información relativa a cada *frame* particular, es posible hacer búsquedas por unidad léxica, obteniéndose toda la información correspondiente a la unidad y al *frame* que le corresponde.

El proyecto está siendo mejorado, se están incorporando para cada clase numerosos ejemplos de diferentes piezas léxicas, con los elementos del *frame* anotados, entre las cuales hay no solamente verbos sino otras categorías de palabras que también tienen estructura argumental, principalmente nombres y adjetivos. Los ejemplos anotados pertenecen a un corpus, por lo que representan el lenguaje real con sus complejidades.

Existe una versión de FrameNet para el español (Subirats, 2009, Web Spanish FrameNet) que se basa en la misma organización de clases que la versión original, pero es muy incompleta en cuanto a las piezas léxicas que se describen dentro de cada clase.

3.1.4 Clasificación verbal basada en tipos de procesos

Halliday (1985) propone una clasificación basada en tipos de procesos. Allí los procesos se organizan en seis tipos básicos: material, comportamiento, mental, verbal, relacional y existencial. Cada tipo contiene a su vez algunos sub-tipos, que mostramos en la tabla 3.4.

Dentro de esta clasificación, los verbos de opinión se encuentran entre los procesos verbales y mentales. Describimos estas clases analizando los argumentos típicos de cada una.

Tipo de proceso	Subtipos
material	general
	acción
	evento
comportamiento	
mental	
	percepción
	cognición
	deseo
	emoción
verbal	
relacional	general
	atribución
	identificación
existencial	

Tabla 3.4: Clases verbales propuestas por Halliday

a) Procesos verbales

Dentro de los procesos verbales, si bien Halliday no define subtipos, sí describe diferentes grupos de verbos:

1. 'targeting verbs' (verbos con meta): *insultar, culpar, regañar, criticar, elogiar*.
2. 'talking verbs' (verbos de conversación): *hablar, conversar*.
3. 'neutral quoting' (verbos de cita neutros): *decir, contar*.
4. 'indicating verbs' (verbos de explicación): *contar, anunciar, explicar, convencer*.
5. 'imperating verbs' (verbos imperativos): *ordenar, requerir, implorar*.

Los verbos del primer grupo siempre implican una postura subjetiva del emisor, cuya orientación está transmitida directamente por el verbo. En los casos restantes, se trata de verbos neutros. Para los verbos de conversación, es muy posible que ni siquiera se incluyan las palabras del emisor, como ya se mencionó en la sección 3.1.1, se trata de un discurso referido.

Los participantes que Halliday reconoce para los procesos verbales son 'Sayer' (emisor), 'Receiver' (receptor), 'Verbiage' (elemento cercano a lo que llamamos asunto, no es el mensaje) y 'Target' (meta u objetivo).

El participante 'Sayer' está presente para todos los grupos mencionados dentro de este tipo de proceso, es la fuente de la opinión.

El participante 'Receiver', si bien Halliday no hace ninguna aclaración sobre esto, no parece tener sentido para los verbos del primer grupo (*criticar, elogiar*). Nosotros no lo

hemos considerado como un elemento de la opinión en nuestra definición, justamente porque para muchos de nuestros predicados de opinión no tiene sentido hablar de un receptor.

El participante 'Verbiage' representa el contenido o el nombre de lo dicho: *hablar de la familia, explicó la decisión del presidente, hizo una pregunta, hablar árabe*. En el ejemplo con el verbo *hacer*, en nuestro enfoque es el nombre *pregunta* el que realmente expresa el contenido verbal, funcionando *hacer* como verbo soporte (ver sección 3.2), mientras que en este marco se considera *hacer* como un ejemplo de proceso verbal.

Es importante notar que Halliday no incluye ningún elemento similar a lo que nosotros llamamos mensaje entre los participantes de los procesos verbales. Desde un punto de vista sintáctico, la cláusula introductora del discurso reproducido y la cláusula que contiene el discurso reproducido son consideradas en este marco como cláusulas separadas, no se analiza la segunda como subordinada a la primera. La cláusula que contiene la cita, ya sea que se trate de discurso directo o de discurso indirecto, no constituye un participante del proceso verbal.

El último participante mencionado, el 'Target', aparece solamente cuando el verbo es del primer grupo ('targeting verbs') ya que representa el objetivo hacia el cual apunta el emisor. Observamos que este participante puede ser visto como la contraparte del participante 'Verbiage', dado que para los verbos que pueden tener el participante 'Verbiage' no es posible el participante 'Target' y, a la inversa, los verbos que aceptan un participante 'Target' no aceptan un participante 'Verbiage'. Como ya dijimos, tampoco es posible para estos verbos el participante 'Receiver'.

b) Procesos mentales

En cuanto a los procesos mentales, nos interesan como posibles verbos de opinión los correspondientes a los tipos cognición (*pensar, creer, suponer*) y emoción (*gustar, agradar, molestar*).

Los participantes propuestos para los procesos mentales son 'Senser' (experimentante) y 'Phenomenon' (fenómeno). En este caso, a diferencia de lo que sucede con los procesos verbales, el 'Phenomenon' puede ser una cláusula (*I regret very much [that I was away from home]*) y esta cláusula, finita o no finita, sí es considerada parte de la oración principal.

Es difícil definir cuáles verbos de emoción pueden expresar opiniones y cuáles expresan emociones de otro tipo, como *divertir, amar, temer*, etc. En principio, podemos pensar que los verbos que transmiten una posición positiva o negativa respecto a un fenómeno son de opinión, mientras que los que se mueven en otras escalas no lo son.

El tratamiento diferente de la cláusula subordinada, en los casos de verbos de comunicación verbal y verbos de creencia o sensación, no resulta apropiado para nuestro trabajo ya que, por un lado, nos interesa incluir dentro de la opinión las palabras expresadas por la fuente y, por otro lado, pretendemos hacer un tratamiento homogéneo de todas las clases verbales que integremos a la clase más general de los verbos de opinión.

3.1.5 Clases verbales y recurso léxico para el español

Inspirada en la clasificación de procesos de Halliday (1985), existe para el español una propuesta de clasificación semántica de los verbos, acompañada por la descripción de los esquemas sintáctico-semánticos de cada clase: el proyecto ADESSE (Albertuz, 2007, García-Miguel, 2005, Web ADESSE). Las clases semánticas de este proyecto coinciden en gran medida con las de Halliday.

En (Albertuz, 2007) se justifica la decisión de utilizar la clasificación de Halliday, ante la alternativa de tomar como base la de Levin, por el tipo de clases verbales que cada una propone. Las de Halliday están motivadas por criterios puramente semánticos mientras que la de Levin, al ser motivada por propiedades sintácticas, lleva a un número de clases demasiado grande en donde, en algunos casos, verbos semánticamente emparentados quedan en clases diferentes por aceptar alternancias diatéticas diferentes.

El proyecto se completa dando para cada clase el esquema argumental prototípico de la clase. Dicho esquema se especializa para las sub-clases, renombrándose los argumentos heredados, si se considera conveniente, y agregándose otros, si corresponde.

Al igual que en la descripción de los procesos de Halliday, los nombres que se dan para los participantes son específicos para cada clase (emisor, receptor, evaluador, evaluado, valoración, etc.), pero a diferencia de lo que allí se propone, para cada sub-clase de procesos, se especializan aún más los participantes: algunos nombres de participantes son modificados, algunos participantes son eliminados y otros son agregados. Por ejemplo, los verbos de comunicación tienen cinco participantes: emisor, mensaje, receptor, asunto, código; mientras que su sub-clase mandato tiene cuatro: emisor, petición (nuevo nombre para un participante similar al mensaje), receptor, causa (nuevo participante).

ADESSE ofrece una base de datos realmente extensa en la cual es posible encontrar numerosos verbos de cada clase, ordenados por frecuencia y acompañados por ejemplos anotados con los argumentos definidos para la clase junto con su función y forma sintáctica.

La clasificación completa de ADESSE¹⁴ se muestra en la tabla 3.5.

Como se observa en la tabla, entre las clases de procesos del primer nivel de clasificación, si comparamos las clases de ADESSE con las de Halliday, vemos que se agrega la clase modulación y se elimina la clase comportamiento (que se incluye como sub-clase de la clase material).

A un segundo nivel, se observan varias diferencias entre las dos propuestas. Además, ADESSE incluye un tercer nivel de clasificación lo cual no hace Halliday.

A continuación analizamos las clases que contienen verbos de opinión y sus sub-clases. Al igual que en el caso de la clasificación de Halliday, tomamos los procesos verbales y mentales, pero analizamos, además, los verbos de aceptación, que es una sub-clase de modulación. Describimos los participantes, en este marco llamados argumentos nucleares típicos, definidos para cada clase. En base a la observación de los numerosos ejemplos que ofrece la web del proyecto, podemos identificar los argumentos que corresponden a los elementos de la opinión

¹⁴ Mostramos la última versión de la propuesta publicada en la página web del proyecto a la fecha de redacción de este documento.

que queremos modelar. Los ejemplos muestran diferentes construcciones en las cuales aparecen los argumentos definidos para las clases, aunque no siempre están todos, como es natural.

Tipo de proceso	Subtipo
material	espacio
	cambio
	otros hechos
	comportamiento
modulación	causación
	disposición
	aceptación
	verbos de apoyo
mental	percepción
	cognición
	elección
	sensación
verbal	comunicación
	valoración
relacional	general
	atribución
	posesión
existencial	existencia
	fase-tiempo
	vida

Tabla 3.5: Clases verbales de ADESSE

a) Procesos verbales de ADESSE

Los procesos verbales se subdividen como sigue:

verbal — comunicación
— valoración

Para los verbos de comunicación se definen cinco argumentos, que ilustramos a través del ejemplo (3.36): emisor (*el socialista*), mensaje (*que Portomeñe hinchó un globo*), receptor (no aparece en 3.36), asunto (*sobre el Jacobeo*), código (no aparece en 3.36, se trata del código o instrumento utilizado para la comunicación, por ejemplo, el idioma).

(3.36) Sobre el Jacobeo, el socialista **dijo** que Portomeñe hinchó un globo ...

Para los verbos de valoración (3.37) se definen los argumentos: evaluador (*la Caja de Ronda*), evaluado (entidad que es evaluada o sobre la que se emite un juicio de valor, en

el ejemplo 3.37: *al PSOE*), asunto (factor desencadenante del proceso valorativo o finalidad del mismo, en el ejemplo 3.37: *311 millones de pesetas*), valoración (juicio expresado, no aparece en 3.37).

(3.37) La Caja de Ronda **había perdonado** al PSOE 311 millones de pesetas.

En base a la definición de los diferentes argumentos, y a la observación de diversos ejemplos, concluimos que el argumento que coincide con lo que nosotros llamamos asunto es el aquí llamado evaluado, y no el llamado asunto.

b) Procesos mentales de ADESSE

Los procesos mentales se subdividen como sigue:

- mental — sensación
 - volición
 - percepción
 - cognición
 - conocimiento
 - creencia
 - elección

De todas las subclases de verbos mentales, consideramos predicados de opinión: sensación (3.38) (incluyendo su subclase volición) y creencia (3.39), que es sub-clase de cognición.

(3.38) En el estadio de Santa Isabel la grúa se llevó varios coches que no **molestaban** a nadie.

(3.39) Yo **creo** que es demasiado tarde para su proyecto.

Los argumentos definidos para la clase sensación son experimentador (en 3.38: *nadie*) y estímulo (en 3.38: *varios coches*). Para su sub-clase volición se mantienen los mismos argumentos. Entre los ejemplos de ADESSE, para la mayoría de los verbos de esta clase el experimentador es un complemento del verbo y el estímulo es el sujeto.

Para la clase creencia se definen: conocedor (en 3.39: *yo*), contenido (en 3.39: *que es demasiado tarde para su proyecto*), asunto (no aparece en 3.39), iniciador (entidad que hace posible que el conocedor experimente un proceso cognitivo, no aparece en 3.39), fuente (entidad a través de la cual se obtiene la creencia, no aparece en 3.39), contenido2 (la creencia como predicación secundaria, no aparece en 3.39).

Es importante destacar que el argumento de la clase creencia que en ADESSE llaman fuente (como el grupo *tu abuelo* en la oración *no debes creer a tu abuelo*), no coincide que con lo que en nuestro trabajo llamamos fuente. Esta última coincide con el argumento conocedor.

c) Procesos de aceptación de ADESSE

La clase aceptación (3.40), que es sub-clase de modulación, acepta los argumentos aceptador (en el ejemplo 3.40 está omitido, corresponde a la tercera persona de singular), aceptado (en el ejemplo 3.40: *la frialdad del lenguaje*), origen (entidad que produce lo aceptado o lo pone a disposición del aceptador, en 3.40 no aparece), rol (el estatus por el cual o para el cual lo aceptado es aceptado, en el ejemplo 3.40: *como natural y nada ofensiva*).

(3.40) Acabó **aceptando** como natural y nada ofensiva la frialdad del lenguaje.

La principal riqueza de ADESSE se encuentra en la gran cantidad de información asociada a cada clase y sub-clase: definición de la clase, especificación de sus argumentos, verbos de la clase, verbos más frecuentes, esquemas sintácticos posibles (indicando cuáles son los más frecuentes), función y forma sintáctica de cada argumento en cada esquema posible, ejemplos anotados asociados a cada esquema.

En su mayoría, se trata de información obtenida a partir del análisis de un importante corpus¹⁵, lo cual constituye una ventaja para nuestro trabajo ya que nuestro objetivo es encontrar este tipo de estructuras en textos reales. El análisis de ejemplos sencillos que presentan algunos de los recursos estudiados, creados artificialmente, resulta muy útil para comprender los conceptos pero no se adapta a la realidad de los textos que tenemos que procesar.

Una desventaja de ADESSE es que aborda solamente el estudio de los verbos. Para nuestro trabajo son igual de importantes ciertos nombres por ser también predicados de opinión. Asimismo, muchos de los ejemplos provienen de textos literarios, cuyo estilo es bien diferente al de los textos de prensa. De todos modos, también es posible encontrar numerosos ejemplos tomados de textos periodísticos.

En resumen, el proyecto ADESSE brinda para este trabajo particular la mayor cantidad de información, en forma ordenada y bastante completa. La clasificación verbal que allí se propone se ajusta a nuestras necesidades ya que los verbos que identificamos como de opinión pertenecen a pocas clases, las cuales contienen en su totalidad verbos de opinión. Una excepción es la clase 'Sensación' que abarca una gama de sensaciones muy diversa, entre las cuales consideramos solamente las que transmiten posiciones positivas o negativas hacia algún tema.

En la tabla 3.6 vemos de qué modo se mapean los argumentos de las clases seleccionadas de ADESSE a los elementos que componen nuestra opinión.

Clase semántica del predicado	Fuente	Asunto	Mensaje
verbal:comunicación	emisor	asunto	mensaje
verbal:valoración	evaluador	evaluado	valoración
mental:sensación	experimentador	estímulo	-
mental:cognición:creencia	conocedor	asunto	contenido
modulación:aceptación	aceptador	aceptado	-

Tabla 3.6: Mapeo entre los argumentos de ADESSE y los elementos de la opinión

¹⁵ Corpus Arthus (<http://www.bds.usc.es/corpus.html>).

3.2 Los predicados de opinión nominales

Los predicados de opinión nominales son, en su gran mayoría, nominalizaciones derivadas de bases verbales (o nombres deverbales).

Al igual que para los verbos, nos interesa reconocer los argumentos de los nombres de opinión. A partir la estructura argumental de los verbos de opinión y las posibles realizaciones sintácticas de cada argumento, necesitamos saber también de qué forma los mismos argumentos se manifiestan cuando en vez de un verbo tenemos una nominalización.

Creemos que para este punto no es de utilidad estudiar los procesos de nominalización del inglés, ya que estos dependen en gran medida de las características particulares de cada pieza léxica. Analizamos las nominalizaciones, por lo tanto, en base a algunos artículos de la Gramática Descriptiva de la Lengua Española (Bosque y Demonte, 1999).

Picallo (1999) dice que los sintagmas nominales nucleados por nominalizaciones "son construcciones similares a oraciones" por la relación semántica que se da entre el núcleo y sus complementos o argumentos. Los complementos del nombre, que siempre son grupos preposicionales, se corresponden con el sujeto y los complementos del verbo del cual ese nombre deriva.

En muchos casos, uno de los argumentos de una nominalización puede interpretarse como agente, rol que en el caso de los nombres derivados de verbos de opinión se corresponde con la fuente. Picallo ilustra diferentes formas de expresar la fuente a través de un complemento o modificador de una nominalización, algunas son aplicables a los nombres de opinión: un complemento con *de* (3.41), un determinante posesivo (3.42), un adjetivo de relación (3.43) o un complemento con *por* (o *por parte de*).

(3.41) No, no creo en el contenido de la declaración de Vicente.

(3.42) Esta conclusión fue decisiva para su rechazo a las pretensiones de la planificación central.

(3.43) gracias al apoyo internacional, al apoyo de la prensa guatemalteca y ...

El complemento con *de* suele ser ambiguo, ya que puede representar tanto la fuente, como ya se mostró en (3.41), como el asunto (3.44).

(3.44) ... palabras que muchos han interpretado como un anuncio de la venida de Cristo.

Además de estos nombres deverbales, observamos que se pueden transmitir opiniones con otros nombres como *palabras* (3.45).

(3.45) La escenografía que se requiere, en palabras de Luis López, coordinador artístico de la ABAO, "es muy particular".

3.3 Los predicados de opinión preposicionales

Al igual que en la sección 3.2, en esta sección nos basamos en artículos de la Gramática Descriptiva de la Lengua Española (Bosque y Demonte, 1999).

La preposición *según* es el predicado de opinión preposicional más común. Pavón Lucero (1999) considera que se trata de una preposición imperfecta, por poseer algunas características que la diferencian de las demás preposiciones, aunque habitualmente es clasificada dentro de las preposiciones del español.

Una de las características particulares de esta preposición imperfecta es que rige caso nominativo (*según tú*) en vez de caso oblicuo (*para ti*) como las demás preposiciones. Como esta particularidad solo se manifiesta en el caso de los pronombres, en cualquier otro caso *según* se comporta como otras preposiciones tomando como complemento un sintagma nominal.

Según Pavón Lucero, uno de los valores de *según* es la introducción de un sintagma nominal que "representa a una fuente de información o de opinión", como en (3.46). En estos casos, el sintagma introducido por *según* es un elemento periférico a la oración, separado de esta por una pausa, pudiendo ir intercalado como un inciso (3.47).

(3.46) Según estos especialistas, la administración de corticoides debe limitarse a casos de asma moderadamente grave ...

(3.47) La respuesta, según Barbacid, surgió hace unos años ...

Además de la partícula *según*, tienen el mismo valor de introductores de fuente de opinión las locuciones prepositivas *de acuerdo con*, *de acuerdo a* y *a juicio de*. También la preposición *para* puede usarse con este valor (3.48). Es llamativo que entre los diversos valores que se describen en De Bruyne (1999) para esta preposición no se mencione el de introductor de fuente.

(3.48) Para Guillermo Castro, director de servicios de la IMM, se cuenta con el instrumental necesario para prevenir el inicio de la epidemia ...

3.4 Síntesis

Realizamos un estudio lingüístico de las expresiones que permiten transmitir opiniones, centrándonos principalmente en los predicados de opinión verbales.

Describimos diferentes trabajos que permiten analizar la información que necesitamos conocer para poder identificar opiniones en textos: cuáles son las clases de verbos que permiten expresar opiniones, cuáles son los argumentos de esos verbos y sus propiedades sintáctico-semánticas y qué estructuras sintácticas son posibles para cada verbo. También analizamos, aunque en forma menos detallada, los nombres que expresan opiniones y las preposiciones y locuciones preposicionales como *según*, *de acuerdo con*, etc.

En primer lugar, estudiamos los verbos que introducen el discurso reproducido en español, en base al trabajo de Maldonado (1999). Además, estudiamos diferentes propuestas de clasificación verbal: las clases verbales de Levin (1993), basadas en alternancias de diátesis, y su adaptación al español (Vázquez et al, 2000); la propuesta de clases de verbos basadas en *frame features* de Fillmore (1968) que da lugar a un importante recurso léxico (Web FrameNet) para el cual existe

una adaptación al español (Web Spanish FrameNet); y la clasificación de verbos en base a diferentes tipos de procesos de Halliday (1985), en la cual se basa el recurso léxico para el español ADESSE (Albertuz, 2007, Garcia-Miguel, 2005, Web ADESSE). Todos estos trabajos, además de proponer clases verbales, analizan los argumentos de los verbos de cada clase. Nosotros estudiamos, en particular, los argumentos que coinciden con los elementos definidos para la opinión (fuente, asunto y mensaje).

El estudio realizado constituye la base del modelo para los predicados de opinión que presentamos en el capítulo 5. Como punto de partida para definir los diferentes tipos de predicados verbales que integran nuestro modelo tomamos las clases de ADESSE que contienen verbos de opinión: comunicación, creencia, valoración, aceptación y sensación. La decisión de tomar como base la propuesta de ADESSE se justifica por diferentes motivos:

- El recurso fue creado originalmente para el español, lo cual garantiza que las particularidades que allí se describen para cada clase y para cada pieza léxica son aplicables a nuestros textos.
- Las clases semánticas, basadas en la clasificación de Halliday, están bien definidas y es posible identificar cuáles de ellas corresponden a verbos de opinión (las 5 clases ya mencionadas). Esto no sucede con las clases de Levin que son más numerosas y no se encuentran organizadas en forma jerárquica en su totalidad. Tampoco es clara la organización de los *frames* de FrameNet, entre los cuales se definen relaciones muy variadas formando un grafo de gran complejidad.
- El recurso brinda numerosos ejemplos tomados de textos reales en español, ordenados según la clase semántica del verbo y además según los argumentos expresados en cada caso y las estructuras sintácticas en las que estos aparecen. También se pueden obtener las piezas léxicas más frecuentes de cada clase.

Una desventaja que presenta ADESSE, en función de nuestros objetivos, es que muchos de los ejemplos pertenecen a textos literarios, siendo imposible filtrar solamente los ejemplos de textos periodísticos, que son el objeto de nuestro estudio.

TERCERA PARTE: Nuestra propuesta

4. Metodología de trabajo y análisis de corpus

Describimos a continuación la metodología seguida para el desarrollo de este trabajo. Luego, presentamos la definición de opinión que elaboramos y el análisis de corpus llevado adelante. A partir de este análisis, confeccionamos un conjunto de ejemplos que abarcan la gran mayoría de fenómenos lingüísticos a tener en cuenta para el desarrollo de los sistemas informáticos.

4.1 Metodología

El trabajo realizado se organiza en diferentes etapas sucesivas que describimos a continuación:

- Definición de la opinión y sus elementos: La definición de opinión elaborada surge de la observación de ejemplos en el corpus y de otras definiciones propuestas en diferentes trabajos del área.
- Modelo de predicados de opinión y sus argumentos: Especificamos las propiedades sintáctico-semánticas de los predicados de opinión y sus argumentos mediante Estructuras de Rasgos Tipificadas. Este formalismo permite definir rasgos de forma clara y, además, permite organizar los tipos de predicados en una jerarquía, de modo que las propiedades compartidas por diferentes tipos son heredadas por sus sub-tipos.
- Elaboración de un repertorio de predicados de opinión: Se extrajeron del corpus analizado los predicados de opinión más frecuentes y se asoció a cada uno de ellos el tipo que le corresponde en el modelo definido. También se definió para cada uno información léxica específica.
- Primer sistema informático: Inicialmente, para el reconocimiento automático se pensó en aplicar métodos simbólicos, principalmente por la disponibilidad de un sistema para la escritura y la aplicación de reglas contextuales. El sistema basado en reglas desarrollado abarca el reconocimiento de las opiniones y sus elementos, para lo cual explota el repertorio de predicados y la información relativa al tipo que les corresponde en el modelo. En algunos casos, el sistema determina la orientación semántica de las opiniones reconocidas.
- Segundo sistema informático: Dado que los métodos estadísticos han demostrado ser mejores que los simbólicos en diversas aplicaciones del PLN, consideramos interesante realizar un segundo sistema informático basado en aprendizaje automático. Realizamos, por lo tanto, algunos experimentos basados en el modelo *Conditional Random Fields* para el reconocimiento de fuentes de opiniones. La información que utilizamos para el aprendizaje es prácticamente la misma que utiliza el sistema de reglas.
- Tercer sistema informático: Habiendo obtenido valores bastante similares en el reconocimiento de fuentes con los dos sistemas anteriores, buscamos la manera de combinar los sistemas con la expectativa de lograr resultados superiores. El sistema combinado resultó muy exitoso, elevando de manera significativa los resultados anteriores.

4.2 Las opiniones en textos de prensa del español

En esta sección, en primer lugar, retomamos la definición de opinión que adelantamos en la introducción, de modo de presentar el contexto en el cual fue establecida. Como se verá, esta definición toma elementos de diferentes propuestas estudiadas. Luego, presentamos un estudio de las expresiones que transmiten opiniones en los textos, basado en el análisis de corpus.

A partir de este análisis confeccionamos un conjunto de ejemplos representativos de las diferentes formas en que las opiniones son expresadas en textos periodísticos del español. Estos ejemplos servirán para ilustrar diversas explicaciones que se darán a lo largo del documento, tanto para las descripciones teóricas como para las secciones destinadas a las implementaciones informáticas.

4.2.1 Definición de opinión

Uno de los trabajos más importantes del área es el de Wiebe, Wilson y Cardie (2005), que proponen un esquema de anotación para estados internos (*private states*), expresión que abarca opiniones y emociones, definiéndolos de este modo:

*"We can further view private states in terms of their functional components as states of experiencers holding attitudes, optionally toward targets"*¹⁶.

Por otro lado, en (Bethard et al, 2004, 2005) se define la opinión del siguiente modo:

*"We define opinion as a sentence, or part of a sentence, that would answer the question 'How does X feel about Y?'"*¹⁷.

Liu (2010) define la opinión como una tupla de cinco elementos, (o, f, oo, h, t) , en la cual o es el objeto sobre el cual se opina, f es un conjunto de características (*features*) del objeto, oo es la orientación de la opinión, h es la fuente (*holder*) y t es la fecha (*time*) de emisión de la opinión.

Para el español, la definición de opinión que se encuentra en el Diccionario de la Real Academia Española (Web DRAE) es la siguiente:

"Dictamen o juicio que se forma de algo cuestionable".

En el marco de nuestro trabajo, por un lado, coincidimos con (Bethard et al, 2004, 2005) en considerar la opinión como una expresión lingüística. Ellos hablan de oración o parte de una oración, pero, debido a las diversas formas de expresar opiniones que observamos en los corpus, optamos por hablar de "segmento de texto". Por otro lado, en vez de definir la opinión en función del tipo de pregunta que esta puede responder, la definimos en función del contenido que estas expresiones transmiten, acercándonos a la definición del DRAE. Por último, nos parece apropiado incluir en la definición algunos de los componentes de la opinión, por lo tanto, al igual que lo hacen (Wiebe et al, 2005), hacemos mención al participante del discurso o fuente que opina, y al asunto sobre el cual este opina. Nuestra definición es la siguiente:

16 Nuestra traducción: Podemos definir los estados internos en términos de sus componentes funcionales como estados de experimentantes demostrando actitudes, opcionalmente respecto a algo.

17 Nuestra traducción: Definimos opinión como una oración, o parte de una oración, que podría responder la pregunta "¿Qué siente/opina X en relación a Y?".

Opinión: Segmento de texto que transmite las expresiones o posturas de alguna fuente sobre algún asunto.

Para referirnos al contenido transmitido por la opinión, es decir, las expresiones que mencionamos en la definición, utilizamos el término mensaje.

Los nombres que elegimos para hacer referencia a cada elemento de la opinión fueron tomados de diferentes propuestas, por un lado, (Wiebe et al, 2005) hablan de fuente (*source*), por otro lado, los términos asunto y mensaje fueron tomados de ADESSE (Albertuz, 2007).

En nuestro trabajo, nos basamos principalmente en la identificación de opiniones de diferentes fuentes, no partimos de la detección de subjetividad. Luego de encontrar segmentos de diferentes fuentes buscamos establecer su orientación semántica que, en algunos casos, podrá ser neutra.

Por esta razón, necesitamos definir un elemento más, que llamamos predicado de opinión, que es la expresión que el autor del texto utiliza para hacer referencia a una opinión de alguna fuente.

El hecho de centrarnos en la identificación de segmentos atribuibles a diferentes fuentes pone en el centro de nuestra investigación al fenómeno del discurso reproducido, pero estudiamos también, como posibles formas de transmitir opiniones, otros mecanismos: elementos introductores de fuente como *según*, verbos de creencia, valoración o sentimiento, uso de nombres, en general derivados de verbos de discurso o de las otras clases semánticas mencionadas.

Esto nos diferencia de muchos trabajos del área, que buscan principalmente elementos de contenido subjetivo para, a partir de ellos, identificar la opinión, que incluye en general al segmento que llamamos mensaje, y determinar su fuente y su asunto (Choi et al, 2005 y 2006, Kim et al, 2006).

Además de identificar los elementos de la opinión, nos interesa conocer su orientación semántica, es decir, queremos saber si la fuente se muestra a favor o en contra del asunto sobre el que opina. Hablamos de orientación semántica siguiendo a (Hatzivassiloglou et al, 1997, Turney, 2002) que utilizan la expresión *semantic orientation*, otros autores se refieren a esta propiedad como polaridad (*polarity*) (Wiebe et al, 2005, Esuli, 2006), valencia (*valence*) (Polanyi et al, 2004) u orientación de la opinión (*opinion orientation*) (Liu, 2010).

Nuestra propuesta consiste en determinar en primer lugar la presencia de una opinión, identificando sus elementos, y luego deducir el valor de orientación semántica de dicha opinión. Para esto buscamos indicios de este valor en cada elemento identificado. Esto nos lleva a definir la propiedad orientación semántica para cada elemento de la opinión, además de definirla para la opinión completa.

A continuación, describimos el corpus de trabajo y luego presentamos un conjunto de ejemplos seleccionados. Posteriormente explicamos las particularidades de cada ejemplo.

4.2.2 Corpus de trabajo

Trabajamos con textos de prensa en español tomados de diferentes fuentes:

- textos de prensa de Uruguay, corpus Cor In (Grassi, 2001) y conjuntos de textos de la prensa digital (www.montevideo.com.uy y www.observa.com.uy).
- corpus del español de diversos países de habla hispana (Davies, 2002), disponible para hacer consultas en línea, se puede consultar por palabra, lema y categoría gramatical. También es posible crear listas personalizadas de palabras, lo cual fue de gran utilidad para este trabajo ya que pudimos definir listas como verbos de opinión, verbos positivos, verbos negativos, etc., que fueron fundamentales para poder realizar las búsquedas de ejemplos.

4.2.3 Conjunto de ejemplos seleccionados

Presentamos un conjunto de ejemplos que ilustran las variantes más frecuentes en la forma de expresar opiniones en los textos de prensa. Numeramos cada ejemplo con la notación Ei, de modo de poder referirnos a ellos desde cualquier parte del documento. En el Anexo A se incluye este mismo conjunto de ejemplos. En cada ejemplo, indicamos cada elemento de la opinión con la notación utilizada anteriormente: predicado en negrita, fuente subrayada, asunto en cursiva y mensaje con fondo gris. A continuación de los ejemplos, comentamos sus características más relevantes.

- (E1) Consultado *sobre la lentitud de los procesos judiciales uruguayos* Carranza **respondió**: "Hay una situación de un muy alto número de presos sin condena, hay que agilizar los procesos".
- (E2) El vicepresidente del Instituto Nacional de Carnes de Uruguay, Fernando Pérez Abella, **dijo** que las conclusiones son muy alentadoras.
- (E3) El 56% de los estadounidenses **creo** que la invasión de Irak en 2003 fue un error.
- (E4) No existe más durante este gobierno el vocero presidencial como figura de estructura presidencial, **afirmó** Menem.
- (E5) La iglesia católica tomó partido durante el proceso revolucionario y **apoyó a** *Victoriano Huerta*.
- (E6) Mientras que otros sectores como el Partido Comunista, Asamblea Uruguay, el Partido Socialista y Alianza Progresista **rechazaron** la propuesta.
- (E7) Clinton formuló su **declaración** durante rueda de prensa gráfica junto con el primer ministro irlandés Bertie Ahern.
- (E8) Nosotros coordinamos esta relación con la Comisión Nacional Bancaria y tenemos mucho **apoyo** *de la Comisión*.
- (E9) Este sistema se utiliza en Estados Unidos desde 1982, **según** Roque Pifarré.
- (E10) **De acuerdo a** un informe de la Dirección de Prensa de Palacio de Gobierno, la reunión organizada por el Ministerio de la Presidencia se realizará en el Hotel París desde las primeras horas de hoy.

- (E11) Entrevistado por el semanario Crónicas, el abogado Gustavo Salle **habló** *del pedido de extradición de Henry Kinssinger; sus discrepancias con el Gobierno del Frente Amplio, el caso Young y el trabajo del Secretario de Presidencia, Gonzalo Fernández durante la dictadura.* "Este Frente Amplio vende una imagen, escenifica determinada postura ideológica y en la práctica edifica otra totalmente opuesta. Se vende como socialista y es ortodoxamente neoliberal. Si yo hubiera estado preso y hubiese sido torturado y hoy tuviera que enfrentar con esa pesada carga existencial y afectiva, la traición política e ideológica que está consumando la cúpula del Frente Amplio no sé cuál sería mi reacción".
- (E12) Mahuad **dijo** *que su país no rechazaba la idea de crear un parque binacional en la zona en disputa.*¹⁸
- (E13) La Alianza Nacional de Agricultores Independientes de Cuba **se pronuncia** *en contra de tales prácticas de corrupción.*
- (E14) Los optimistas **sostienen** *que la probabilidad de que esto suceda es muy baja pues el comportamiento que han mostrado los embalses hasta el momento es un indicador de que San Pedro será benévolo y los niveles de agua no llegarán a extremos.*

La definición que dimos para la opinión en la sección 4.1 abarca tanto citas a palabras pronunciadas por las fuentes como referencias a lo que estas piensan. En (E1) se transmite lo que *Carranza* dijo respecto a *la lentitud de los procesos judiciales uruguayos*, el verbo *respondió* indica que esa persona se pronunció sobre ese tema, mientras que en (E3) se transmite lo que *el 56% de los estadounidenses* creen, sin que se haga mención a la existencia de un acto verbal por parte de ellos. En (E6) se menciona una postura positiva de *la iglesia católica* respecto a un tema, pero tampoco se hace referencia a la existencia de un acto verbal.

El recurso más usual para transmitir opiniones de otros es el discurso reproducido, tanto discurso directo (E1) como discurso indirecto (E2). En los casos típicos de discurso reproducido no hay una mención explícita al asunto, como en (E2). De todos modos, es posible encontrar casos de discurso reproducido que incluyen un segmento especial para introducir el asunto, como (E1). Los verbos introductores de discurso reproducido suelen aparecer como incisos finales, después de una coma que los separa del mensaje, como en (E4).

El hecho de que no haya un segmento específico que introduzca el asunto de la opinión, como en (E2), no significa que no se pueda saber sobre qué asunto se opina. Analizando el interior del mensaje podemos saber que en (E2) se habla sobre *las conclusiones* (aunque es necesaria más información contextual). Como dijimos, en nuestro trabajo solo identificamos menciones al asunto de la opinión cuando este aparece en forma explícita, nos referimos con esto a los segmentos como *sobre la lentitud de los procesos judiciales uruguayos*, de (E1).

Es posible que el que escribe haga mención a la opinión de otro utilizando un verbo que resume de algún modo lo que este opina, como *apoyó* en el ejemplo (E5) o *rechazaron* en (E6). Estos verbos transmiten la orientación semántica de la fuente, mientras que los verbos de

18 En este ejemplo, la anidación de opiniones vuelve un poco confusa la notación utilizada para indicar los elementos de cada opinión. Dentro del mensaje (en fondo gris) de la opinión correspondiente a la fuente *Mahuad*, hay otra opinión con fuente *su país* (subrayado), predicado *no rechazaba* (en negrita) y asunto *la idea de crear un parque binacional en la zona en disputa* (en cursiva).

comunicación verbal, en este sentido, son neutros. Como vimos en el estudio de las clases verbales, particularmente en las de ADESSE, encontramos verbos de este tipo en las clases valoración, aceptación y sensación. Por lo general, en estos casos no se transmite el mensaje de la opinión. El asunto sobre el que se opina es el complemento directo del verbo, como *Victoriano Huerta* en (E5) y *la propuesta* en (E6).

Para referirnos a todos los verbos que permiten la expresión de opiniones, verbos de comunicación, creencia, valoración, aceptación y sensación, hablamos de verbos de opinión.

Otra forma de transmitir opiniones es a través de la utilización de nombres, como *declaración* en (E7) o *apoyo* en (E8), se trata por lo general de nombres derivados de verbos de opinión. En estos casos, la fuente se expresa generalmente a través de un complemento del nombre introducido por *de*, como *de la Comisión* en (E8). Es común también la utilización de estos nombres en estructuras con verbos soporte o de apoyo, como *formuló su declaración* de (E7) en donde la fuente es el sujeto del verbo soporte. Otras construcciones de este tipo son *emitir una opinión*, *realizar un comentario*, *eleva una declaración*.

En los ejemplos (E9) y (E10) vemos que la expresión de opiniones se basa en la utilización de ciertos elementos introductores de fuente, como *según* y *de acuerdo a*, muy comunes en los textos de prensa. En estos casos, la fuente es el segmento que sigue a estas expresiones, el resto de la oración constituye el mensaje. La fuente introducida por estos elementos suele estar al principio de la oración, como en (E10) o al final, como en (E9). También puede aparecer como inciso en la mitad de la oración.

La opinión no necesariamente se circunscribe al ámbito de la oración, los elementos que la componen pueden estar en oraciones diferentes, como se muestra en (E11). En particular, el mensaje puede abarcar varias oraciones.

En todo texto existe un primer nivel de cita que es la mención a opiniones de otros por parte del autor del texto. Pero dentro de las opiniones de otros participantes puede haber también citas, por lo que las opiniones pueden estar anidadas dentro de otras. Es el caso del ejemplo (E12) en el cual se transmite una opinión de la fuente *Mahuad*, cuyo mensaje es la subordinada completa, pero dentro de la subordinada se hace mención a otra opinión, con fuente *su país* y asunto *la idea de crear un parque binacional en la zona en disputa*. La primera fuente es introducida por el predicado *dijo* y la segunda por el predicado *rechazaba*.

La orientación semántica de la opinión

Dijimos anteriormente que calculamos la orientación semántica de las opiniones a partir de la orientación semántica de sus elementos. Veremos cómo se aplica esta idea a nuestros ejemplos.

La mayor parte de los predicados de opinión de nuestros corpus de análisis son verbos de comunicación y creencia o sus nominalizaciones (E1, E2, E3, E4, E7, E11, E12, E13) y predicados preposicionales como *según* y similares (E9, E10). Por lo tanto, se trata de predicados neutros. También el asunto es neutro en la mayoría de los casos, ya que, o bien son introducidos por expresiones neutras como *sobre*, *en cuanto a*, *en lo referente a* (E1, E11), o bien se trata de grupos nominales, como el asunto de *rechazaba* en (E12). Las fuentes, en general, son grupos nominales que hacen mención a alguna persona o institución, por lo que suelen ser también neutras.

El elemento que permite determinar la orientación semántica en todos estos casos es el mensaje, ya que allí se transmite el contenido de la opinión. Igualmente, existe la posibilidad de que este elemento sea también neutro, como en (E9) y (E10), caso en el cual la opinión completa es neutra, de modo que se estaría transmitiendo información objetiva. En nuestro trabajo preferimos hablar de opiniones neutras u objetivas, similares al discurso reproducido objetivo de (Wiebe et al, 2005), y no excluirlas de nuestro estudio por tratarse, de todos modos, de información que se atribuye a una fuente diferente al autor del texto.

En la mayoría de los casos, la orientación semántica de la opinión se obtiene, por lo tanto, a partir del mensaje. Sabemos que en (E2) el mensaje es positivo y en (E11) el mensaje es claramente negativo, pero determinar estos valores en forma automática no es tan evidente. No incluimos en nuestro trabajo el estudio de la orientación semántica del mensaje, como ya se dijo, pero es una de nuestras líneas de investigación para trabajar en el futuro.

En cuanto a la orientación semántica contenida en los restantes elementos de la opinión, podemos decir que la opinión de (E5) es positiva y la de (E6) es negativa, dado que los predicados utilizados por el autor del texto para transmitirlos son positivo (*apoyó*) y negativo (*rechazaron*), respectivamente. En cuanto a la orientación del asunto, en (E13) la expresión *en contra de* permite deducir que la opinión es negativa. La fuente puede contener también información sobre la orientación subjetiva, como la palabra *optimistas* en (E14), determinando así la orientación total de la opinión, pero es muy poco frecuente.

En el capítulo que sigue presentamos un modelo teórico para las opiniones que intenta contemplar todas sus posibles realizaciones lingüísticas. El modelo se centra en los predicados de opinión y considera a los restantes elementos -fuente, asunto y mensaje- argumentos de estos predicados. Dado que no siempre se trata de argumentos en el sentido estricto, es decir, constituyentes requeridos por los predicados, en nuestro modelo los llamamos argumentos de opinión.

4.3 Síntesis

Explicamos la metodología seguida para desarrollar nuestra propuesta para el reconocimiento de opiniones de diferentes fuentes en textos periodísticos. A grandes rasgos, nuestro trabajo se organizó en varias etapas consecutivas: definición de opinión, análisis de corpus, modelo para los predicados de opinión y sus argumentos, creación de un repertorio de predicados de opinión, sistema informático basado en reglas, sistema informático basado en aprendizaje automático y sistema informático combinado.

De cada etapa de trabajo surgen elementos que son retomados en las etapas posteriores. El modelo elaborado se basa en la definición de opinión y en el análisis del corpus, también toma muchos elementos de los estudios lingüísticos del capítulo 3. El repertorio de predicados incluye información sobre la jerarquía de tipos y sus restricciones, definidas en el modelo, y es utilizado por los sistemas informáticos desarrollados.

Además de explicar la metodología de trabajo, en este capítulo presentamos la definición de opinión, que se basa en diferentes propuestas del área, y mostramos diferentes ejemplos de opiniones tomados de nuestro corpus de trabajo. Confeccionamos un conjunto de ejemplos representativos de los principales fenómenos que hay que tener en cuenta para el reconocimiento

automático de las opiniones. Este conjunto se incluye en el anexo A, ya que los ejemplos allí contenidos son mencionados en diferentes capítulos posteriores a este.

5. Un modelo para los predicados de opinión y sus argumentos

Como vimos en los capítulos precedentes, nuestra concepción de opinión se basa en la presencia de un predicado de opinión que puede ser un verbo de opinión, un nombre, muchas veces derivado de uno de estos verbos, o una preposición como *según* o locución prepositiva como *de acuerdo con*. Los demás componentes de la opinión, fuente, asunto y mensaje, son generalmente argumentos del predicado.

Nuestro modelo establece las propiedades sintáctico-semánticas de las piezas léxicas que pueden constituir predicados de opinión. A través de ciertos rasgos de estas piezas se establece la forma que toman los diferentes argumentos de la opinión y las diferentes estructuras sintácticas en las que los predicados y sus argumentos pueden participar. Las propiedades que comparten diferentes piezas léxicas dan lugar a diferentes clases de verbos. Esta información es utilizada luego por los sistemas informáticos que desarrollamos.

Las propiedades sintáctico-semánticas que incluimos en el modelo para cada tipo de predicado, si bien están basadas en las propiedades presentadas en los diferentes enfoques lingüísticos que estudiamos, también se nutren del análisis de ejemplos de nuestro corpus de trabajo.

En relación al modelo que aquí presentamos, un antecedente importante es el estudio de los verbos de sentimiento de Mathieu (2000). Allí los verbos son analizados y organizados según un criterio semántico, teniendo en cuenta también propiedades sintácticas. Se definen clases de verbos y para cada clase se incorporan propiedades sintácticas al estilo del *Lexique-Grammaire* (Gross, 1975), en donde los verbos son clasificados en diferentes tablas, según las estructuras sintácticas básicas en las cuales estos participan. Mathieu identifica los argumentos de cada verbo y determina los roles semánticos característicos de los verbos de sentimiento (experimentante, causa, objeto), basándose en los trabajos de Fillmore (1968). Además, las clases semánticas están conectadas por relaciones de intensidad y de antonimia. Por ejemplo, se indica que hay un grado creciente de intensidad entre las siguientes clases: *interés* -> *amor* -> *pasión* -> *fascinación*. Entre *amor* y *odio* se identifica una relación de antonimia.

Para especificar nuestro modelo utilizamos estructuras de rasgos tipificadas (*Typed Feature Structures*, TFS) (Carpenter, 1992). Muchas de las decisiones que tomamos en la definición del modelo están inspiradas en las gramáticas de tipo HPSG (*Head-driven Phrase Structure Grammar*) (Pollard et al, 1987, 1994), en particular la forma en que se relacionan los rasgos sintácticos y los rasgos semánticos.

Es importante aclarar que utilizamos TFS solamente como formalismo para el modelado de piezas léxicas, no con la intención de definir una gramática completa, en particular, no vamos a definir reglas de formación de frases.

El modelo especificado mediante TFS no se integra directamente a los sistemas informáticos que se describen en la cuarta parte de este documento, sino que se toma la información en él contenida, y se aplica de diferentes maneras, según la forma en que se implementa cada sistema.

Entre diversos trabajos basados en TFS, comentamos, por un lado, la propuesta de Asher, Benamara y Mathieu (2008) en el área Minería de Opiniones y, por otro lado, el estudio de los verbos de ciertas clases semánticas del español de Taulé (1995).

Asher, Benamara y Mathieu (2008) utilizan TFS para representar la opinión, aunque con características diferentes de lo que aquí planteamos. En ese trabajo, el objeto opinión es representado mediante una TFS, dentro de la cual se definen rasgos semánticos (categoría, polaridad, intensidad) y los componentes de la opinión (fuente, asunto y mensaje). Se definen reglas para la combinación de esos objetos, en base al tipo de relación retórica existente entre ellos. El objetivo principal de la propuesta es obtener el valor de polaridad de las combinaciones de diferentes opiniones. No se modelan propiedades sintácticas.

El trabajo de Taulé (1995), si bien no tiene relación con el análisis de opiniones, se asemeja a nuestra propuesta porque se utilizan TFS para especificar las piezas léxicas correspondientes a verbos del español pertenecientes a determinadas clases semánticas (cocción, ingestión, movimiento y meteorología). Las estructuras de rasgos que allí se utilizan están más cercanas a las que proponemos nosotros, apuntando a modelar las combinaciones posibles entre ciertos verbos y sus argumentos. Las piezas léxicas incluyen una gran cantidad de información sintáctica y semántica: clase semántica, estructura argumental y eventiva, posibles diátesis, entre otras. Este trabajo forma parte del proyecto Acquilex (Copestake, 1992, Web Acquilex), en el marco del cual se han desarrollado léxicos basados en TFS para diversos idiomas. Este proyecto dio lugar al desarrollo del sistema LKB (*Lexical Knowledge Base*) (Copestake, 1992, 2002), que permite la generación y aplicación de gramáticas basadas en TFS.

En lo que sigue presentamos en primer lugar una breve introducción a las estructuras de rasgos tipificadas. Luego describimos el modelo definido para los predicados de opinión.

5.1 Breve introducción a las estructuras de rasgos tipificadas

Las estructuras de rasgos (*Feature Structures*, FS) pueden verse como matrices de pares atributo-valor (*Attribute Value Matrix*, AVM). En este marco se habla de *features*, término que se traduce en general como rasgo, más que de atributos. Como los valores asociados a cada rasgo son tipos, organizados en una estructura de tipos, hablamos de Estructura de Rasgos Tipificadas (*Typed Feature Structures*, TFS).

Una jerarquía de tipos debe cumplir las propiedades siguientes:

- Existe un elemento único del cual todos los demás son descendientes (o subtipos).
- No hay ciclos.
- Para todo subconjunto de tipos, si existen descendientes comunes, entonces hay un único elemento entre ellos que es más general que todos los demás.

Cada tipo puede tener asociada una restricción, representada como TFS. Como los tipos están organizados en la jerarquía, las restricciones son heredadas por los subtipos. Esta es una propiedad importante ya que permite definir restricciones específicas para cada tipo, sin necesidad de incluir las restricciones definidas para tipos más generales.

Una introducción a las TFS puede leerse en (Copestake, 2002).

5.2 Representación de los predicados de opinión y sus argumentos

Entre los componentes de la opinión, el predicado ocupa un lugar destacado por ser el elemento predicativo que rige los argumentos a través de los cuales se expresan los componentes restantes. Algunos predicados rigen otros argumentos, por ejemplo, algunos verbos de comunicación toman como argumento un receptor (*X le dijo a Y que ...*), pero en nuestro modelo no tendremos en cuenta argumentos que no expresen los componentes que definimos para la opinión (fuente, asunto y mensaje).

Para algunos de los verbos que consideramos de opinión, alguno de los componentes de la opinión pueden no estar seleccionado por el verbo. Por ejemplo en ADESSE, para los verbos de aceptación (*aceptar, rechazar*) o sensación (*gustar, molestar*) no se define un argumento que coincida con lo que llamamos mensaje. Sin embargo, encontramos ejemplos como (5.1) en donde uno de estos verbos, *apoyar*, tiene asociada una cita directa.

- (5.1) ... hasta que el visitador examinase la conducta de cuantos habían intervenido en fundar las colonias, y lo apoyaba de este modo: "Si se hallare que hemos malversado, seremos dignos del mayor castigo: si se viere que hemos malogrado la empresa ... "

Comenzamos a construir nuestra jerarquía de tipos a partir del tipo genérico *top*, que es el tipo más general de la jerarquía, requerido por la definición de jerarquía de tipos dada en la sección 6.1. En un primer nivel está el tipo *componente_opinión* que tiene como subtipos, por un lado, el predicado de opinión y, por otro lado, los restantes componentes, que llamamos argumentos de opinión. Una primera versión de nuestra jerarquía de tipos se muestra en la figura 5.1.

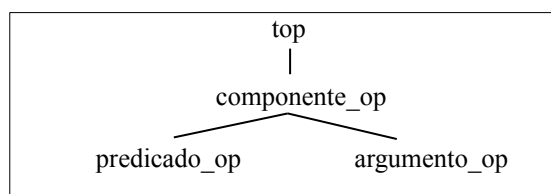


Figura 5.1 Primera versión de la jerarquía de tipos

Hablamos de argumento de opinión, y no de argumento a secas, ya que, como dijimos, estamos dejando de lado otros posibles argumentos, por no estar relacionados con la información que queremos modelar y además porque estamos considerando argumento de opinión a ciertos segmentos que no son estrictamente argumentos del verbo (como el mensaje del ejemplo 5.1).

Para todos los componentes de la opinión, predicado y argumentos, definimos rasgos sintácticos y rasgos semánticos. Más adelante veremos cómo se especifica la asignación de roles semánticos a los argumentos de los predicados, es decir, las relaciones entre propiedades sintácticas y semánticas, gracias a la posibilidad que brindan las TFS's de coindexar valores. Incorporamos estos nuevos rasgos y sus tipos a la jerarquía, como muestra la figura 5.2.

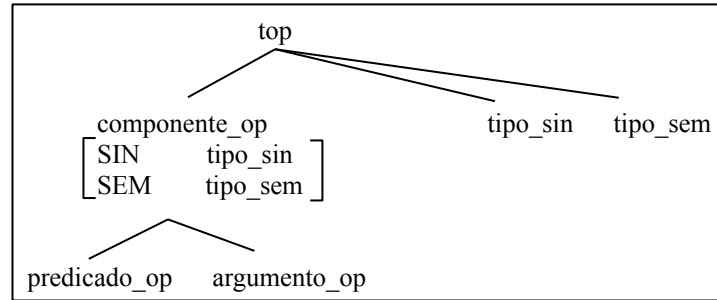


Figura 5.2 Restricciones del tipo `componente_op`

Es importante recordar que las restricciones de un tipo cualquiera de la jerarquía (tfs asociada a un tipo) son heredadas por todos sus subtipos, por lo que los tipos *predicado_op* y *argumento_op* heredan los rasgos SIN y SEM del tipo *componente_op*.

Definimos un rasgo sintáctico, la categoría gramatical, y dos rasgos semánticos, la orientación semántica y un índice, que comparten todos los componentes de la opinión. Estos rasgos se definen como restricciones de los tipos *tipo_sin* y *tipo_sem*.

Luego, definimos algunos rasgos sintácticos y semánticos más, que modelan propiedades específicas de cada tipo de componente. Para esto debemos especializar los tipos *tipo_sin* y *tipo_sem*, definiendo subtipos: *tipo_sin_pred*, *tipo_sin_arg*, *tipo_sem_pred* y *tipo_sem_arg*. Estos tipos se asignan a los rasgos SIN y SEM de los dos subtipos de *componente_op*: *predicado_op* y *argumento_op*. En la figura 5.3 mostramos la jerarquía de tipos con estas incorporaciones.

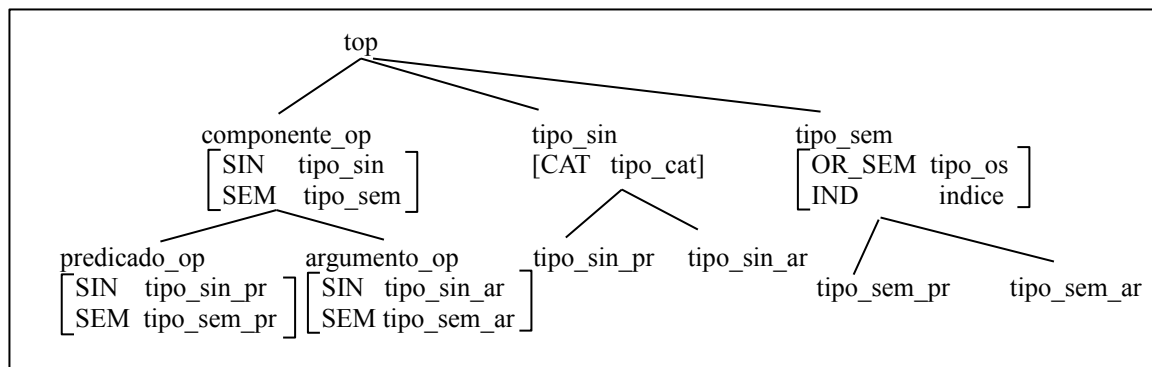


Figura 5.3 Restricciones de los tipos `predicado_op` y `argumento_op`

El tipo *tipo_os* tiene como subtipos tres valores posibles para la orientación semántica: *neutro*, *negativo* y *positivo*.

El tipo *indice* tiene como subtipos las letras que normalmente se utilizan como índices: *i*, *j*, *k*, etc. El rasgo IND nos permitirá establecer correspondencias entre elementos.

El tipo correspondiente a la categoría, *tipo_cat*, tiene varios subtipos que son necesarios para todas las posibles categorías del predicado de opinión y de los argumentos de la opinión, algunas son: nombre (nom), verbo (ver), preposición (prep), proposición (prop), cita directa (citaD, representa una proposición entre comillas), cita indirecta (cital, representa una proposición subordinada con *que*).

A continuación, presentamos las propiedades de los argumentos de opinión y luego nos concentramos en las propiedades del predicado de opinión, ocupándonos principalmente del verbo, por ser este el elemento más complejo dentro de nuestro modelo.

5.2.1 Los argumentos de la opinión

Como propiedades sintácticas de los argumentos, además de la categoría gramatical, necesitamos establecer la función sintáctica que cada argumento cumple respecto al verbo. Esta información se codifica mediante la definición del rasgo FUN para el tipo *tipo_sin_ar*.

El tipo de FUN es *tipo_fun*, que puede tomar los valores: *suj* (sujeto), *cd* (complemento directo), *ci* (complemento indirecto), *cp* (complemento preposicional). Para este último valor, es necesario incluir la lista de preposiciones regidas por cada pieza léxica. Los tipos *tipo_sin_ar* y *tipo_fun* quedan como se muestra en la figura 5.4.

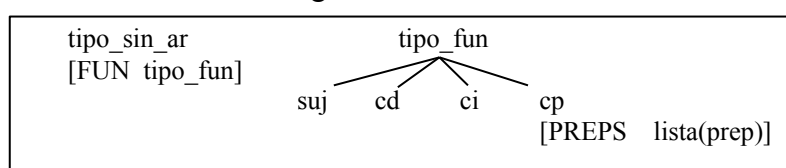


Figura 5.4 Tipos *tipo_sin_ar* y *tipo_fun*, subtipos y restricciones

Definimos el tipo genérico *lista(t)* a partir del cual, instanciando *t*, se obtienen listas de elementos de tipos determinados, por ejemplo *lista(prepp)*. Una lista se define en forma recursiva como se muestra a continuación:

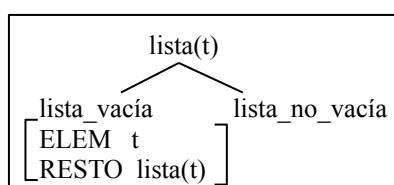


Figura 5.5 Tipo lista

En nuestro trabajo nos centramos principalmente en los predicados de opinión, no definimos, por lo tanto, más propiedades semánticas para los argumentos que la orientación semántica (a través del rasgo OR_SEM). Además establecemos el rol semántico de cada argumento respecto al predicado mediante coindexación entre su índice y los rasgos semánticos de los predicados.

5.2.2 El predicado de opinión

En primer lugar, necesitamos un rasgo que permita establecer qué argumentos toma el predicado. Este rasgo es una lista de elementos de tipo *argumento_op*, ya que, como dijimos anteriormente, solamente incluimos en nuestro modelo los argumentos de los predicados que representan algún componente de la opinión.

En este punto, en lo que respecta a los verbos, seguimos a Pollard y Sag (1987) que incluyen al sujeto dentro de la lista de argumentos del verbo. Otra posibilidad interesante sería tomar la propuesta de Sag, Wasow y Bender (2003), más cercana a la teoría de la X Barra, en donde se establece para las piezas léxicas un rasgo valencia que tiene definidos a su vez dos rasgos: uno

para el especificador (rasgo SPR) y otro para los complementos (rasgo COMPS). En ese marco el sujeto se modela como especificador del verbo, quedando separado de los complementos.

En principio estamos considerando que todos los argumentos de opinión son opcionales. Si bien en algunos casos algunos de los argumentos deberían estar requeridos por los predicados (*hablar sobre algo*, *rechazar algo*, *decir algo*) en textos reales existen variadas situaciones en las cuales los argumentos pueden omitirse siendo posible recuperarlos en el contexto.

Las restantes propiedades sintácticas de los predicados dependen de su categoría gramatical: verbo, nombre y preposición (o locución prepositiva). Necesitamos, por lo tanto, un nuevo nivel de especialización en la jerarquía para lo cual definimos tres subtipos para *predicado_op*. Estos subtipos quedan especificados por la asignación de valores al rasgo CAT.

Además, para definir las propiedades particulares de cada tipo de predicado, volvemos a especializar *tipo_sin_pr*.

Ilustramos, en la figura 5.6, el resultado de introducir en la jerarquía las modificaciones que se desprenden de estos últimos comentarios, mostrando solamente los nodos *predicado_op* y *tipo_sin_pr*.

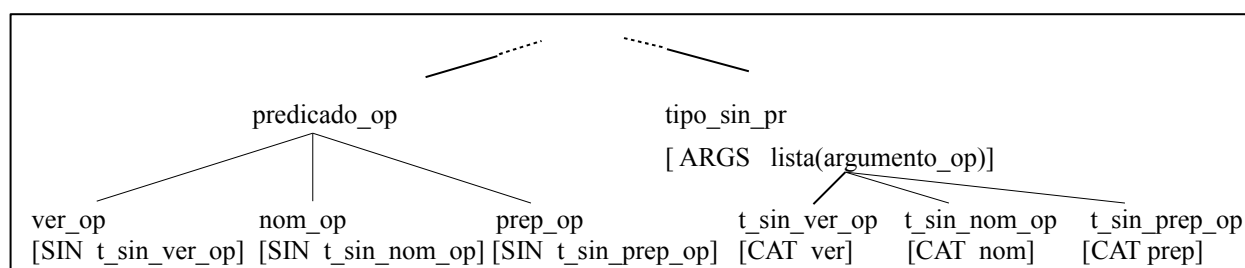


Figura 5.6 Fragmento de la jerarquía de tipos que muestra los tipos *predicado_op* y *tipo_sin_pr*

Los tipos *ver*, *nom* y *prep* son subtipos de *tipo_cat*, como ya se dijo, de otro modo la jerarquía no sería correcta, ya que el rasgo CAT tiene asignado el tipo *tipo_cat*.

En cuanto a las propiedades semánticas de los predicados, nos interesa especialmente modelar los roles semánticos que corresponden a los componentes de la opinión, por lo cual dentro de un rasgo que denominamos PRED (por 'predicación') especificamos los tres argumentos que participan en la predicación que establecen estos predicados: fuente, asunto y mensaje. Estas propiedades se definen como restricciones del tipo *tipo_sem_pr*.

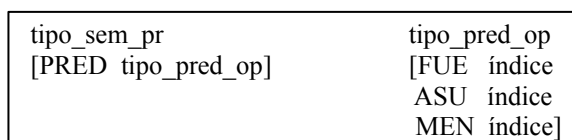


Figura 5.7 Tipos *tipo_sem_pr* y *tipo_pred_op*

Para continuar con la presentación del modelo para opiniones, describimos, en primer lugar, los verbos de opinión, luego vemos cómo se modelan los nombres de opinión y los predicados de opinión preposicionales.

5.2.2.1 Los verbos de opinión

Las diversas clases de verbos que incluimos dentro de la clase más amplia de los verbos de opinión, presentados en el capítulo 3, presentan diferentes propiedades sintácticas y semánticas que creemos importante modelar.

En el plano sintáctico incluimos el mapeo entre funciones sintácticas y roles semánticos, por lo que es más exacto hablar de propiedades sintáctico-semánticas, por un lado, y propiedades semánticas, por otro. En este trabajo solamente modelamos una propiedad semántica, la orientación semántica.

Propiedades sintáctico-semánticas de los verbos de opinión

Las propiedades tenidas en cuenta para los verbos, además de las ya mencionadas para los predicados en general, son:

- Mapeo entre rol semántico y función sintáctica de los argumentos de opinión.
- Categoría gramatical de los argumentos de opinión.
- Aceptación de voz pasiva (*Y fue criticado por X*).
- Posibilidad de topicalización del asunto, fenómeno de alta frecuencia en los textos de prensa (*Sobre el tema, X opina que ...*), mediante el uso de la preposición *sobre* o de alguna locución prepositiva característica (*en cuanto a, en lo que concierne a, en lo que respecta a, etc.*).
- Aceptación de inversión de sujeto dentro de un inciso de discurso final o intercalado (*..., dijo X,*).

El mapeo entre roles semánticos y funciones sintácticas, propiedad (a), se representa mediante la coindexación entre los argumentos del verbo, que tienen definido un rasgo FUN para establecer su función sintáctica, y los roles semánticos posibles para ese verbo, definidos dentro del rasgo PRED del tipo *tipo_sem_pr*.

La categoría gramatical de los argumentos de opinión, propiedad (b), se establece a través del rasgo CAT de *tipo_sin* que está definido para todos los componentes de la opinión (predicado y argumentos), por lo tanto también para su subtipo *tipo_sin_ar*.

Las últimas tres propiedades, (c), (d) y (e), se representan mediante la introducción de varios rasgos para la restricción del tipo *t_sin_ver_op* (que es uno de los subtipos de *t_sin_pr*, como se muestra en la figura 5.6). La restricción resultante para el tipo *t_sin_ver_op* se muestra en la figura 5.8:

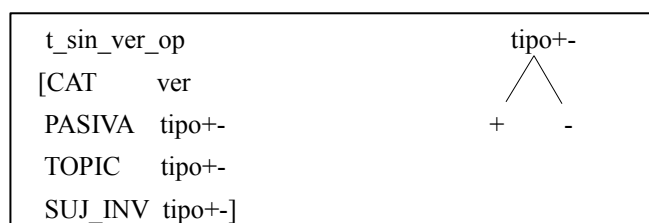


Figura 5.8 Tipos *t_sin_ver_op* y *tipo+-*

En la tabla 5.1 mostramos cómo se comportan respecto a las propiedades definidas (a - e) los verbos de las diferentes clases de ADESSE seleccionadas para formar nuestra clase de verbos de opinión.

A partir de los datos de la tabla, buscamos la manera de organizar las diferentes clases de verbos definiendo tipos y subtipos de modo de factorizar hacia los tipos más generales las propiedades compartidas por diferentes clases, logrando así una representación económica.

clase	subclase	verbos	argumento opinión			pasiva	asunto topicalizado	sujeto invertido
			rol sem	función sint	forma sint			
C1: comunicación	C11: disc reprod	<i>decir</i> <i>declarar</i>	fuente	sujeto	gn	-	+	+
			asunto	cp/cd	gp/gn			
			mensaje	cd	citaI/citaD			
	C12: disc referido	<i>hablar</i> <i>referirse</i>	fuente	sujeto	gn	-	-	-
			asunto	cp/cd	gp/gn			
			mensaje		citaD			
C2: creencia		<i>creer</i> <i>considerar</i>	fuente	sujeto	gn	-	+	+
			asunto	cp/cd	gp/gn			
			mensaje	cd	citaI/citaD			
C3: evaluación		<i>felicitar</i> <i>condenar</i>	fuente	sujeto	gn	+	-	-
			asunto	cd/cp	gn/gp/prop			
			mensaje		citaD			
C4: aceptación		<i>aceptar</i> <i>rechazar</i>	fuente	sujeto	gn	+	-	-
			asunto	cd/cp	gn/gp/prop			
			mensaje		citaD			
C5: sensación	C51	<i>apreciar</i> <i>detestar</i>	fuente	sujeto	gn	+	-	-
			asunto	cd/cp	gn/gp/prop			
			mensaje		citaD			
	C52	<i>gustar</i> <i>molestar</i>	fuente	ci	gp	-	-	-
			asunto	sujeto	gn/prop			
			mensaje		citaD			

Tabla 5.1: Propiedades sintáctico-semánticas de los verbos de opinión

Tomamos como primer criterio de clasificación el tipo de argumento que cumple la función sujeto. Para las clases comunicación (C1), creencia (C2), evaluación (C3), aceptación (C4) y un subconjunto de verbos de sensación (C51), cuyo sujeto es la fuente de la opinión, definimos el tipo *vo_sf* (verbos de opinión con sujeto fuente), mientras que para los restantes verbos de sensación (C52), cuyo sujeto es el asunto, definimos el tipo *vo_sa* (verbos de opinión con sujeto asunto).

Dentro del grupo de verbos cuyo sujeto tiene rol fuente (*vo_sf*), definimos un nuevo nivel de especialización que agrupa, por un lado, los verbos de las clases C1 (comunicación) y C2 (creencia), bajo el tipo *vo_sf_cree_com*; y, por otro lado, los verbos de las clases C3 (evaluación), C4 (aceptación) y C51 (sensación), bajo el tipo *vo_sf_no_neu* (verbos no neutros).

La propiedad sintáctica que permite esta distinción es la aceptación de voz pasiva (propiedad c): las clases C1 y C2 no la aceptan, mientras que las clases restantes sí.

Por último, definimos un nuevo nivel de especialización, dentro del tipo *vo_sf_cree_com* para marcar las diferencias entre las clases C11 y C12, que en la tabla llamamos de discurso reproducido y de discurso referido. Los primeros, junto con los de creencia (C2), quedan bajo el tipo *vo_sf_cree_drep*. Los de discurso referido quedan solos bajo el tipo *vo_sf_dref*. Esta subdivisión se basa en las propiedades siguientes:

- Forma sintáctica del mensaje (propiedad b): Los verbos introductores de discurso reproducido (C11), junto con los de creencia (C2), aceptan una cita indirecta (mensaje como subordinada con *que*, como *X dijo/cree que ...*). Los verbos de discurso referido (C12) no la aceptan (* *X habló que ...*).
- Topicalización del asunto (propiedad d): Los verbos de C11 y C2 aceptan esta propiedad (*En cuanto a este tema, X dijo/cree que ...*). Los verbos de la clase C12 no la aceptan (* *En cuanto a este tema, X habló.*).
- Inversión del sujeto en un inciso de discurso (propiedad e): Es posible para los verbos de las clases C11 y C2 ("*.....*", *dijo/cree X*) pero no para los de la clase C12 (* "*.....*", *habló X*).

Propiedades semánticas de los verbos de opinión

Dentro del marco de este trabajo decidimos modelar una única propiedad semántica, la orientación semántica. Esta propiedad aporta información que ayuda a establecer si una opinión es positiva, negativa o neutra respecto al tema sobre el que se opina.

Las propiedades sintáctico-semánticas analizadas en el apartado anterior nos permitieron crear una jerarquía de verbos de opinión que dio lugar a cuatro tipos diferentes: verbos de opinión de creencia y discurso reproducido (*vo_sf_cree_drep*), verbos de opinión de discurso referido (*vo_sf_dref*), verbos de opinión no neutros con sujeto fuente (*vo_sf_no_neu*) y verbos de opinión con sujeto asunto (*vo_sa*).

Los primeros dos tipos tienen orientación semántica neutra. Los tipos *vo_sf_no_neu* y *vo_sa*, en cambio, tienen orientación semántica positiva o negativa. Para cada uno de ellos definimos, por lo tanto, dos subclases, dadas por el valor de orientación semántica: *vo_sf_pos*, *vo_sf_neg*, *vo_sa_pos* y *vo_sa_neg*.

La jerarquía de verbos de opinión

Una vez establecida esta organización de los verbos de opinión según sus propiedades sintáctico-semánticas, la jerarquía de verbos queda como lo muestra la figura 5.9.

Por razones de espacio, las restricciones para cada tipo de verbo no se incluyen en la jerarquía sino que se presentan aparte, con una notación extendida, tomada de Copestake (2002) y (Sag et al, 2003), que se explica a continuación:

- # → coindexación
- < > → notación simplificada para listas
- a | b → alternativa
- (a) → elemento opcional

Además, para algunos rasgos cuyo tipo es una lista, se utiliza la notación $\langle \dots, elem \rangle$, tomada de Pollard y Sag (1987), para indicar que se agrega el elemento 'elem' a una lista que puede contener otros elementos definidos para ese mismo rasgo en tipos más generales.

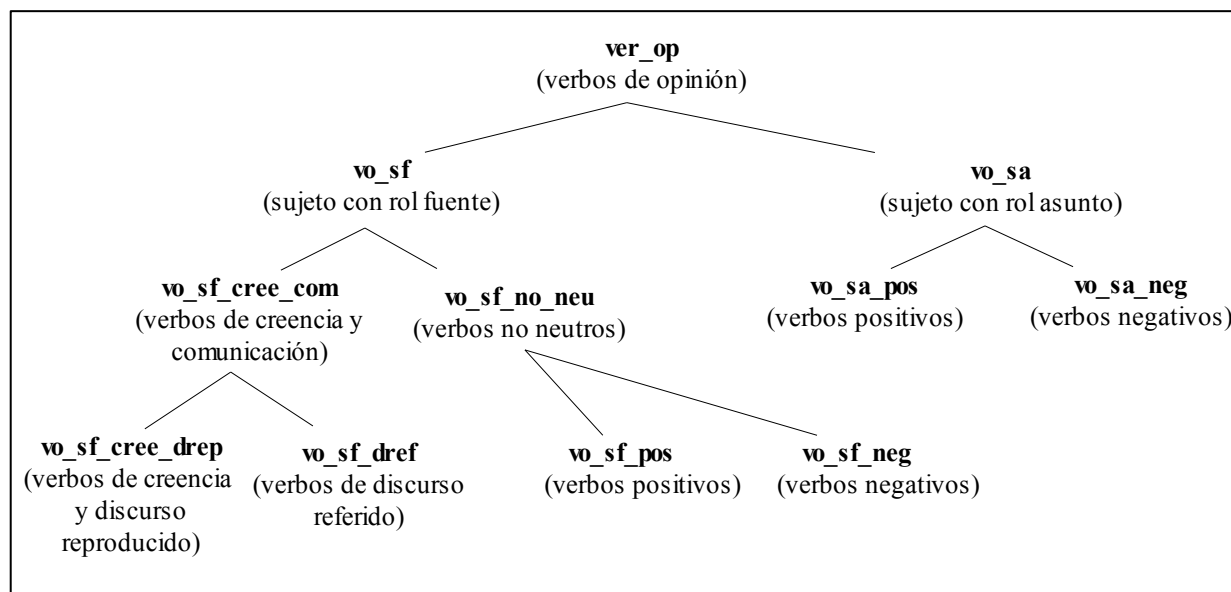


Figura 5.9: Jerarquía de tipos de verbos de opinión

Restricciones para el tipo *vo_sf*

(verbos de opinión con sujeto fuente)

Clases: C11 (*decir*), C12 (*hablar*), C2 (*creer*), C3 (*criticar*), C4 (*aceptar*) y C51 (*odiar*)

Subtipo de *ver_op*.

Solamente se indica que hay un argumento que es un grupo nominal con función sujeto cuyo rol semántico respecto al verbo es fuente.

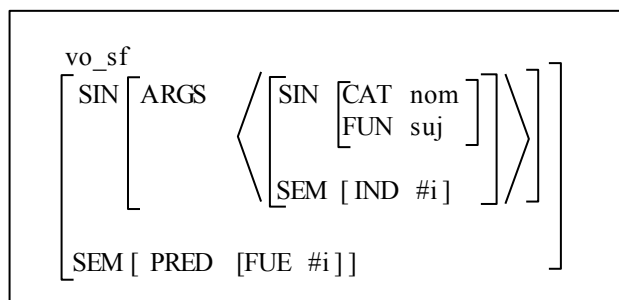


Figura 5.10: Tipo *vo_sf* (sujeto con rol fuente)

En español el sujeto puede estar omitido, por simplicidad asumimos que cualquier sujeto omitido puede ser materializado mediante el pronombre personal que concuerda con la conjugación verbal.

Restricciones para el tipo *vo_sf_cree_com*

(verbos de opinión con sujeto fuente, clases creencia y comunicación)

Clases: C11 (*decir*), C12 (*hablar*), C2 (*creer*)

Hereda las restricciones del tipo *vo_sf*.

Además se indica que este tipo no acepta la voz pasiva y que hay un nuevo argumento que es

un grupo nominal o preposicional con función complemento directo o preposicional cuyo rol semántico respecto al verbo es asunto.
Por último, se asigna orientación semántica neutra.

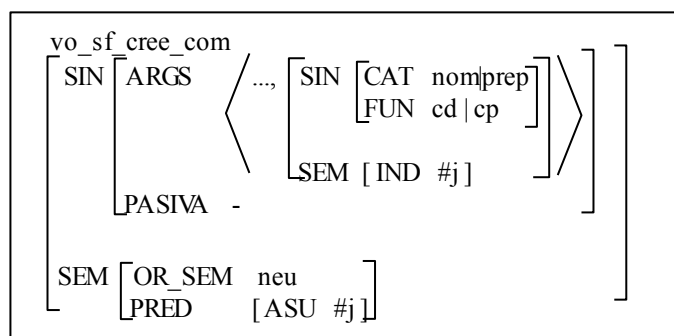


Figura 5.11: Tipo *vo_sf_cree_com* (sujeto con rol fuente, verbos de creencia y de comunicación)

Restricciones para el tipo *vo_sf_cree_drep*

(verbos de opinión con sujeto fuente, clases creencia y discurso reproducido)

Clases: C11 (*decir*), C2 (*creer*).

Hereda las restricciones de *vo_sf_cree_com*.

Se indica que aceptan tanto cita directa como cita indirecta con rol mensaje; topicalización del asunto e inversión del sujeto.

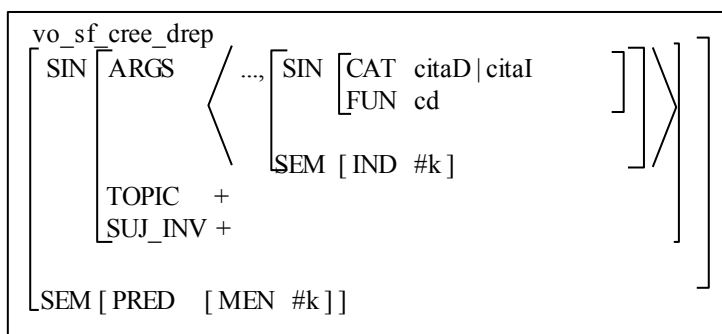


Figura 5.12: Tipo *vo_sf_cree_drep* (sujeto con rol fuente, verbo de creencia o discurso reproducido)

Una restricción que no es posible modelar en el esquema que estamos definiendo es el hecho de que cuando hay topicalización del asunto, no puede haber inciso de discurso con inversión de sujeto (* *Sobre el tema, hay que discutir más, dijo el doctor*).

Restricciones para el tipo *vo_sf_dref*

(verbos de opinión con sujeto fuente, clase discurso referido)

Clases: C12 (*hablar*).

Hereda las restricciones de *vo_sf_cree_com*.

Se indica que solo aceptan cita directa con rol mensaje y no aceptan topicalización del asunto ni inversión del sujeto.

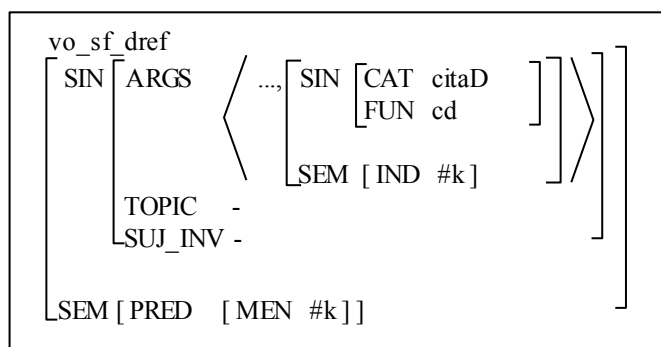


Figura 5.13: Tipo vo_sf_dref (sujeto con rol fuente, verbo de discurso referido)

Restricciones para el tipo vo_sf_no_neu

(verbos de opinión con sujeto fuente y orientación semántica no neutra)

Clases: C3 (*criticar*), C4 (*aceptar*) y C51 (*odiar*).

Hereda las restricciones de ver_op_sf.

Tienen un argumento, que puede ser un grupo nominal, un grupo preposicional o una proposición, que tiene rol asunto respecto al verbo. También un argumento con forma de cita directa que es mensaje.

Aceptan pasiva (Y fue rechazado por X).

No es claro que acepten topicalización de asunto, mientras no se observe esta estructura en los corpus se considera que no corresponde.

La orientación semántica es positiva o negativa, se especifica en sus subclases.

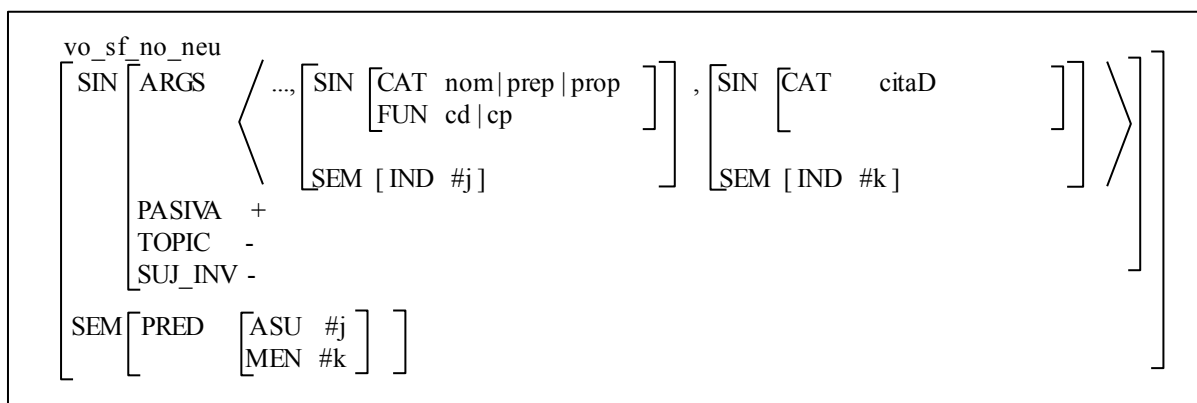


Figura 5.14: Tipo vo_sf_no_neu (sujeto con rol fuente, orientación semántica no neutra)

Restricciones para el tipo vo_sf_pos

(verbos de opinión con sujeto fuente y orientación semántica positiva)

Clases: C3 (*apoyar*), C4 (*aceptar*) y C51 (*querer*).

Hereda las restricciones de vo_sf_no_neu.

La orientación semántica es positiva.

Restricciones para el tipo *vo_sf_neg*

(verbos de opinión con sujeto fuente y orientación semántica negativa)

Clases: C3 (*criticar*), C4 (*rechazar*) y C51 (*odiar*).

Hereda las restricciones de *vo_sf_no_neu*.

La orientación semántica es negativa.



Figura 5.15: Tipos *vo_sf_pos* (sujeto con rol fuente, orientación semántica positiva) y *vo_sf_neg* (sujeto con rol fuente, orientación semántica negativa)

Restricciones para el tipo *vo_sa*

(verbos de opinión con sujeto asunto)

Clases: C52 (*gustar*, *molestar*).

Hereda las restricciones de *ver_op*.

Uno de los argumentos es un grupo nominal con función sujeto y rol asunto.

El sujeto puede estar invertido, aunque no se trata exactamente de la misma estructura que para los verbos de discurso reproducido o creencia, utilizamos el mismo rasgo para indicar esta propiedad.

No aceptan pasiva ni topicalización de asunto.

La orientación semántica es positiva o negativa, se especifica en sus subclases.

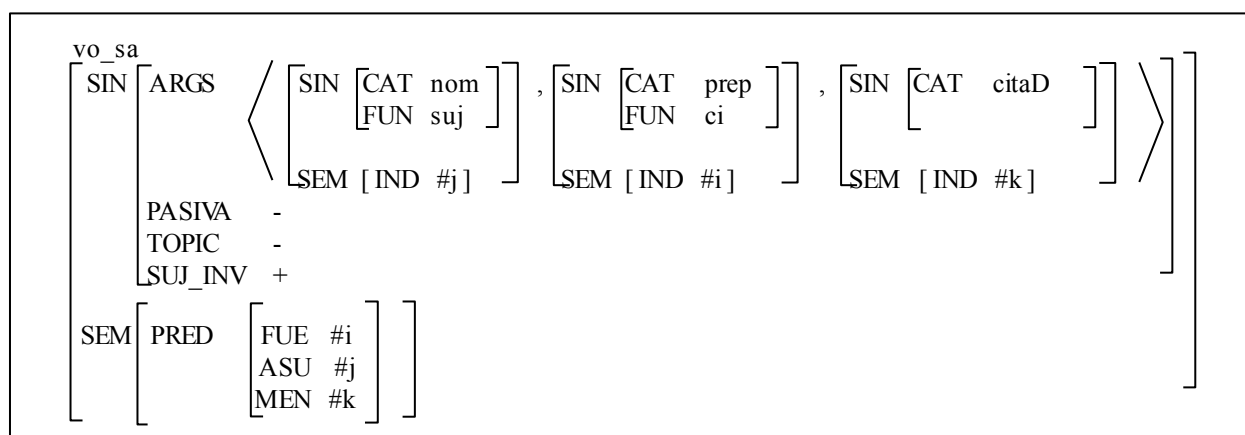


Figura 5.16: Tipo *vo_sa* (sujeto con rol asunto)

Restricciones para el tipo *vo_sa_pos*

(verbos de opinión con sujeto asunto y orientación semántica positiva)

Clases: C52 (*gustar*).

Hereda las restricciones de *vo_sa*.

La orientación semántica es positiva.

Restricciones para el tipo *vo_sa_neg*

(verbos de opinión con sujeto asunto y orientación semántica negativa)

Clases: C52 (molestar).

Hereda las restricciones de *vo_sa*.

La orientación semántica es negativa.

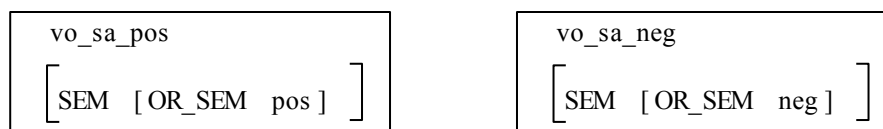


Figura 5.17: Tipos *vo_sa_pos* (sujeto con rol asunto, orientación semántica positiva) y *vo_sa_neg* (sujeto con rol asunto, orientación semántica negativa)

5.2.2.2 Los nombres de opinión

La gran mayoría de nuestros nombres de opinión son deverbales. Este tipo de nombres heredan de sus verbos de origen su estructura argumental, como vimos en la sección 3.2.2 del capítulo 3. Es posible modelar los nombres deverbales como el producto de la aplicación de reglas léxicas a los verbos, especificando los procesos morfológicos que permiten obtener las formas nominales y los mecanismos mediante los cuales se produce la herencia de argumentos.

Sin embargo, para nuestro trabajo decidimos no seguir este enfoque, principalmente por la constatación de que, si bien los verbos de las diferentes clases que consideramos de opinión muestran comportamientos diferentes en cuanto a sus propiedades sintáctico-semánticas, los nombres que de ellos se derivan muestran un comportamiento mucho más homogéneo. Los argumentos de opinión se manifiestan de manera muy similar para todos los nombres deverbales de opinión.

Por otra parte, en cuanto a la morfología, los mecanismos de derivación son muy variados y transversales a las clases semánticas definidas. Por ejemplo, se agrega el sufijo *-ión* en derivados de verbos pertenecientes a diferentes clases semánticas, como *expresar/expresión*, *declarar/declaración*, *aprobar/aprobación*, *aceptar/aceptación*, *preocupar/preocupación*.

También tuvimos en consideración una razón práctica, relacionada con el objetivo final de nuestro trabajo, que es la implementación de un sistema automático. En ese sentido, no constituye un aporte interesante el análisis de estas unidades como derivadas de los verbos ya que el *tagger* con el que trabajamos los considera unidades independientes. Para vincular un nombre con el verbo del cual deriva deberíamos generar un proceso especial para obtener esa relación.

Como primera aproximación al modelado de los nombres de opinión proponemos un solo tipo, *nom_op*, con ciertos rasgos para la representación de los argumentos de opinión, y algunos subtipos definidos en base a los valores posibles de orientación semántica. Un estudio más detallado de los nombres de opinión podrá dar lugar a una jerarquía de nombres más compleja y la consecuente reducción de redundancia en las entradas léxicas.

El tipo *nom_op* hereda los rasgos SIN y SEM de tipos más generales. Al rasgo SIN le corresponde el tipo *t_sin_nom_op* y al rasgo SEM el tipo *tipo_sem_pr*.

En el rasgo ARGS especificamos las realizaciones posibles de los argumentos de opinión y, a través de la coindexación entre estos argumentos y los rasgos semánticos, establecemos los roles semánticos de los argumentos. La TFS que representa esto es la siguiente:

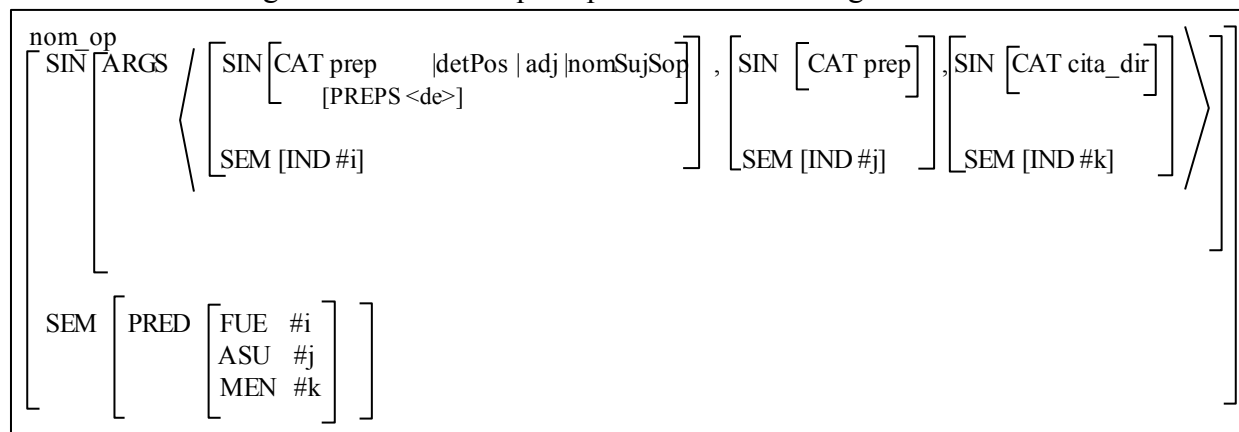


Figura 5.18: Tipo *nom_op* (predicados nominales)

Como se muestra en la figura 5.18, la fuente (argumento con índice 'i') puede tener diferentes realizaciones sintácticas:

- complemento con de (*la opinión de X*)
- determinante posesivo (*su/mi/nuestra opinión*) (indicado por la categoría *detPos*)
- adjetivo relacional (*la opinión internacional*)
- sujeto de un verbo soporte (*X realizó una declaración*) (categoría *nomSujSop*)

Los primeros tres puntos fueron mencionados en la sección 3.2.2 de este documento. El último punto surge del análisis de corpus, en donde encontramos con frecuencia ejemplos como (5.2) con un verbo soporte cuyo sujeto es la fuente del nombre de opinión.

(5.2) Sólo una opinión arriesga Trindade: "Por el tiempo y el esfuerzo que demandó el trabajo, considero que se trata de elementos ideológicos, no diseñados por mero placer".

A partir de esta clase general para los nombres de opinión, definimos dos subclases, *nom_op_neu* y *nom_op_no_neu*, para nombres neutros (*opinión, afirmación*) y no neutros (*rechazo, apoyo*), respectivamente. Para la clase *nom_op_no_neu*, definimos dos subclases más: *nom_op_pos* (nombres positivos, como *apoyo*) y *nom_op_neg* (nombres negativos como *rechazo*). Estas clases se determinan asignando los valores correspondientes al rasgo OR_SEM dentro de los rasgos semánticos (SEM).

Uno de los datos que más interesa sobre los argumentos de los nombres de opinión es qué preposición es posible para el asunto. Esta información vendrá especificada en las entradas léxicas, como sucede con los verbos.

Cabe mencionar una situación particular que se da respecto a la preposición *de*. Normalmente, esta preposición introduce la fuente de la opinión, pero para muchos nombres es posible utilizarla también como introductora de asunto. En los ejemplos (6.3) y (6.4) los nombres

anuncio y *rechazo* admiten el asunto introducido por *de*, en cambio en (6.5) y (6.6) los nombres *comunicado* y *apoyo* solo admiten *de* como introductor de fuente.

- (6.3) a. Como prueba de esto anotan que el anuncio de la Fiscalía sobre la investigación fue dado a conocer a la opinión pública ...
b. ... criticó al gobierno de George Bush por el anuncio de levantar las sanciones económicas que se había impuesto al régimen de Pretoria.
- (6.4) a. Hay un rechazo de las bases a sus dirigentes, rechazo normal y natural ...
b. Así que, frente a esa desmesura, que puede provocar cierto rechazo de su obra, yo creo que hoy tenemos que hacer una lectura entre líneas ...
- (6.5) Aunque el comunicado de la Cancillería dominicana trata de quitar significación a la inasistencia de Aristide al encuentro ...
- (6.6) Carlos Álvarez cuenta con el apoyo de Ricardo Muti y el equipo de la Scala para debutar en el primer coliseo ...

Como lo muestran los ejemplos, esta característica no depende de la clase semántica de los verbos y sus nombres derivados.

En cuanto al mensaje, lo incluimos como un argumento más, a pesar de que, al igual que en el caso de los verbos, no se trata estrictamente de un argumento. En nuestro enfoque, es un elemento más de la opinión, al mismo nivel que la fuente y el asunto. Encontramos varios casos de citas directas introducidas por nombres de opinión de diferentes clases semánticas, como *declaración* (6.7) y *crítica* (6.8).

- (6.7) Esa carencia de humanismo conduce a Mann a una declaración trágica: “Después de aquel viaje, me sentí un extraño en mi patria”.
- (6.8) Marías practica la crítica al incipiente turismo: “Corren camellos y asnos, entre gritos, con una falsa alegría turística, ...”

5.2.2.3 Los predicados de opinión preposicionales

En el caso de los predicados de opinión preposicionales, consideramos que solamente existen dos posibles argumentos de opinión: la fuente y el mensaje. El mensaje corresponde a la oración principal y la fuente es el grupo introducido por la preposición o locución prepositiva de opinión.

Como observamos en la sección 3.3 del capítulo 3, el segmento introducido por el predicado de opinión preposicional puede aparecer al principio de la oración, intercalado (en forma de inciso) o al final de la oración. Como nuestro modelo no da cuenta del orden en el cual aparecen los argumentos en la oración, representamos de un mismo modo cualquiera de los casos mencionados.

Los predicados de opinión preposicionales que pudimos relevar a partir del estudio de los artículos citados en la sección 3.3 y del análisis de nuestro corpus de desarrollo son los siguientes: *según*, *para*, *de acuerdo con*, *de acuerdo a* y *a juicio de*. Como se ve, los cuatro tienen orientación semántica neutra.

Todos los aspectos mencionados quedan representados mediante la TFS de la figura 5.19.

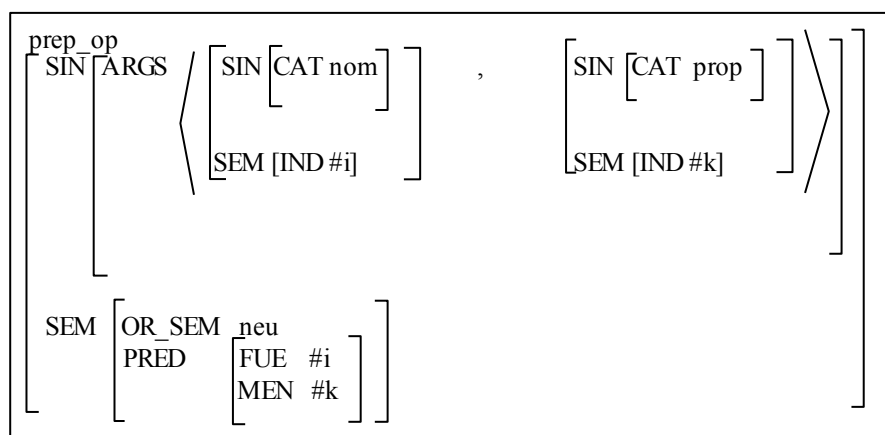


Figura 5.19: Tipo *prep_op* (predicados preposicionales)

Como ya lo hemos visto para los verbos y los nombres, en este caso, el tipo *prep_op* hereda los rasgos SIN y SEM de tipos más generales. Al rasgo SIN le corresponde el tipo *t_sin_prep_op* y al rasgo SEM el tipo *tipo_sem_pr*. En el rasgo ARGS especificamos las realizaciones posibles de los argumentos de opinión correspondientes a predicados preposicionales.

5.3 Utilización del modelo en el desarrollo de los sistemas informáticos

El modelo para los predicados de opinión que especificamos mediante las TFS se completa mediante la definición de entradas léxicas, cada una de las cuales debe llevar asociado el tipo que le corresponde en la jerarquía de tipos. En el caso de los verbos, esos tipos, como vimos, determinan diferentes clases de predicados.

Por ejemplo:

- al verbo *decir* le corresponde el tipo o la clase *vo_sf_cree_drep* (es un verbo con sujeto fuente, además es un verbo de creencia o de discurso reproducido).
- al verbo *referirse* le corresponde el tipo o la clase *vo_sf_dref* (es un verbo con sujeto fuente, además introduce discurso referido).
- al verbo *apoyar* le corresponde el tipo o la clase *vo_sf_pos* (es un verbo con sujeto fuente y además es un verbo de una de las clases subjetivas, con orientación semántica positiva).

Los sistemas informáticos que desarrollamos necesitan, como insumo fundamental, un repertorio de predicados de opinión que permitan detectar la presencia de opiniones en los textos. Además, cada sistema, en diferente medida, utiliza parte de la información que nuestro modelo teórico permite asociar a cada predicado.

Principalmente, el sistema basado en reglas utiliza gran parte de la información sintáctico-semántica que se representa en el modelo. Dado un verbo de opinión determinado, según la clase a la cual este pertenece, se pueden aplicar diferentes reglas. Por ejemplo, en algunos casos se aplica una regla que marca el sujeto como fuente (para verbos cuyo tipo en la jerarquía es algún subtipo de *vo_sf*), en otros casos la fuente es un complemento (verbos cuyo tipo es algún subtipo de *vo_af*). También puede aplicarse una regla que incorpora una subordinada (con *que*) como el

mensaje, si el verbo tiene el tipo *vo_sf_cree_drep* (verbo de creencia o de discurso reproducido). En otros casos, puede aplicarse una regla que reconoce como asunto un grupo nominal (el objeto directo del verbo), si el verbo tiene asociado algún subtipo de *vo_sf_no_neu* (verbos no neutros).

Como parte de la implementación, creamos un repertorio de predicados de opinión en donde especificamos para cada entrada la clase (el tipo) que le corresponde en el modelo. Describimos dicho repertorio en el capítulo 6.

El tipo que debe asociarse a cada entrada puede determinarse en base a la clase que le corresponde a cada verbo en ADESSE, dado que de ahí tomamos las clases iniciales que dieron lugar al modelo. Sin embargo, como los verbos a ser incluidos en el repertorio pertenecen a los corpus de desarrollo, su tipo fue determinado a partir del comportamiento observado en sus contextos de ocurrencia. Esto se debe a la posibilidad de ambigüedad y, además, a que la mayoría de los verbos que estudiamos son utilizados de forma bastante homogénea en los textos periodísticos (recordamos que los ejemplos de ADESSE pertenecen en muchos casos a textos literarios).

5.4 Síntesis

Definimos un modelo basado en estructuras de rasgos tipificadas (TFS) en el cual mostramos cómo se organizan los predicados de opinión según diferentes propiedades sintácticas y semánticas. El modelo incluye información sobre la forma en que se realizan los argumentos que coinciden con los elementos que definimos para la opinión (fuente, asunto y mensaje).

Para los predicados verbales definimos diferentes tipos (o clases) que se organizan en forma jerárquica, heredando cada sub-tipo las propiedades (o restricciones) de los tipos más generales. Algunas de las propiedades sintáctico-semánticas que incluimos son: rol semántico del sujeto del verbo de opinión (fuente o asunto), aceptación de una subordinada que contiene el mensaje de la opinión, aceptación del asunto como objeto directo, posibilidad de voz pasiva, posibilidad de topicalización del asunto.

Para los predicados nominales y preposicionales la organización es mucho más simple que para los verbales, se definieron solamente algunos pocos tipos.

Para todos los predicados se definió un rasgo puramente semántico: la orientación semántica. En el caso de los verbos y los nombres, hay tipos con orientación neutra (*decir, creer*) y tipos con orientación no neutra que se dividen en positivos (*gustar, felicitar, apoyo*) y negativos (*rechazar, molestar, crítica*). Los predicados preposicionales son todos neutros (*según, de acuerdo con, a juicio de*).

La jerarquía de tipos y las restricciones definidas para cada tipo constituyen información importante que debe ser tomada en cuenta para el reconocimiento automático, por lo que es incluida en el repertorio de predicados (ver capítulo 6) que utilizan los sistemas desarrollados.

CUARTA PARTE: Sistemas informáticos

6. Repertorio de predicados de opinión

Para el reconocimiento de las opiniones es fundamental consultar un diccionario o repertorio de las expresiones que transmiten opiniones en los textos. Como vimos en el estudio de trabajos relacionados, se ha trabajado mucho en la construcción de diccionarios subjetivos, especialmente para el inglés, pero también en alguna medida para el español.

En este trabajo, debido a nuestro enfoque centrado en los predicados de opinión verbales y nominales, necesitamos un repertorio de este tipo de predicados.

Decidimos construir un recurso propio para una primera etapa de trabajo, de modo de evaluar el grado de cobertura que este repertorio permite alcanzar. Para esto, consideramos los verbos y nombres usados frecuentemente en contextos de opinión en nuestros corpus de análisis, obteniendo un total de 86 verbos y 42 nombres de opinión.

La evaluación del sistema de reglas (que presentamos en el capítulo 8) mostró que, para el tipo de textos con los que estamos trabajando, estas cantidades son apropiadas, dado que se obtuvo un alto porcentaje de recuperación de los predicados de opinión (91%). En un total de 281 ocurrencias de predicados de opinión en el corpus de evaluación (de 13.000 palabras), solamente 25 no fueron reconocidas.

Posteriormente a la evaluación mencionada, se extendió el repertorio de modo que cuenta actualmente con 100 verbos y 55 nombres de opinión.

La cantidad de elementos incluidos en el repertorio es comparable a las cantidades utilizadas en algunos trabajos cercanos al nuestro, todos para el inglés:

- Kim y Hovy (2006) utilizan una lista de 151 verbos positivos y 69 verbos negativos, indicadores de opinión. No trabajan con verbos de comunicación, que son neutros, ya que su trabajo se orienta exclusivamente al reconocimiento de contenido subjetivo.
- En el trabajo de (Krestel et al, 2008), se utiliza una lista de 54 verbos de discurso reproducido. En este caso, no se incluyen verbos de clases valoración, aceptación, sensación, ya que el objetivo es el reconocimiento exclusivamente del discurso reproducido.
- Saurí (2008) en su tesis sobre la factividad de los eventos, crea un repertorio de verbos, nombres y adjetivos indicadores de fuentes (*Source Introducing Predicates*, SIP). Saurí necesita esta información ya que considera que la atribución de segmentos a diferentes fuentes es un factor importante en el cálculo de la factividad de los eventos. El repertorio de SIP incluye, entre muchos otros, los verbos: *say, express, consider, think, accept, accuse, hate, worry*. Como se ve, se abarcan verbos pertenecientes a clases similares a las que consideramos en nuestro trabajo (comunicación, creencia, valoración, aceptación, sensación). Las piezas léxicas tienen asociada una clase semántica (*say, announce, conjecture, know, disclose, etc.*) e información sintáctica, como función sintáctica de los argumentos y algunas preposiciones aceptadas. El repertorio cuenta con unos 200 verbos, 60 nombres y 30 adjetivos. Es el recurso que más se asemeja a nuestro repertorio de predicados de opinión, salvo por el hecho de que nosotros no incluimos adjetivos. Es

importante señalar que este recurso no fue aplicado, hasta el momento de la publicación de la tesis de Saurí, al reconocimiento automático de las fuentes.

6.1 Representación interna de las piezas léxicas

En base al modelo especificado mediante estructuras de rasgos tipificadas (ver capítulo 5) se asoció cierta información a cada pieza léxica del repertorio de predicados de opinión.

Cada pieza es representada como una lista en donde el primer elemento es un lema de un verbo o nombre de opinión y los demás elementos representan el tipo al que pertenece cada lema en el modelo y las propiedades que deben ser especificadas a nivel léxico, por ejemplo, preposición aceptada por cada verbo.

Por razones prácticas, relacionadas con la implementación del sistema de reglas contextuales, incluimos para cada pieza, además del tipo específico al cual pertenece, todos los tipos que lo subsumen. Esta decisión implica cierta redundancia en la información asociada a las piezas léxicas, pero simplifica significativamente la operativa de las reglas.

Esto sucede porque los tipos intermedios de la jerarquía nos permiten referirnos a un conjunto más amplio de predicados que comparten una cierta propiedad que los tipos más específicos. Por ejemplo, si se va a aplicar una regla que marca como fuente un grupo nominal que precede un predicado de opinión en forma finita, es suficiente verificar que este pertenezca al tipo *vo_sf* (verbo de opinión con sujeto fuente). Sería más costoso, desde el punto de vista de la ejecución del motor de reglas, verificar si pertenece a alguno de los tipos subsumidos por *vo_sf*, que son *vo_sf_cree_drep* (verbos con sujeto fuente de creencia y discurso reproducido), *vo_sf_dref* (verbos con sujeto fuente de discurso referido), *vo_sf_pos* (verbos con sujeto fuente positivos), *vo_sf_neg* (verbos con sujeto fuente negativos).

En la tabla 6.1 mostramos algunos ejemplos de verbos y en la tabla 6.2 ejemplos de nombres de nuestro repertorio.

En cuanto a la información léxica, tanto para los verbos como para los nombres se registra una lista (representada por corchetes) de preposiciones aceptadas como introductoras de asunto, que puede ser una lista vacía. Si se incluye la preposición *sobre*, se asume que también se aceptan otras locuciones típicas introductoras de asunto como *en lo que respecta a*, *en cuanto a*, etc.

Para los verbos, en particular, se indica con el símbolo Φ la aceptación de asunto sin preposición (es decir, un grupo nominal con función complemento directo) y por medio de la palabra 'prop' se indica la aceptación de una proposición como asunto.

Para los nombres se indica, además de la lista de preposiciones que introducen el asunto, el rol semántico que debe asignarse al complemento con *de*. En algunos pocos casos este complemento corresponde siempre al asunto (*la difusión de material prohibido*). En la mayoría de los casos o bien se trata siempre de la fuente (*la afirmación del senador*), o bien el rol es ambiguo (*el planteo del senador / el planteo de la formación de una comisión*).

El repertorio completo de nombres y verbos de opinión se incluye en el Anexo B.

verbo de opinión	clase dentro de la jerarquía	forma del asunto
decir	vo_sf_cree_drep, vo_sf_cree_com, vo_sf	[Φ , de, sobre]
advertir	vo_sf_cree_drep, vo_sf_cree_com, vo_sf	[sobre]
hablar	vo_sf_dref, vo_sf_cree_com, vo_sf	[de, sobre]
referirse	vo_sf_dref, vo_sf_cree_com, vo_sf	[a]
apoyar	vo_sf_pos, vo_sf_no_neu, vo_sf	[Φ , a]
coincidir	vo_sf_pos, vo_sf_no_neu, vo_sf	[con, en]
rechazar	vo_sf_neg, vo_sf_no_neu, vo_sf	[Φ , a]
gustar	vo_sa_pos, vo_sa	[] ¹⁹
molestar	vo_sa_neg, vo_sa	[]

Tabla 6.1: Algunos verbos de opinión incluidos en nuestro repertorio

nombre de opinión	clase dentro de la jerarquía	forma del asunto	rol del complemento con <i>de</i>
afirmación	nom_op_neu, nom_op	[sobre]	fuelle
difusión	nom_op_neu, nom_op	[]	asunto
planteo	nom_op_neu, nom_op	[sobre]	ambiguo
interés	nom_op_pos, nom_op_no_neu, nom_op	[hacia, por, en]	fuelle
aceptación	nom_op_pos, nom_op_no_neu, nom_op	[a]	ambiguo
agravio	nom_op_neg, nom_op_no_neu, nom_op	[hacia, a]	fuelle
crítica	nom_op_neg, nom_op_no_neu, nom_op	[habia, a]	fuelle

Tabla 6.2: Algunos nombres de opinión incluidos en nuestro repertorio

6.2 Síntesis

Para el reconocimiento automático de las opiniones, nuestros sistemas informáticos necesitan identificar los predicados de opinión, ya que estos indican, como se dijo, que se está transmitiendo una opinión de alguna fuente. Para esto es necesario contar con un repertorio de predicados de opinión, recurso que construimos especialmente para nuestro trabajo.

El repertorio, en su versión actual, contiene 100 verbos y 55 nombres de opinión, en general extraídos de nuestro corpus de desarrollo. La evaluación del sistema de reglas (capítulo 8) muestra que con estas cantidades se logró una buena cobertura (91%) del corpus de evaluación. De todos modos, es necesario seguir incorporando elementos al repertorio.

Además de los lemas de los verbos y nombres de opinión, incluimos en el repertorio los tipos asociados a cada pieza léxica, según el modelo definido para los predicados de opinión. De este modo, los sistemas informáticos cuentan con información sintáctica y semántica para cada verbo

¹⁹ Para los verbos con sujeto asunto, esta lista pierde sentido, por lo que queda vacía.

o nombre particular, que les permite identificar los restantes elementos de la opinión (fuente, asunto y mensaje) dentro de las estructuras sintácticas en donde los predicados ocurren.

También se incluye para cada pieza del repertorio información sobre el tipo de preposición que esta acepta para introducir el asunto.

El repertorio completo puede verse en el Anexo B.

7. Sistema basado en reglas contextuales

El primer enfoque que aplicamos para el reconocimiento de opiniones en textos se basa en la escritura manual de reglas, deducidas a partir de un análisis de corpus.

El sistema toma como entrada un texto pre-procesado por un POS-tagger, Freeling (Atserias et al, 2006, WebFreeLing1) y posteriormente por el sistema Clatex (Wonsever et al, 2006) que segmenta las oraciones en proposiciones. Luego se aplican varios módulos de reglas para el reconocimiento de opiniones, obteniéndose como salida el texto con anotaciones en formato xml que muestran las opiniones encontradas y los elementos que las componen. Además, se muestran los valores de orientación semántica, mediante el atributo 'os'. La salida del sistema se ilustra en el siguiente ejemplo:

```
<opinion os="pos">
  <fuente os="neu">Mujica</fuente>
  <predicado os="pos">respaldó</predicado>
  <asunto os="neu">importante inversión minera</asunto>
</opinion>
```

Las reglas que integran el sistema se basan en el formalismo de reglas contextuales definido por Wonsever y Minel (2001), incluyendo algunas extensiones posteriores. Este tipo de regla permite la especificación de contextos, zonas de exclusión, operadores de opcionalidad, negación, eliminación de etiquetas ya existentes, entre otros. Además, cada regla ofrece la posibilidad de chequear condiciones diversas de los elementos que la componen, como por ejemplo, la pertenencia a una lista de palabras, funcionalidad imprescindible para este trabajo.

Las reglas diseñadas para este sistema se agrupan en módulos según el elemento que reconocen: predicado, fuente, asunto, mensaje. Hay, además, un módulo final que arma la opinión completa y algunos módulos de reglas accesorias, como el módulo que arma grupos nominales complejos del tipo [*El director del Hospital Maciel, Daniel Parada*] o los módulos que identifican y combinan elementos subjetivos y los operadores que actúan sobre ellos. El sistema contiene un total de 173 reglas.

Muchas de las reglas consultan diferentes recursos léxicos. El principal es el repertorio de predicados de opinión desarrollado para este trabajo (capítulo 6), pero también es necesario otro tipo de información: lista de indicadores de persona (*señor, doctor, candidato*), lista de indicadores de institución (*institución, hospital, diario*), lista de verbos soporte (*realizar, emitir*) y lista de preposiciones o locuciones prepositivas que introducen el asunto (*sobre, en cuanto a, respecto a, en lo que respecta a*).

Los recursos son consultados durante la aplicación de las reglas para verificar la pertenencia de algún elemento a alguna de las listas mencionadas y, si corresponde, tomar de allí la información asociada al elemento buscado. Esta consulta se realiza mediante la funcionalidad de invocación a predicados que extiende la potencialidad del formalismo de reglas contextuales.

Algunas de las propiedades que incluimos en nuestro modelo del capítulo 5, como aceptación de voz pasiva o posibilidad de topicalización, no son tenidas en cuenta por las reglas. Como el sistema se utiliza exclusivamente para reconocimiento y no para generación de texto, estas

formas particulares solo aparecen para los verbos que las aceptan, dado que asumimos textos bien escritos, razón por la cual no es imprescindible chequearlas.

En cambio, otras propiedades sí deben ser verificadas. Por ejemplo, la aceptación de mensaje proposicional (cita indirecta) debe chequearse necesariamente para evitar marcar como mensaje una subordinada con otro significado, como es el caso de *rechazaron que se clausurara el local*, en donde la proposición con *que* no constituye una cita indirecta sino el asunto, es decir, lo rechazado.

Al explicar cómo funcionan las reglas, veremos qué información proveniente del modelo teórico basado en TFS se utiliza en cada caso.

La operativa general del sistema se ilustra en la figura 7.1.

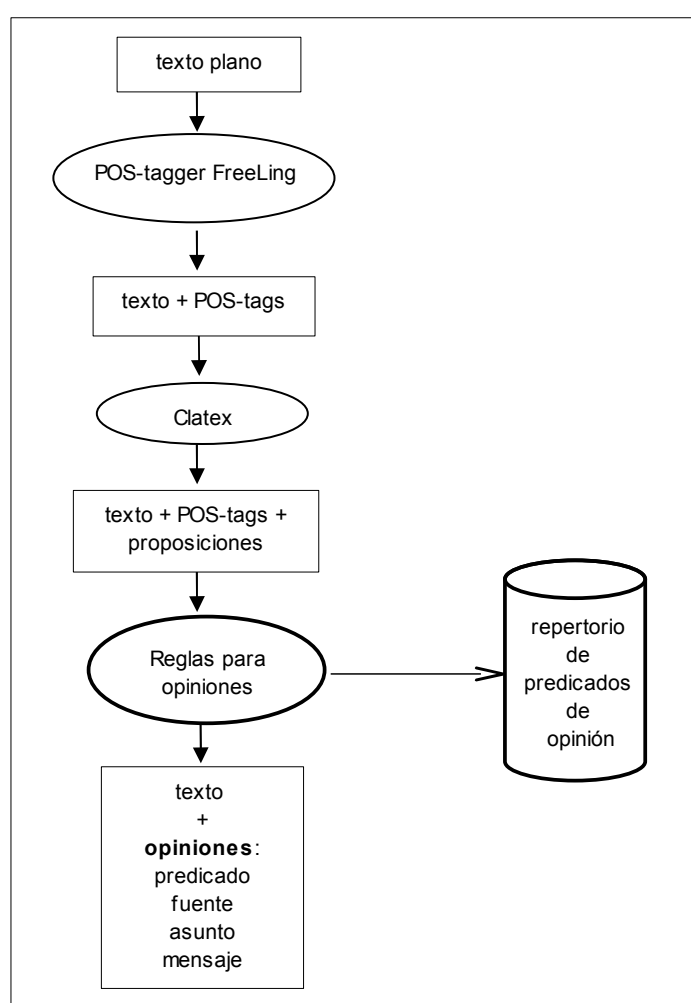


Figura 7.1: Operativa del sistema de reglas contextuales

En la figura se muestra en trazo oscuro el módulo de reglas y el repertorio de predicados de opinión, ambos creados en el marco de este trabajo.

Las reglas se aplican sobre los textos utilizando un intérprete de reglas contextuales en Prolog desarrollado por el grupo PLN. Este intérprete requiere que las reglas se escriban en un formato

interno poco amigable, por lo que utilizamos un generador de reglas, también desarrollado por el grupo PLN, que requiere cierta sintaxis que especificamos en la sección 7.1.

La versión disponible del generador de reglas al momento de comenzar este trabajo tenía algunas limitaciones importantes: solo permitía trabajar con etiquetas de tres atributos, no tenía incorporada la funcionalidad de llamadas a predicados Prolog para chequear condiciones, no aceptaba algunos de los operadores especiales como el que permite exigir que una etiqueta no esté contenida en otras.

Debido a lo anterior, en una primera etapa de implementación, extendimos las potencialidades del generador de reglas para poder trabajar con un número variable de atributos, utilizar todos los operadores soportados por el intérprete de reglas y, fundamentalmente, incluir chequeo de condiciones en las reglas, lo cual nos permite verificar si las palabras pertenecen a nuestro repertorio de predicados de opinión y también a algunas listas que definimos.

A continuación describimos la sintaxis de las reglas contextuales y luego los módulos de reglas para la identificación de la opinión y sus elementos.

El conjunto completo de reglas se incluye en el Anexo C.

7.1 El formalismo de reglas contextuales

Este formalismo permite la definición de reglas, agrupadas en módulos. Todas las reglas de un mismo módulo se aplican en una sola recorrida del texto. Las reglas se aplican sobre un texto que contiene ciertas etiquetas. Cada etiqueta es una lista de valores y abarca un segmento de texto. Cada vez que una regla es exitosa, se asigna una nueva etiqueta a un segmento de texto.

El texto de entrada, que está preprocesado por el *tagger* FreeLing, contiene una etiqueta por cada token del texto. El primer atributo de cada etiqueta es la categoría gramatical, el segundo atributo es la palabra y el tercero es el lema. Los restantes atributos dependen de la categoría.

Cabe aclarar que las etiquetas que asigna FreeLing siguen el formato EAGLES (Web EAGLES). Por ejemplo, para el nombre común masculino plural *árboles*, la etiqueta es NCMP00. Toda la información contenida en dicha etiqueta se mantiene en la etiqueta interna del sistema, en forma de lista, que para la palabra *árboles* sería: ['N','árboles','árbol','C','M','P','0','0']. La especificación de las etiquetas de FreeLing para cada categoría pueden consultarse en (FreeLing2).

Cada vez que se aplica una regla se asigna una nueva etiqueta sobre algún segmento de texto, el primer atributo de esta nueva etiqueta es una categoría, dentro de un conjunto de categorías definidas para cada aplicación particular de este sistema de reglas contextuales. En nuestro trabajo, por ejemplo, algunas de las categorías son predicadoOpinion, fuente, asunto, mensaje y numerosas categorías auxiliares más.

Sintaxis de las reglas

La sintaxis de las reglas es la siguiente:

```
nomRegla# etiqueta => contIzq >> cuerpo << contDer // conds .
```

Cada elemento significa lo siguiente:

- nomRegla: Nombre de la regla.
- etiqueta: Etiqueta a asignar al segmento de texto reconocido por la regla, es decir, el segmento que concuerda con el cuerpo de la regla. Esta etiqueta tiene la forma [cat,atr₁,atr₂, ...].
- contIzq: Contexto izquierdo; es una secuencia de etiquetas que se exigen a la izquierda del segmento que la regla reconoce.
- contDer: Contexto derecho; es una secuencia de elementos que se exigen a la derecha del segmento que la regla reconoce.
- cuerpo: Secuencia de etiquetas que abarcan el segmento de texto que va a ser etiquetado por la regla.
- conds: Condiciones que deben cumplirse para que la regla sea exitosa. Las condiciones pueden aplicarse sobre cualquier elemento de la regla, tanto elementos pertenecientes a los contextos como al cuerpo.

Los dos contextos pueden ser vacíos, el cuerpo de la regla siempre debe contener al menos una etiqueta.

Sobre las etiquetas de los contextos o del cuerpo pueden aplicarse algunos operadores que aumentan la expresividad de las reglas:

- op(et) La etiqueta 'et' es opcional.
- el(et) La aplicación de la regla, además de asignar la etiqueta que aparece del lado izquierdo de la regla (la que se indica antes del signo =>), tiene como efecto eliminar la etiqueta 'et'.

Sobre las etiquetas de los contextos se puede aplicar además un operador de negación:

- no(et) Se exige que no aparezca la etiqueta 'et'.

Para cada etiqueta se pueden exigir valores para cualquiera de sus atributos, por ejemplo, la expresión:

['c' , 'v2' , _ , _ , 'v5' , _ , _ , _]

indica que se pide una etiqueta con la categoría 'c', el valor 'v2' en el segundo atributo y el valor 'v5' en el quinto atributo. Los lugares ocupados por infra-guiones indican que se admite cualquier valor para esos atributos.

La expresión anterior es equivalente a la siguiente:

['c' , 'v2' , _ , _ , 'v5' | _]

en donde se indica que a partir del quinto atributo, los atributos restantes pueden tener cualquier valor.

En la tabla 7.1 mostramos diferentes expresiones para pedir diferentes tipos de elementos, en base a las etiquetas correspondientes al texto de entrada.

Notación	Tipo de palabra	Ejemplos
['S', 'de' _]	únicamente la palabra <i>de</i>	<i>de</i>
['S' _]	cualquier preposición	<i>de, a, del, al, contra, desde</i>
['V' _]	cualquier forma verbal	<i>ir, voy, cantar, vivido</i>
['V', _, 'caminar' _]	cualquier forma del verbo <i>caminar</i>	<i>camino, caminamos, caminando</i>
['V', _, _, _, 'N' _]	cualquier verbo en infinitivo	<i>cantar, comer, servir, limpiar</i>
['N', _, _, _, 'M' _]	cualquier nombre masculino	<i>árbol, árboles, temor, doctor</i>

Tabla 7.1 Ejemplos de especificación de elementos

Ejemplos de reglas sin condiciones

```
reg1# [ 'GV', noRealis, advNeg ] => >> [ no | _ ], [ 'V' | _ ] <<.
reg2# [ 'GV', noRealis, advImp ] => [ _, imposible | _ ] >> [ 'V', _, _, _, 'N' | _ ] <<.
```

En reg1 los dos contextos son vacíos y el cuerpo tiene dos elementos (separados por coma): la palabra *no* y un verbo (en cualquier forma). Esto quiere decir que la regla reg1 asigna la etiqueta *noRealis* con el valor *advNeg* para el primer atributo a todo segmento de texto compuesto por *no* y un verbo. Por ejemplo, para:

(7.1) Manuel no quiere ir a la escuela.

se agrega la etiqueta:

(7.2) Manuel <noRealis atr1="advNeg"> no quiere </noRealis> ir a la escuela.

En reg2 el contexto izquierdo debe contener la palabra *imposible* y el cuerpo debe ser un verbo en infinitivo. Por ejemplo, para:

(7.3) Julia dijo que es imposible entrar al estadio.

se agrega la etiqueta:

(7.4) Julia dijo que es imposible <noRealis atr1="advImp"> entrar </noRealis> al estadio.

Variables

Se pueden usar variables para tomar el valor de algún atributo de alguna etiqueta del lado derecho de la regla (contIzq, cuerpo o contDer) para ponerlo como valor del algún atributo de la etiqueta que la regla asigna (lado izquierdo de la regla). Cualquier palabra comenzada en letra mayúscula, sin comillas, es una variable. A continuación se muestra un ejemplo en donde *Modo* es una variable.

```
reg1# [ 'GV', noRealis, advNeg, Modo ] => >> [ no | _ ], [ _, _, 'V', _, Modo | _ ] <<.
```

La aplicación de esta regla al mismo ejemplo que se dio, da como resultado el texto anotado siguiente:

(7.5) Manuel <noRealis atr1="advImp" atr2="I"> no quiere </noRealis> ir a la escuela.

La letra "I" indica modo Indicativo, es el modo del verbo *quiere* que se transfiere a la expresión completa *no quiere*.

Condiciones

Se pueden imponer condiciones que deben cumplir algunos elementos de las etiquetas. Al final de la regla, luego del símbolo //, se indica la condición que debe cumplir algún elemento. Por ejemplo, en la regla siguiente se reconocen grupos nominales formados por un determinante, un nombre cuyo lema sea un título como *señor*, *doctor*, *presidente*, y un nombre propio. Para verificar la restricción impuesta al primer nombre, se aplica la condición 'título' aplicada al lema (representado por la variable Lema).

```
reg3# ['NP',tit] => >> [_ ,el|_],[_ ,Lema,'N'|_],[_ ,_, 'N', 'P'|_] << // titulo(Lema).
```

Los elementos que cumplen con la condición 'título' se especifican como hechos Prolog, del siguiente modo:

```
titulo('señor').
titulo('doctor').
titulo('presidente').
titulo('senador').
titulo('ingeniero').
...
```

Esto significa que la regla reg3 reconocerá todos los ejemplos siguientes:

[El señor Juan Pérez] será llamado ...

[La doctora Rodríguez] vendrá

Estamos con [la ingeniera Sara López]

Notar que FreeLing reconoce nombres propios compuestos por varias palabras, como *Juan Pérez*, y les asigna la etiqueta NP00000. De modo que, inicialmente, el segmento de texto *Juan Pérez* lleva la etiqueta ['Juan_Pérez','Juan_Pérez','N','P','0','0','0','0','0'].

Zonas de exclusión

Otra potencialidad de este formalismo de reglas consiste en la posibilidad de establecer zonas de exclusión, tanto en los contextos como en el cuerpo de la regla. Una zona de exclusión definida mediante la especificación de un conjunto de etiquetas C y un número N, indica que se permite una secuencia de hasta N palabras dentro de la cual no ocurra ninguna de las etiquetas de C.

La regla reg4 muestra un ejemplo de zona de exclusión, indicada por el *:

```
reg4# [NP,tit] => >>[_ ,el|_],*(C,3),[_ ,Lema,'N'|_],[_ ,_, 'N', 'P'|_] << // titulo(Lema).
C={['F'|_],['N'|_],['C'|_]}
```

La categoría 'F' de FreeLing significa signo de puntuación, 'N' nombre y 'C' conjunción.

Algunas de las expresiones reconocidas por reg4 son:

[La Ministra Rodríguez]

[La flamante Ministra Rodríguez]

[La recientemente designada ministra Sandra Rodríguez]

7.2 Los módulos de reglas

7.2.1 Reglas para el reconocimiento del predicado

Para identificar los predicados de opinión se realizaron reglas que para todos los verbos y nombres del texto chequean, mediante la definición de condiciones como las detalladas en la sección 7.1, si el lema correspondiente pertenece al repertorio de predicados de opinión.

En caso afirmativo, se asigna la etiqueta 'predOpinion' con algunos atributos. Un atributo indica si se trata de un verbo o un nombre, el resto de los atributos guardan la información asociada a cada entrada en el repertorio.

Para algunos pocos verbos se diseñaron reglas especiales de desambiguación dado que se usan frecuentemente con significados no relacionados con la expresión de opiniones. La tabla 7.2 muestra las reglas de desambiguación de verbos de opinión.

Verbo	Regla	Ejemplo
<i>contar</i>	Se exige que luego del verbo no venga la preposición <i>con</i> .	<i>el salón cuenta con un PC</i> <i>Juan cuenta que vino ... / cuenta sus experiencias</i>
<i>continuar</i>	Se exige que luego del verbo venga un grupo nominal o un signo de puntuación.	<i>continúan arreglando</i> <i>..., continúa el doctor.</i>
<i>pronunciarse</i>	Solo es predicado de opinión en su uso pronominal.	<i>pronuncia muy bien el inglés</i> <i>se pronunció sobre el tema</i>
<i>confirmar</i>	Solo es predicado de opinión en su uso no pronominal.	<i>se confirmó la huelga</i> <i>confirmó que no va a participar</i>
<i>afirmar</i>	Solo es predicado de opinión en su uso no pronominal.	<i>se afirmó sobre el mostrador</i> <i>afirmó que no va a participar</i>
<i>decir</i>	Se descarta la locución <i>es decir</i> .	<i>me considero de izquierdas, es decir, un hombre ...</i> <i>lo escuché decir que no va a participar</i>

Tabla 7.2 Desambiguación de verbos de opinión

En la sección de evaluación se analiza la cantidad de verbos de opinión que son marcados erróneamente como predicados de opinión por no haber sido desambiguados.

7.2.2 Reglas para el reconocimiento de la fuente

Mostramos en la tabla 7.3 algunas reglas (simplificadas) que integran el módulo para la identificación de la fuente. Estas reglas asignan la etiqueta fuente a los segmentos de texto que

son reconocidos por el cuerpo de la regla. Además de asignar la etiqueta fuente, como se muestra en la tabla, las reglas asignan valores a algunos atributos definidos para esta etiqueta, no mostrados en la tabla por simplicidad. Estos atributos son:

- código de la regla que asignó la etiqueta

- estructura sintáctica (sujeto antepuesto / sujeto pospuesto / complemento con *de*, etc.)

f1a# [fuente] => no(prep) >> gn << (zona,3), vOpFinSF	X dijo / X también opinó
f1b# [fuente] => punt, vOpFinSF, (zona,3) >> gn <<	..., dijo ayer X
f1c# [fuente] => punt, vOpFinSF, (zona,3), prep, gn >> gn <<	..., dijo a Brecha X
f2# [fuente] => vOpPartSF, "por" >> gn <<	expresado por X
f3a# [fuente] => nOp, "de" >> gn <<	apoyo de X
f3b# [fuente] => >> gn << vSopFin, op(det), nOp	X emitió una opinión
f3c# [fuente] => nOp, vSopPart, "por" >> gn <<	apoyo brindado por X
f4a# [fuente] => "según", op(vOp) >> gn <<	según X, ... / según dijo X, ...
f4b# [fuente] => finOr, "para" >> gn <<	. Para X, ...
f4c# [fuente] => introdAut >> gn <<	a juicio de X, ...
f5a# [fuente] => "a" >> gn << (zona,3), vOpFinSA	a X le molesta ...
f5b# [fuente] => vOpFinSA, "a" >> gn <<	... agrada a X

Tabla 7.3: Reglas simplificadas para el reconocimiento de la fuente. Notación utilizada:

gn - grupo nominal

zona,x - zona de exclusión de hasta x palabras

vOpFinSF, vOpFinSA - verbo de opinión, forma finita, SF: sujeto fuente, SA: sujeto asunto

vOpPartSF - verbo de opinión, participio, SF: sujeto fuente

nOp - nombre de opinión

vSopFin, vSopPart - verbo soporte, forma finita o participio

finOr - fin de oración

det - determinante

op - operador de opcionalidad

Las tres reglas f1 (f1a, f1b y f1c) identifican fuentes que son sujeto (antepuesto o pospuesto) de un verbo de opinión. Se permite que entre el sujeto y el verbo existan hasta 3 palabras que no pueden ser ni verbo, ni grupo nominal, ni puntuación, ni conjunción (*el senador este martes dijo ...*) y en el caso de la regla f1c se permite, además de la zona de exclusión, un grupo preposicional (prep+gn) entre la fuente y el verbo (*dijo ayer a la prensa el senador*). Como se dijo, las reglas mostradas están simplificadas, las reglas reales incluyen otras restricciones, como chequeo de concordancia entre sujeto y verbo. La regla f1a reconoce las fuentes de los ejemplos (E1, E2, E3, E6, E11, E12 y E14) y la regla f1b reconoce la fuente de (E4).

La regla f2 permite encontrar la fuente de un verbo de opinión en participio, que en estos casos es el complemento agente del verbo (*las palabras expresadas por el senador*).

Las tres reglas f3 (f3a, f3b y f3c) tratan los casos de nominalizaciones. La fuente suele encontrarse en el complemento con *de* (*las opiniones del senador*) pero también es común que la nominalización se encuentre en una construcción con verbo soporte (*el senador emitió una declaración*). Las reglas contemplan también los casos de verbos soporte en voz pasiva (*la declaración emitida por el senador*). En todos los casos se chequean las concordancias que correspondan. La regla f3a identifica la fuente de (E8) y la regla f3b la fuente de (E7).

Para la aplicación de la regla f3a se chequea, en el repertorio de predicados, el atributo de los nombres de opinión que indica el rol del complemento con *de*. La regla tiene dos versiones: si el atributo mencionado tiene valor 'fuente', la regla marca como fuente todo gn que cumpla las condiciones de la regla; si el valor del atributo es 'ambiguo', se exigen algunas propiedades al gn: tener como núcleo un nombre propio, contener un indicador de persona o contener un indicador de institución. La regla no se aplica en ningún caso si el valor del atributo es 'asunto'.

Las tres reglas f4 (f4a, f4b y f4c) marcan fuentes introducidas por elementos como *según*, *para*, *de acuerdo con*, *de acuerdo a*, *a juicio de*. La regla f4a se aplica cuando el introductor de fuente es *según*, caso en el cual es posible encontrar un verbo de opinión entre *según* y la fuente (*según dijo el senador*). La regla f4b se aplica para el introductor *para*, que introduce mucha ambigüedad por ser una preposición con muchos otros usos, por lo que se exige que esté en principio de oración. Para los introductores restantes se aplica la regla f4c. La regla f4a marca la fuente del ejemplo (E9) y la regla f4c la fuente de (E10).

Finalmente, las reglas f5 (f5a y f5b) identifican las fuentes de verbos cuyo sujeto tiene rol asunto y por lo tanto la fuente se expresa a través del complemento indirecto (*La propuesta (les) gusta a los senadores* / *A los senadores les gusta la propuesta*).

La clase a la cual pertenece cada verbo de opinión, información contenida en el repertorio de predicados, es tomada en cuenta por las reglas f1, f2, y f5 ya que es necesario saber qué rol corresponde al sujeto de cada verbo para poder identificar las fuentes. Las reglas f1 y f2 exigen sujeto con rol fuente, por lo que solo se aplican a verbos que tienen como clase más general *vo_sf*. Las reglas f5 exigen sujeto con rol asunto, por lo que chequean que los verbos pertenezcan a la clase *vo_sa*.

El ejemplo (E5) del Anexo A tiene un verbo con sujeto omitido, recuperable en la proposición anterior. Nuestro sistema de reglas no resuelve estos casos. En el capítulo 12, comentamos una extensión a nuestro trabajo que busca recuperar las fuentes omitidas y resolver correferencias entre fuentes.

7.2.3 Reglas para el reconocimiento del asunto

Esquemizamos las reglas para el asunto en la tabla 7.4.

La regla a1 marca como asunto los grupos preposicionales introducidos por preposiciones o locuciones prepositivas típicas de asunto o tema (*sobre*, *en relación a*, etc.), tomadas de una lista. Se pide que antes de la preposición haya un predicado de opinión, verbo o nombre, y se permite una zona de hasta 4 palabras entre el predicado y la preposición (*habló en las últimas horas sobre el tema*).

Las tres reglas a2 marcan el asunto teniendo en cuenta el valor del atributo del repertorio de verbos de opinión que indica la lista de preposiciones posibles del asunto, siempre que este tenga

sujeto con rol fuente. La regla a2a marca un grupo preposicional introducido por una preposición perteneciente a la lista registrada en el atributo mencionado. La regla a2b marca como asunto un grupo nominal, si el verbo pertenece a alguna de las clases que aceptan el asunto como complemento directo. El chequeo de esta condición se indica luego del símbolo //, mediante el texto *asuntoGN(vOp)* (el verbo de opinión acepta un grupo nominal como asunto). La regla a2c marca una proposición como asunto, si el verbo pertenece a alguna de las clases que aceptan un asunto preposicional. En este caso se toman las proposiciones marcadas previamente por ClaTex. Se indica este chequeo mediante el texto *asuntoProp(vOp)* (el verbo de opinión acepta una proposición como asunto).

a1# [asunto] => pOp, (zona,4) >> prepAs, gn <<	opinó sobre Y / opinión respecto a Y
a2a# [asunto] => vOpSF >> prep, gn << // listaPrep(pre)	criticó a Y
a2b# [asunto] => vOpSF >> gn << // asuntoGN(vOp)	rechazaron Y
a2c# [asunto] => vOpSF >> prop << // asuntoProp(vOp)	rechazaron que Y
a3# [asunto] => nOp >> "de", gn <<	anuncio de Y
a4a# [asunto] => >> prepAs, (zona,20) << vOpSF	En lo que respecta a Y, explicó ..
a4b# [asunto] => >> prepAs, (zona,20) << fuente	En cuanto a Y, X dijo ...
a5# [asunto] => >> gn << vOpPartSF	Y rechazado por X
a6a# [asunto] => no("a") >> gn << op(pronDat), vOpFinSA	Y (le) gusta a X
a6b# [asunto] => vOpFinSA >> gn <<	a X le molesta Y

Tabla 7.4: Reglas simplificadas para el reconocimiento del asunto. Notación utilizada:

gn - grupo nominal
zona,x - zona de exclusión de hasta x palabras
pOp - predicado de opinión (puede ser verbal o nominal)
vOpSF, vOpSA - verbo de opinión, SF: sujeto fuente, SA: sujeto asunto
vOpPartSF - verbo de opinión participio, SF: sujeto fuente
vOpFinSF, vOpFinSA - verbo de opinión finito, SF: sujeto fuente, SA: sujeto asunto
nOp - nombre de opinión
prepAs - preposición (o locución) introductora de asunto
prep - preposición
prop - proposición
pronDat - pronombre dativo
op - operador de opcionalidad
no(x) - operador de negación

La regla a3 marca como asunto un grupo introducido por *de*, siempre que no se haya marcado ese grupo como fuente en el módulo anterior.

Las reglas a4 marcan el asunto topicalizado (*En lo que respecta al tema debatido, el senador opinó ...*), introducido por alguna de las preposiciones de la lista de preposiciones de asunto (*sobre, en lo que respecta a, etc.*), permitiendo una zona de hasta 20 palabras como contenido del asunto en vez de restringirse a un grupo nominal como en los casos anteriores.

La regla a5 marca como asunto un grupo nominal que precede a un verbo de opinión en participio, como en *la situación denunciada por X*.

Por último, las reglas a6 (a6a y a6b) marcan el asunto de verbos que tienen sujeto con rol asunto, para lo cual se exige que el verbo tenga como clase más general a *vo_sa*. El asunto en estos casos es un grupo nominal, el sujeto del verbo.

Entre los ejemplos del Anexo A, en 6 casos el asunto se expresa en forma explícita. El asunto de (E1) es reconocido por la regla a4a, el de (E5) por la regla a2a, el asunto de (E6) y el de (E12) por la regla a2b, el asunto de (E11) es reconocido por la regla a3 y, por último, el asunto de (E13) es reconocido por la regla a1. En este último caso, como la locución introductora del asunto (*en contra de*) es negativa, el asunto queda con ese mismo valor de orientación semántica.

7.2.4 Reglas para el reconocimiento del mensaje

Como paso previo a la identificación del mensaje se marcaron con la etiqueta 'cita' todos los segmentos entrecomillados del texto. Para esto se pre-procesaron las comillas asignándole a cada una un atributo que toma los valores 'abre' o 'cierra'. La regla que marca los segmentos entrecomillados simplemente asigna la etiqueta cita a una zona de exclusión precedida por comillas con valor 'abre' y seguida por comillas con valor 'cierra'. La zona de exclusión tiene largo máximo 200, dado que en algunos casos la cita contiene varias oraciones. La única etiqueta que se excluye de la zona es el signo de puntuación comillas.

Las reglas para marcar el mensaje son de tres tipos:

- Reglas que marcan como mensaje cualquier segmento marcado entre comillas. Si dos o más citas aparecen en forma consecutiva quedan marcadas dentro de un único mensaje. Estas reglas marcan el mensaje en los ejemplos (E1) y (E11).
- Reglas que marcan como mensaje una proposición, segmentada previamente por ClaTex, precedida por un verbo de opinión perteneciente a alguna clase que acepte una cita indirecta, información tomada del repertorio de predicados. Estas reglas marcan el mensaje en los ejemplos (E2), (E3), (E12) y (E14).
- Reglas que marcan como mensaje la oración completa hasta la coma, cuando hay un inciso final de discurso con sujeto invertido. El mensaje de (E4) es reconocido por estas reglas.
- Reglas que marcan como mensaje las oraciones que contienen un indicador de fuente tipo *según*, cuya fuente fue reconocida previamente por el módulo para el reconocimiento de fuente. El segmento introducido por el indicador de fuente queda afuera del mensaje. Estas reglas reconocen el mensaje de los ejemplos (E9) y (E10).

7.2.5 Reglas para armar la opinión completa

El módulo de reglas que arma las opiniones en base a los elementos identificados contiene diferentes reglas para los distintos esquemas posibles observados, como por ejemplo: [asunto, fuente, predicado, mensaje], [fuente, predicado, mensaje], [fuente, predicado, asunto], [mensaje, predicado, fuente]. El total de esquemas posibles es 29, teniendo en cuenta, además del orden y la presencia o ausencia de los elementos, otras características como el tipo de predicado (nominal, verbal o preposicional).

Una regla particular de este módulo incorpora una oración o párrafo completo entrecomillado a la opinión inmediata anterior, siempre que no haya otro texto entre ambos. El esquema es [[opinion] + [“”]]. De este modo, el mensaje queda asociado a los elementos (fuente y asunto) de la opinión anterior, si estos son mencionados²⁰.

7.3 La orientación semántica

Las reglas, además de reconocer las opiniones y sus elementos, asignan valores de orientación semántica (OS) para cada elemento identificado y también para la opinión completa.

Nuestra propuesta consiste en calcular el valor final de OS de la opinión a partir de los valores de OS de sus elementos. Partimos de la hipótesis siguiente: la OS de la opinión es positiva si alguno de sus elementos tiene OS positiva; la OS de la opinión es negativa si alguno de sus elementos tiene OS negativa; y la OS la opinión es neutra si todos sus elementos son neutros.

Podría haber problemas si dos elementos de la opinión tuvieran valores de OS diferentes, pero creemos que esto no debería suceder si los valores son calculados correctamente y, de hecho, no encontramos hasta el momento ningún caso en ningún texto en donde esto suceda.

En el ejemplo (7.6) podemos pensar que el predicado positivo *apoyar* es contradictorio con los elementos subjetivos negativos *farsa*, *sectaria* y *censuradas*, presentes en el asunto. La contradicción desaparece si se atribuyen esas palabras negativas a su fuente, que es el autor del texto, mientras que la fuente de la opinión que se analiza, introducida por el predicado *apoyar*, es *Carlos Fuentes*, *García Márquez* y *Fernando del Paso*. En un análisis correcto de este ejemplo se debería asignar el valor neutro al asunto, ya que estamos reconociendo la opinión de la fuente que no es el autor del texto²¹.

(7.6) Carlos Fuentes, García Márquez y Fernando del Paso no han tenido la más ligera vacilación en **apoyar** esta *farsa sectaria donde otras voces mexicanas han sido censuradas*.

En nuestro sistema de reglas, la OS del predicado y del asunto se obtienen a partir de valores que estos ya tienen establecidos en los recursos léxicos utilizados. Para los predicados, la OS se deduce de su clase, que está registrada en el repertorio de predicados; los valores posibles son neutro, positivo o negativo. Para el asunto, la OS solamente es positiva o negativa si su introductor lo es, por ejemplo, si el introductor es *en contra de*, el asunto es negativo, si es *a favor de*, el asunto es positivo. En cualquier otro caso la OS del asunto es neutra. No tenemos en cuenta palabras de contenido subjetivo que puedan ocurrir dentro del asunto, ya que estas son expresadas por el autor del texto y no por la fuente citada, como mostramos en (7.6).

La fuente en general es neutra, salvo cuando se incluyen palabras como *optimista*, *decepcionado*, *seguidor*, *detractor*. Para estos casos, luego de la aplicación del módulo para el reconocimiento de fuentes, se ejecuta un procedimiento especial que busca ocurrencias de palabras como las mencionadas dentro de las fuentes y asigna un valor de OS positivo o negativo, según el caso.

20 En su análisis sobre fuente y asunto de las opiniones, Stoyanov et al (2008) comentan un caso como este.

21 En (Stoyanov et al, 2008) se analizan en forma similar algunos ejemplos para el inglés.

Solo si el predicado y el asunto son neutros, se toma el valor de la fuente, si esta es no neutra, como valor final de OS para la opinión. Las fuentes no neutras resultaron poco frecuentes, de hecho, en el corpus de evaluación no hubo ninguna ocurrencia de fuentes no neutras, por lo que no se evaluó el efecto que tiene la OS de la fuente sobre el valor final.

Creemos que, para casos como *el entusiasmado ministro dijo ...* o *los optimistas creen ...*, es bastante probable que la opinión completa mantenga la OS dada por la fuente. En cambio, para fuentes que incluyan palabras como *seguidor* o *detractor*, no necesariamente la opinión completa tendrá la misma OS que la fuente. Por ejemplo, en (7.7) se expresa una opinión positiva y en (7.8) una opinión es negativa, por más que en los dos casos se utiliza como fuente la palabra *seguidores*.

(7.7) Con su nueva imagen Charanga Habanera ha ganado adeptos, **han opinado seguidores de su música.**

(7.8) De otro lado, cabría añadir que, por su parte, los seguidores del shiísmo **rechazaron** por completo *a los Abasíes*.

No trabajamos con la OS del mensaje, como se aclaró en la introducción, de modo que este elemento siempre tiene OS no determinada en nuestro sistema. Definimos un valor especial para representar la OS no determinada ('nd').

La orientación semántica final de la opinión se calcula según el siguiente algoritmo:

```

entrada: <OS_fuen,OS_pred,OS_asun,OS_mens>
salida : OS_op

si (OS_pred <> neu) => OS_op=OS_pred
si no, si (OS_asun <> neu) => OS_op=OS_asun
    si no, si (OS_fuen <> neu) => OS_op=OS_fuen
    si no, OS_op=OS_mens

Notación:
    OS_op = OS de la opinión,
    idem para los componentes
    <> significa "distinto de"

```

Figura 7.2: Algoritmo para la determinación de la orientación semántica de la opinión completa

Si bien no calculamos el valor de OS del mensaje, lo incluimos en el algoritmo ya que creemos que, una vez que este sea resuelto, quedará resuelta la OS de la opinión completa. En la salida actual de nuestro sistema, si el predicado, el asunto y la fuente son neutros, sea cual sea el contenido del mensaje, la opinión total toma el valor no determinado.

El orden en que se toma la OS de cada elemento está determinado por la certeza con la que es posible obtener la OS de cada uno. En primer lugar, el predicado y el asunto tienen una OS determinada por las entradas léxicas por lo que no habría posibilidad de errores. Entre los dos elementos le damos prioridad al predicado ya que es más frecuente encontrar predicados no neutros que introductores de asunto no neutros. Luego miramos la fuente, que en algunos casos puede llevar a errores, como en los ejemplos que mostramos antes. Finalmente, consideramos el mensaje (a incorporar en el futuro) ya que es el elemento más difícil de analizar y es más probable que se cometan errores al determinar su OS. Esto se debe a que se trata de un segmento

extenso, con estructura oracional, en la cual pueden ocurrir diversos elementos subjetivos y modificadores de subjetividad que no es sencillo combinar.

Si alguno de los cuatro elementos que intervienen en el cálculo de la OS de la opinión no está presente, se asume para él el valor de OS neutro, de modo que no influya en el valor final. En particular, si la opinión no contiene mensaje, se considera que el mensaje es neutro, en vez de considerarlo no determinado.

7.4 Aplicación del sistema de reglas. Ejemplos

El sistema de reglas genera una salida que consiste en un archivo con el texto original al cual se le agregan etiquetas xml. Existen cinco etiquetas posibles: opinion, predicado, fuente, asunto y mensaje. Cada elemento tiene un solo atributo, llamado 'os', que representa la orientación semántica. Los valores posibles para este atributo son: 'neu' (neutro), 'pos' (positivo), 'neg' (negativo) y 'nd' (no determinado). El valor 'nd' solo se utiliza para el mensaje y la opinión.

Aplicamos el sistema de reglas al conjunto de ejemplos del Anexo A. La tabla 7.5 presenta las salidas correspondientes a E1, E2, E3 y E4, que incluyen verbos de comunicación y creencia.

E1	Consultado sobre la lentitud de los procesos judiciales uruguayos Carranza respondió: "Hay una situación de un muy alto número de presos sin condena, hay que agilizar los procesos".	Consultado <opinion os="nd"> <asunto os="neu">sobre la lentitud de los procesos judiciales uruguayos</asunto>, <fuente os="neu">Carranza</fuente> <predicado os="neu">respondió</predicado>: <mensaje os="nd"> "Hay una situación de un muy alto número de presos sin condena, hay que agilizar los procesos"</mensaje> </opinion>.
E2	El vicepresidente del Instituto Nacional de Carnes de Uruguay, Fernando Pérez Abella dijo que las conclusiones son muy alentadoras.	<opinion os="nd"> <fuente os="neu">El vicepresidente del Instituto Nacional de Carnes de Uruguay, Fernando Pérez Abella</fuente>, <predicado os="neu">dijo</predicado> <mensaje os="nd">que las conclusiones son muy alentadoras</mensaje> </opinion>.
E3	El 56% de los estadounidenses cree que la invasión de Irak en 2003 fue un error.	<opinion os="nd"> <fuente os="neu">El 56% de los estadounidenses</fuente> <predicado os="neu">cree</predicado> <mensaje os="nd">que la invasión de Irak en 2003 fue un error</mensaje> </opinion>.
E4	No existe más durante este gobierno el vocero presidencial como figura de estructura presidencial, afirmó Menem.	<opinion os="nd"> <mensaje os="nd">No existe más durante este gobierno el vocero presidencial como figura de estructura presidencial</mensaje>, <predicado os="neu">afirmó</predicado> <fuente os="neu">Menem</fuente> </opinion>.

Tabla 7.5: Aplicación del sistema de reglas a los ejemplos (E1) a (E4), que contienen verbos de comunicación y creencia.

Como se ve en la tabla 7.5, todos los casos de discurso reproducido son analizados correctamente por las reglas.

La tabla 7.6 muestra los ejemplos (E5) y (E6). Al igual que en la tabla 7.5, estos ejemplos contienen predicados verbales no neutros.

E5	La iglesia católica tomó partido durante el proceso revolucionario y apoyó a Victoriano Huerta.	La iglesia católica tomó partido durante el proceso revolucionario y <opinion os="pos"> <predicado os="pos">apoyó</predicado> <asunto os="neu">a Victoriano Huerta</asunto> </opinion>.
E6	Mientras que otros sectores como el Partido Comunista, Asamblea Uruguay, el Partido Socialista y Alianza Progresista rechazaron la propuesta.	Mientras que otros sectores como el Partido Comunista, Asamblea Uruguay, <opinion os="neg"> <fuente os="neu">el Partido Socialista y Alianza Progresista</fuente> <predicado os="neg">rechazaron</predicado> <asunto os="neu">la propuesta</asunto> </opinion>.

Tabla 7.6: Aplicación del sistema de reglas a los ejemplos (E5) y (E6), que contienen verbos subjetivos.

En los dos ejemplos de la tabla 7.6 existen problemas para la identificación de la fuente. En el ejemplo (E5) la fuente de la opinión está omitida y debe recuperarse en la proposición anterior (*la Iglesia Católica tomó partido durante el proceso revolucionario*), que está coordinada con la proposición que contiene la opinión. Como el sistema de reglas no abarca el problema de la recuperación de fuentes omitidas, para la evaluación se considera que opiniones como esta no contienen fuente, por lo que la no recuperación de estas fuentes no afecta los valores de *recall*.

En el ejemplo (E6) la fuente es identificada en forma parcial ya que el módulo de reglas que arman grupos nominales no pudo armar el grupo *otros sectores como el Partido Comunista, Asamblea Uruguay, el Partido Socialista y Alianza Progresista* por su estructura compleja. Por un lado, este módulo no reconoce grupos nominales con complementos que no estén introducidos por *de*. Por otro lado, tampoco es capaz de armar series de grupos nominales de más de dos elementos. Por lo tanto, queda marcado como fuente de la opinión el grupo nominal formado por *el Partido Socialista y Alianza Progresista*.

En los dos casos de la tabla 7.6 las opiniones toman el valor de orientación semántica de los predicados que, como ya se dijo, tienen valores no neutros. Estos valores finales de la opinión son correctos.

La tabla 7.7 muestra la salida de la aplicación del sistema de reglas a los ejemplos (E7) y (E8), que contienen predicados nominales, uno neutro y otro positivo.

En el ejemplo (E7), la opinión detectada solo contiene un predicado y una fuente, los dos elementos son bien reconocidos. Para el ejemplo (E8), se marcan el predicado y la fuente. El asunto ocupa el lugar de sujeto del verbo *tenemos*, pero está omitido. Si bien puede ser recuperado de la proposición anterior, en este trabajo no nos propusimos resolver ese problema, como ya comentamos al hablar de la fuente del ejemplo (E5).

E7	Clinton formuló su declaración durante rueda de prensa gráfica junto con el primer ministro irlandés Bertie Ahern.	<opinion os="neu"> <fuente os="neu">Clinton</fuente> formuló su <predicado os="neu">declaración</predicado> </opinion> durante rueda de prensa gráfica junto con el primer ministro irlandés Bertie Ahern.
E8	Nosotros coordinamos esta relación con la Comisión Nacional Bancaria y tenemos mucho apoyo de la Comisión.	Nosotros coordinamos esta relación con la Comisión Nacional Bancaria y tenemos mucho <opinion os="pos"> <predicado os="pos">apoyo</predicado> de <fuente os="neu">la Comisión</fuente> </opinion>.

Tabla 7.7: Aplicación del sistema de reglas a los ejemplos (E7) y (E8), que contienen predicados nominales.

La tabla 7.8 muestra la salida de la aplicación del sistema de reglas a los ejemplos (E9) y (E10), que contienen predicados preposicionales.

E9	Este sistema se utiliza en Estados Unidos desde 1982, según Roque Pifarré.	<opinion os="nd"> <mensaje os="nd">Este sistema se utiliza en Estados Unidos desde 1982</mensaje>, <predicado os="neu">según</predicado> <fuente os="neu">Roque Pifarré</fuente> </opinion>.
E10	De acuerdo a un informe de la Dirección de Prensa de Palacio de Gobierno, la reunión organizada por el Ministerio de la Presidencia se realizará en el Hotel París desde las primeras horas de hoy.	<opinion os="nd"> <predicado os="neu">De acuerdo a</predicado> <fuente os="neu">un informe de la Dirección de Prensa de Palacio de Gobierno,</fuente> <mensaje os="nd">la reunión organizada por el Ministerio de la Presidencia se realizará en el Hotel París desde las primeras horas de hoy</mensaje> </opinion>.

Tabla 7.8: Aplicación del sistema de reglas a los ejemplos (E9) y (E10), que contienen predicados preposicionales.

Como se ve en la tabla 7.8, los dos ejemplos con predicados preposicionales quedaron marcados correctamente.

La tabla 7.9 incluye algunos casos de opiniones con estructura compleja: por un lado, el ejemplo (E11) en el cual la opinión abarca más de una oración, por otro lado, el ejemplo (E12) que contiene dos opiniones anidadas.

En los dos ejemplos de la tabla 7.9 observamos que existen errores en el marcado del asunto. En (E11), el asunto está formado por una serie de grupos nominales separados por comas, de los cuales solo el primero es marcado como asunto. Este error conlleva a un error en la determinación de la opinión completa, ya que los grupos nominales no incluidos en el asunto alejan al predicado del mensaje. Por lo tanto, este último no es incluido dentro de la opinión expresada mediante el predicado *habló*, dando lugar a una nueva opinión. Realizamos una prueba con un ejemplo similar a (E11), simplificando el asunto. La salida obtenida es correcta. Mostramos el ejemplo modificado en (7.9) y la salida en (7.10).

E11	Entrevistado por el semanario Crónicas, el abogado Gustavo Salle habló del pedido de extradición de Henry Kinssinger, sus discrepancias con el Gobierno del Frente Amplio, el caso Young y el trabajo del Secretario de Presidencia, Gonzalo Fernández durante la dictadura. "Este Frente Amplio vende una imagen, escenifica determinada postura ideológica y en la práctica edifica otra totalmente opuesta . Se vende como socialista y es ortodoxamente neoliberal . Si yo hubiera estado preso y hubiese sido torturado y hoy tuviera que enfrentar con esa pesada carga existencial y afectiva, la traición política e ideológica que está consumando la cúpula del Frente Amplio no sé cuál sería mi reacción".	Entrevistado por el semanario Crónicas, <opinion os="neu"> <fuente os="neu">el abogado Gustavo Salle</fuente> <predicado os="neu">habló</predicado> <asunto os="neu">del pedido de extradición de Henry Kinssinger</asunto> </opinion> , sus discrepancias con el Gobierno del Frente Amplio, el caso Young y el trabajo del Secretario de Presidencia, Gonzalo Fernández durante la dictadura. <opinion os="nd"> <mensaje os="nd"> "Este Frente Amplio vende una imagen, escenifica determinada <predicado os="neu">postura</predicado> ideológica y en la práctica edifica otra totalmente opuesta . Se vende como socialista y es ortodoxamente neoliberal . Si yo hubiera estado preso y hubiese sido torturado y hoy tuviera que enfrentar con esa pesada carga existencial y afectiva, la traición política e ideológica que está consumando la cúpula del Frente Amplio no sé cuál sería mi reacción" </mensaje> </opinion>.
E12	Mahuad dijo que su país no rechazaba la idea de crear un parque binacional en la zona en disputa.	<opinion os="nd"> <fuente os="neu">Mahuad</fuente> <predicado os="neu">dijo</predicado> <mensaje os="nd"> que <opinion os="pos"> <fuente os="neu">su país</fuente> <predicado os="pos">no rechazaba</predicado> <asunto os="neu">la idea</asunto> </opinion> de crear un parque binacional en la zona en disputa </mensaje> </opinion>.

Tabla 7.9: Aplicación del sistema de reglas a los ejemplos (E11) y (E12), que contienen opiniones de estructura compleja.

(7.9) Entrevistado por el semanario Crónicas, el abogado Gustavo Salle habló del pedido de extradición de Henry Kinssinger. "Este Frente Amplio vende una imagen, escenifica determinada postura ideológica y en la práctica edifica otra totalmente opuesta. Se vende como socialista y es ortodoxamente neoliberal. Si yo hubiera estado preso y hubiese sido torturado y hoy tuviera que enfrentar con esa pesada carga existencial y afectiva, la traición política e ideológica que está consumando la cúpula del Frente Amplio no sé cuál sería mi reacción".

(7.10) Entrevistado por el semanario Crónicas,

```
<opinion os="nd">
  <fuente os="neu">el abogado Gustavo Salle</fuente>
  <predicado os="neu">habló</predicado>
  <asunto os="neu">del pedido de extradición de Henry
  Kinssinger</asunto>.
  <mensaje os="nd">"Este Frente Amplio vende una imagen, escenifica
  determinada postura ideológica y en la práctica edifica otra totalmente
  opuesta. Se vende como socialista y es ortodoxamente neoliberal. Si yo
  hubiera estado preso y hubiese sido torturado y hoy tuviera que enfrentar
  con esa pesada carga existencial y afectiva, la traición política e ideológica
  que está consumando la cúpula del Frente Amplio no sé cuál sería mi
  reacción"</mensaje>.
</opinion>
```

En la salida del ejemplo (E12), incluido en la tabla 7.9, vemos que el asunto es identificado en forma parcial: no se incluye el segmento *de crear un parque binacional en la zona en disputa* dentro del grupo nominal que constituye el asunto. Al igual que en el caso de la fuente del ejemplo (E6), el error es producido por las reglas para armar grupos nominales. El fragmento que no es incluido en el asunto, queda formando parte del mensaje.

En cuanto a la anidación de opiniones, el sistema se comporta correctamente, marcando los elementos de cada opinión sin problemas.

Finalmente, en la tabla 7.10 mostramos un ejemplo (E13) de opinión con asunto no neutro y un ejemplo (E14) de opinión con fuente no neutra.

E13	La Alianza Nacional de Agricultores Independientes de Cuba se pronuncia en contra de tales prácticas de corrupción.	<pre><opinion os="neg"> <fuente os="neu">La Alianza Nacional de Agricultores Independientes de Cuba</fuente> <predicado os="neu">se pronuncia</predicado> <asunto os="neg">en contra de tales prácticas de corrupción</asunto> </opinion>.</pre>
E14	Los optimistas sostienen que la probabilidad de que esto suceda es muy baja pues el comportamiento que han mostrado los embalses hasta el momento es un indicador de que San Pedro será benévolo y los niveles de agua no llegarán a extremos.	<pre><opinion os="pos"> <fuente os="pos">Los optimistas</fuente> <predicado os="neu">sostienen</predicado> <mensaje os="nd"> que la probabilidad de que esto suceda es muy baja pues el comportamiento que han mostrado los embalses hasta el momento es un indicador de que San Pedro será benévolo </mensaje> </opinion> y los niveles de agua no llegarán a extremos.</pre>

Tabla 7.10: Aplicación del sistema de reglas a los ejemplos (E13) y (E14), que contienen asunto y fuente no neutros.

Para el primer ejemplo de la tabla 7.10, la salida es correcta. Para el segundo ejemplo, se observa un reconocimiento parcial del mensaje. El error es cometido por el sistema para la segmentación en proposiciones (ClaTex), que no resuelve en forma correcta la coordinación

entre la última proposición (*los niveles de agua no llegarán a extremos*) y la proposición inmediatamente anterior (*que San Pablo será benévolo*). Como comentamos anteriormente para un caso de

Como se verá en el capítulo 8, que describe la evaluación de este sistema, los elementos identificados en forma parcial (como la fuente de (E6), el asunto de (E11) y (E12) y el mensaje de (E14)), por un lado, son considerados incorrectos cuando se calculan medidas exactas, pero, por otro lado, son considerados correctos cuando se calculan medidas con solapamientos. En la evaluación de las opiniones completas también se consideran medidas exactas y con solapamientos. En este último caso entran opiniones que tienen algún elemento marcado en forma parcial, como las de (E6), (E11) y (E14) y también las opiniones para las cuales alguno de sus elementos quedó sin marcar, como la opinión de (E12) que quedó sin el mensaje.

Resumiendo, de los 14 ejemplos que mostramos en esta sección, 10 son marcados correctamente en su totalidad y 4 contienen elementos parciales. No contamos entre los elementos no reconocidos las fuentes o asuntos omitidos, por más que estos sean recuperables en el contexto (fuente de (E5) y asunto de (E8)), ya que no nos propusimos resolver elementos anafóricos ni correferencias (ver la descripción de un trabajo relacionado con este problema en el capítulo 12).

7.5 Síntesis

Realizamos un sistema basado en la escritura de reglas contextuales que utiliza el repertorio de predicados de opinión para el reconocimiento de la opinión y sus elementos. El sistema genera como salida un texto con etiquetas xml como el siguiente:

```
<opinion os="pos">
  <fuente os="neu">Mujica</fuente>
  <predicado os="pos">respaldó</predicado>
  <asunto os="neu">importante inversión minera</asunto>
</opinion>
```

Se definieron diferentes módulos de reglas, uno para el reconocimiento de cada elemento (predicado, fuente, asunto, mensaje) y un módulo final para la opinión completa. Para cada elemento, salvo el mensaje, se da un valor de orientación semántica y para la opinión completa se calcula un valor final de orientación semántica, en función de los valores de los elementos.

El sistema fue aplicado a los ejemplos seleccionados del Anexo A, mostrando, en líneas generales, la cobertura que logran las reglas sobre el corpus de desarrollo. Si se toman en cuenta elementos parcialmente reconocidos, el sistema identifica prácticamente todas las opiniones y sus elementos. Este resultado solamente muestra que la gran mayoría de los fenómenos analizados para las expresiones de opinión fueron resueltos correctamente por las reglas. En el capítulo 8 se presenta una evaluación completa del sistema, esta vez sobre un corpus de testeo compuesto por textos no utilizados durante el desarrollo.

8. Evaluación del sistema basado en reglas contextuales

Para la evaluación del sistema trabajamos con un corpus de prensa digital de Uruguay, tomado de las mismas publicaciones de las cuales provienen los textos de prensa digital del corpus de desarrollo. El corpus contiene 38 artículos con un promedio de 300 palabras cada uno. El tamaño total es de aproximadamente 13.000 palabras (sin contar signos de puntuación) y 650 oraciones.

Se aplicaron las reglas al corpus completo y se hizo una revisión manual de la salida con el objetivo de evaluar la identificación de cada elemento de la opinión y, además, la identificación de opiniones completas. También se hizo una evaluación parcial de la asignación de valores de orientación semántica, teniendo en cuenta solamente los valores provenientes del predicado o el introductor de asunto.

Además de realizarse la evaluación de la performance de las reglas, durante la etapa de revisión manual se corrigió el texto anotado de modo que se obtuvo un corpus de 13.000 palabras anotado correctamente con las opiniones y sus elementos.

La tabla 8.1 muestra los resultados obtenidos y los valores de *precision* y *recall* calculados. Se calcularon dos tipos de medidas diferentes. Por un lado se consideraron correctos solamente los segmentos que fueron recuperados completamente por las reglas, llamándolos recuperados exactos. Por otro lado, se calculó una medida más laxa, considerando correctos tanto los segmentos recuperados completos como los que fueron recuperados en forma parcial, es decir, segmentos que se solapan con los segmentos correctos del texto. Estas medidas que incluyen solapamiento son usadas con frecuencia en los trabajos del área (Choi et al, 2005, Lu, 2010).

8.1 Resultados en el reconocimiento del predicado

Los resultados arrojados por la evaluación muestran que la lista de predicados de opinión utilizada es bastante completa: tenemos 256 predicados reconocidos en un total de 281 predicados presentes en el corpus²².

Creemos que podemos lograr mejoras incorporando al sistema más reglas para la desambiguación de los predicados, ya que todos los casos de predicados marcados incorrectamente (23 en un total de 281, un 8%) se deben a ambigüedad en su significado.

Por otro lado, durante el proceso de evaluación se recopilamos varios verbos que no están en el repertorio de predicados de opinión, causantes de los 25 casos de predicados no reconocidos. Estos verbos se incorporaron al repertorio para etapas posteriores del trabajo, estudiando previamente la frecuencia con la que son utilizados en su significado de opinión de modo de no introducir fuentes de ambigüedad.

22 Al momento de realizarse la evaluación del sistema de reglas, los elementos como *según*, *de acuerdo con*, etc., no estaban incluidos entre los predicados. Solamente teníamos predicados verbales y nominales. De todos modos, el sistema ya reconocía las opiniones introducidas por *según* y similares, considerando que no existía en esos casos un predicado. Por esta razón, en la tabla 8.1 se observa que existen más opiniones que predicados. De todo esto se desprende que los números que presentamos para los predicados corresponden solamente a predicados verbales y nominales.

	predicado	fuerza	asunto	mensaje	opinión
total en corpus	281	212	74	243	302
recuperados-ex	256	133	33	140	128
recuperados-so	0	20	13	64	104
no recuperados	25	57	28	39	70
incorrectos	23	11	2	10	14
precision-ex	92%	81%	67%	65%	52%
precision-so	92%	93%	96%	95%	94%
recall-ex	91%	63%	45%	58%	42%
recall-so	91%	72%	62%	84%	77%
medida F-ex	91.5%	71%	54%	61%	47%
medida F-so	91.5%	81%	75%	89%	85%

Tabla 8.1: Resultados de la evaluación del sistema de reglas. Las filas representan:

- *total*: cantidad total de elementos en el texto
- *recuperados-ex*: cantidad de elementos recuperados exactos
- *recuperados-so*: cantidad de elementos recuperados incluyendo solapamientos
- *no recuperados*: cantidad de elementos no recuperados
- *incorrectos*: cantidad de segmentos marcados incorrectos
- *precision-ex*: *precision* exacta, calculada usando recuperados-ex
- *precision-so*: *precision* con solapamientos,
calculada usando recuperados-ex + recuperados-so
- *recall-ex*: *recall* exacto, calculado usando recuperados-ex
- *recall-so*: *recall* con solapamientos,
calculado usando recuperados-ex + recuperados-so
- *medida F-ex*: medida F calculada usando *recall-ex* y *precision-ex*,
- *medida F-so*: medida F calculada usando *recall-so* y *precision-so*,

8.2 Resultados en el reconocimiento del mensaje

En cuanto al mensaje, en la mayoría de los casos en los que este fue reconocido en forma parcial encontramos lo que suele denominarse discurso pseudo-directo (ver sección 3.1.1 del capítulo 3), que consiste en citas mixtas directas e indirectas. Este estilo fue contemplado por las reglas pero no en todas sus posibilidades. En particular, los casos no reconocidos en su totalidad consisten en proposiciones introducidas por *que*, que contienen en su interior citas entre comillas que se extienden más allá del límite de la proposición. El mensaje es reconocido solamente hasta el fin de la proposición.

Por otro lado, algunos errores provenientes del segmentador en proposiciones ClaTex causaron el reconocimiento parcial del mensaje en varios casos (11 casos en 64).

Es notorio que este es uno de los elementos de la opinión para el cual se recuperó mayor cantidad de segmentos parciales (64 en 243), por tener estructura oracional suele ser el segmento de mayor extensión, por lo cual resulta más difícil su correcta delimitación.

8.3 Resultados en el reconocimiento de la fuente

Para las fuentes también observamos varios casos de reconocimiento parcial. Como se asume que las fuentes son grupos nominales, en muchos casos quedan algunos complementos afuera del grupo, como por ejemplo, las subordinadas que modifican al nombre y varios complementos preposicionales (solo se incluyen complementos con *de*).

Cuando la fuente no está mencionada explícitamente en la oración, fenómeno muy frecuente en español, ya que los verbos pueden tener sujeto omitido, es importante recuperar la fuente en el contexto. Como ya dijimos, este es un problema que no abordamos en este trabajo, por lo que no consideramos esas fuentes como elementos no recuperados. De todos modos, durante la evaluación se tomó nota de los casos para los cuales la fuente no mencionada en una opinión coincide con la fuente de la opinión inmediatamente anterior. Observamos que de 68 opiniones que no tienen fuente explícita, en 58 casos (85%) la fuente correspondiente coincide con la de la opinión anterior. No estamos teniendo en cuenta los casos en los cuales la fuente está expresada mediante un elemento anafórico, en esos casos marcamos como fuente dicho elemento.

La recuperación de fuentes omitidas y la resolución de correferencias para fuentes anafóricas se estudiaron en un trabajo posterior a este que comentamos en la sección sobre extensiones y trabajos futuros.

8.4 Resultados en el reconocimiento del asunto

Para el asunto, que en general es un grupo nominal, se dan varios casos de reconocimiento parcial debido a la no incorporación de algunos complementos, al igual que para la fuente. Observamos que la cantidad de opiniones que contienen menciones explícitas al asunto es bastante baja, solo 74 en 302 (un 24.5%). Por esta razón creemos que la evaluación realizada en cuanto al reconocimiento de este elemento no es suficiente.

8.5 Resultados de la determinación de la orientación semántica

Se evaluaron los valores de orientación semántica asignados por el sistema a las opiniones cuyos predicados o introductores de asunto son no neutros. De los 281 predicados del texto, 26 tienen asignados valores positivos (12) o negativos (14). Solo un caso, el verbo *especular*, que estaba clasificado como negativo dentro de nuestra lista de verbos, se consideró incorrecto. Si bien este verbo parece tener una connotación negativa, su ocurrencia en el corpus procesado es neutra (... *es otro de los próximos nombres que se especula renunciará para dedicarse a la campaña electoral*). Las opiniones que contienen estos predicados toman de ellos los valores de orientación semántica. En todos los casos se verificó que la extensión de estos valores a toda la opinión es correcta.

En el caso de los introductores de asunto, solamente se encontraron 3 con valores no neutros, todos negativos: 2 ocurrencias de *contra* y una de *en contra de*. En los tres casos las opiniones correspondientes quedaron con valores negativos, lo cual resulta correcto.

Si bien nuestra hipótesis sobre el efecto del valor de orientación semántica del predicado y del asunto sobre el valor final de la orientación semántica de la opinión parece, en principio, verificarse, la cantidad de elementos no neutros es claramente muy escasa como para sacar

conclusiones definitivas. Es necesaria una evaluación particular de estos fenómenos, para lo cual habría que confeccionar un corpus que contenga mayor cantidad de opiniones con este tipo de elementos.

Como ya se mencionó, no se encontraron fuentes no neutras. Además, no se trabajó sobre la orientación semántica del mensaje.

8.6 Síntesis

El sistema de reglas fue evaluado sobre un corpus de 13.000 palabras. Se calcularon medidas exactas y medidas con solapamiento. En el primer caso, solo se consideran correctos los elementos que son reconocidos completamente por el sistema. En el segundo caso, se incluyen entre los elementos reconocidos aquellos que fueron marcados en forma parcial, quedando algunas palabras pertenecientes al elemento sin marcar. Los resultados obtenidos se muestran en la tabla 8.2 (versión reducida de la tabla 8.1).

	predicado	fuentes	asunto	mensaje	opinión
<i>precision-exacta</i>	92%	81%	67%	65%	52%
<i>precision-solapamientos</i>	92%	93%	96%	95%	94%
<i>recall-exacta</i>	91%	63%	45%	58%	42%
<i>recall-solapamientos</i>	91%	72%	62%	84%	77%
<i>medida F-exacta</i>	91.5%	71%	54%	61%	47%
<i>medida F-solapamientos</i>	91.5%	81%	75%	89%	85%

Tabla 8.2: Resultados obtenidos para el sistema de reglas

Como se ve en la tabla, la cantidad de elementos marcados en forma parcial es importante para el asunto y el mensaje, elementos que suelen ser más extensos que el predicado y la fuente.

Los resultados para el predicado muestran que el repertorio tiene muy buena cobertura, aunque es posible mejorarlo agregando más predicados y estudiando casos de ambigüedad.

En el caso de la fuente, los elementos incompletos en general pierden algunos de sus complementos, por lo que podemos pensar que la información parcial recuperada es de todos modos de gran utilidad para conocer las fuentes de las opiniones.

La existencia de un segmento en el que se haga mención explícita al asunto es muy poco frecuente en el corpus (74 casos en 302 opiniones), por lo que la evaluación de este elemento no es del todo concluyente. De todos modos, los valores con solapamientos obtenidos pueden ser de utilidad.

En cuanto al mensaje, creemos que es posible mejorar los resultados ya que algunos de los errores detectados tienen que ver con formas mixtas de discurso directo y discurso indirecto que pueden incorporarse a las reglas.

El corpus de testeo fue corregido manualmente durante la evaluación, generándose de este modo un corpus anotado con las opiniones y sus elementos que ya está siendo utilizado en un proyecto sobre la orientación semántica del mensaje (ver capítulo 12).

9. Sistema basado en aprendizaje automático

El segundo sistema informático consiste en la aplicación de técnicas de aprendizaje automático en base al modelo Campos Aleatorios Condicionados (*Conditional Random Fields*, CRF). Este sistema solo se desarrolló para el reconocimiento de fuentes de opiniones.

Para trabajar con técnicas de aprendizaje automático es necesario contar con ejemplos anotados que permitan entrenar un clasificador. Por esta razón, llevamos adelante una tarea de anotación manual de corpus, marcando las fuentes de opiniones. El corpus obtenido de este modo tiene aproximadamente 40.000 palabras.

Describimos a continuación, luego de una breve introducción al modelo CRF, los experimentos llevados a cabo y los resultados obtenidos.

9.1 Introducción al modelo *Conditional Random Fields*

Conditional Random Fields (Lafferty et al, 2001) es un modelo probabilístico secuencial que ha mostrado ser exitoso en diversas aplicaciones de Procesamiento de Lenguaje Natural, como segmentación de grupos nominales, identificación de entidades nombradas y etiquetado morfológico (Sutton et al, 2010). También se ha utilizado en el área Minería de Opiniones para el reconocimiento de fuentes del inglés (Choi et al, 2005 y 2006) y para la clasificación de oraciones subjetivas del inglés y del chino en positivas y negativas (Nakagawa et al, 2010).

Planteamos el problema de reconocimiento de fuentes como un problema de clasificación secuencial. En este tipo de problema, se estima la probabilidad condicional de una secuencia de valores de la variable de salida (la clase de cada unidad léxica) dada una secuencia de entrada (observaciones).

Como ventaja de este modelo respecto a enfoques alternativos usados para problemas de este tipo, como los Modelos Ocultos de Markov (*Hidden Markov Models*), podemos señalar la posibilidad de usar un amplio espectro de funciones tipo atributo construidas a partir de los datos de entrada y de la variable de salida, sin requerimientos de independencia para distintos atributos entre sí. En relación a los Modelos Markovianos de Máxima Entropía (*Maximum Entropy Markov Models*), se evita el problema conocido como *bias problem* (Lafferty et al, 2001), por el cual se produce un sesgo hacia estados con pocas transiciones de salida.

En lo que sigue, introducimos brevemente el modelo siguiendo a (Lafferty et al, 2001, Wallach, 2004, Sutton et al, 2010 y Mesa, 2010).

Para el caso particular de un *Lineal Chain CRF* (LCCRF), que es un CRF con estructura de cadena lineal, tenemos la siguiente definición (adaptada de Sutton et al, 2010 y Mesa, 2010):

Sean X e Y vectores aleatorios, λ_j un parámetro que toma valores en $\mathbb{R}^m$ y sea $\{f_j(y, y', x, i)\}_{j=1..m}$ un conjunto de funciones características (o funciones de atributos) que toman valores reales. Un LCCRF es una distribución $P(y|x)$ que toma la forma siguiente:

$$\mathbb{P}(y|x) = \frac{1}{Z_{\lambda}(x)} \prod_{i=1}^n \exp \left(\sum_{j=1}^m \lambda_j f_j(y_{i-1}, y_i, x, i) \right)$$

La fórmula muestra que la probabilidad $P(y|x)$ se calcula en función del valor actual de la salida (y_i), del valor de la salida en la posición anterior (y_{i-1}) y de los valores de diferentes posiciones i de la secuencia completa de observaciones (x).

Las f_j son funciones características que pueden tomar valores reales, aunque frecuentemente son funciones booleanas, y permiten modelar propiedades que relacionan la secuencia de observaciones (x) y el valor de salida del estado actual (y_i) y del anterior (y_{i-1}). Por ejemplo, es posible establecer que si en un estado y_i la etiqueta de salida es NPL (nombre propio de lugar), en el estado anterior y_{i-1} la etiqueta de salida es Prep y el valor de x en i es la palabra *en*, entonces la función vale 1, en cualquier otro caso vale 0.

Z es una constante de normalización que se calcula como sigue:

$$Z_{\lambda}(x) = \sum_y \exp \left(\sum_{i=1}^n \sum_{j=1}^m \lambda_j f_j(y_{i-1}, y_i, x, i) \right)$$

El entrenamiento del modelo consiste en estimar los parámetros λ_j de modo de maximizar la probabilidad $P(y|x)$.

9.2 Descripción de los experimentos

Utilizamos la herramienta CRF++ (WebCRF++) para aplicar *Conditional Random Fields* al problema de identificación de las fuentes de las opiniones.

Partimos el corpus anotado de modo de contar con un corpus de entrenamiento, de 30.000 palabras, y un corpus de testeo, de 10.000. Definimos un conjunto básico de atributos de entrada en base al cual realizamos diversos experimentos. Además, realizamos otros experimentos incorporando atributos nuevos o eliminando algunos atributos del conjunto básico.

Para los diferentes experimentos también introducimos variantes en otros elementos: el tamaño del corpus, la forma en que se combinan los atributos de entrada para calcular el modelo, el tipo de segmentación del texto.

En todos los casos, incluimos en el corpus una columna con los valores de salida. Utilizamos el sistema de anotaciones B-I-O que permite anotar secuencias. La etiqueta B indica que el elemento es el inicio (*Begin*) de una secuencia, la etiqueta I indica que el elemento está en el interior (*Inside*) de una secuencia y la etiqueta O indica que el elemento no pertenece a ninguna secuencia (*Outside*).

Como solamente trabajamos en la identificación de fuentes de opiniones, los valores posibles para la salida son: B-FU, inicio de una fuente, I-FU, interior de una fuente, O, otros casos.

Describiremos a continuación las distintas variantes y luego los experimentos realizados en base a diferentes combinaciones de esas variantes.

9.2.1 Atributos de entrada

Conjunto básico de atributos

Los atributos utilizados para el entrenamiento deben poder ser obtenidos en forma automática, de modo que el clasificador que resulte del aprendizaje pueda ser aplicado sobre cualquier texto, aplicando sobre él un pre-procesamiento completamente automático.

Por esta razón, como atributos de entrada incluimos información brindada por el etiquetador FreeLing y por los repertorios de predicados de opinión confeccionados para este trabajo.

El conjunto básico de atributos de entrada, llamado AtrsBas, contiene 8 elementos:

- palabra: palabra del texto, según la segmentación de FreeLing, incluyendo palabras compuestas y locuciones.
- lema: lema de cada palabra. Tener en cuenta el lema como atributo, además de la palabra, permite que se realicen generalizaciones que son fundamentales para el clasificador.
- categoría gramatical: primera letra de la etiqueta asignada por FreeLing a cada palabra, esa letra indica la categoría gramatical, las letras restantes aportan información morfológica²³.
- tipo de nombre: para las palabras cuya categoría es nombre, permite indicar si se trata de un predicado de opinión (como *declaración*, *apoyo*, etc.) o de cualquier otro nombre, en cuyo caso se indica si es nombre común o nombre propio. Para las categorías restantes este atributo vale 0.
- tipo de verbo: para las palabras cuya categoría es verbo, permite indicar si se trata de un predicado de opinión (como *declarar*, *apoyar*, etc.) o de cualquier otro verbo, en cuyo caso se indica si es un verbo principal, un verbo auxiliar o un verbo soporte (*emitir*, *realizar*, etc.). Para las categorías restantes este atributo vale 0.
- número: para las palabras cuya categoría es determinante, nombre, adjetivo o verbo, se indica su número. Para las categorías restantes este atributo vale 0. Esta información es importante para que se respete la concordancia en número tanto al interior de los grupos nominales (forma sintáctica más frecuente de las fuentes) como entre fuente y verbo de opinión, en los casos en que esta es el sujeto de dicho verbo.
- género: para las palabras cuya categoría es determinante, nombre o adjetivo, se indica su género. Para las categorías restantes este atributo vale 0. Esta información es importante para que se respete la concordancia en género al interior de los grupos nominales.
- introductor de fuente: indica si el elemento es un predicado de opinión preposicional, como *según*, *de acuerdo con*, etc.²⁴. En caso afirmativo el atributo vale 1, en otros casos vale 0.

23 Como se aclaró en 7.1, FreeLing utiliza un conjunto de etiquetas basado en EAGLES (WebEAGLES).

24 Llamamos a estos predicados 'introductores de fuente' ya que el grupo nominal que introducen constituye la fuente de la opinión: *Según [el presidente], las medidas adoptadas*

En la tabla 9.1 mostramos los valores posibles para cada atributo y la forma en que estos valores son obtenidos.

Atr	Descripción	Valores posibles	Origen del valor asignado
0	palabra del texto	conj. de pals. del corpus	texto
1	lema	conj. de lemas del corpus	FreeLing
2	categoría	N - nombre V - verbo D - determinante A - adjetivo S - preposición R - adverbio F - puntuación otros	FreeLing
3	tipo de nombre	si categoría = N: C - común P - propio	FreeLing
		Op - opinión	lista de noms de op + manual
		otras categorías: 0	valor por defecto
4	tipo de verbo	si categoría = V: M - verbo principal A - auxiliar haber S - auxiliar ser	FreeLing
		Sop - verbo soporte Op - verbo de opinión	lista de vbs soporte + manual lista de vbs de op + manual
		otras categorías: 0	valor por defecto
5	número	si categoría=N,D A,V: P - plural S - singular	FreeLing
		otras categorías: 0	valor por defecto
6	género	si categoría = N, D, A: M - masculino S - femenino C - común	FreeLing
		otras categorías: 0	valor por defecto
7	introducción de fuente	si palabra = 'según', etc. 1 - introd. de fuente	lista de introductores de fuente
		otras categorías: 0	valor por defecto

Tabla 9.1: AtrsBas: Conjunto básico de atributos de entrada

Cabe aclarar que, si bien los predicados de opinión verbales y nominales fueron tomados del repertorio utilizado por el sistema de reglas, se hizo además una revisión manual de los corpus para detectar elementos relevantes, no incluidos allí. Estos elementos fueron incorporados al

repertorio (lista de nombres de opinión y lista de verbos de opinión) de modo de mejorarlo.

Esto se hizo con el objetivo de poder evaluar los resultados de aplicar CRF al reconocimiento de fuentes, eliminando en lo posible los errores provenientes de factores externos a esta tarea concreta, como la no identificación de algunos predicados de opinión.

En la figura 9.1 mostramos algunos fragmentos del corpus de entrenamiento. Hay tres predicados de opinión: *destacó*, *precisiones* y *según*. El primero, por ser verbal, lleva el valor 'Op' en el atributo 4, correspondiente al 'tipo de verbo'. El segundo, por ser nominal, lleva el valor 'Op' en el atributo 5, correspondiente al 'tipo de nombre'. El tercero es un predicado preposicional o introductor de fuente, por lo que lleva el valor '1' en el atributo 7. El verbo *realizó*, por ser un verbo soporte, lleva el valor 'Sop' en el atributo 4.

Las fuentes a identificar son las secuencias de palabras [*La ministra*], [*La ministra de Salud Pública*] y [*Muñoz*]. A estas secuencias de palabras se les asignan los valores de salida B-FU, para el primer elemento de cada secuencia, e I-FU, para los restantes. Todas las demás líneas del ejemplo tienen O en la última columna, indicando que la palabra no integra ninguna fuente.

0	1	2	3	4	5	6	7	sal
La	el	D	0	0	S	F	0	B-FU
ministra	ministra	N	C	0	S	F	0	I-FU
destacó	destacar	V	0	Op	S	0	0	O
la	el	D	0	0	S	F	0	O
prevención	prevención	N	C	0	S	F	0	O
...								
La	el	D	0	0	S	F	0	B-FU
ministra	ministra	N	C	0	S	F	0	I-FU
de	de	S	0	0	0	0	0	I-FU
Salud_Pública	salud_pública	N	P	0	0	0	0	I-FU
realizó	realizar	V	0	Sop	S	0	0	O
algunas	alguno	D	0	0	P	F	0	O
precisiones	precisión	N	Op	0	P	F	0	O
...								
Las	el	D	0	0	P	F	0	O
medidas	medida	N	C	0	P	F	0	O
,	,	F	0	0	0	0	0	O
según	según	S	0	0	0	0	1	O
Muñoz	muñoz	N	P	0	0	0	0	B-FU
,	,	F	0	0	0	0	0	O
tuvieron	tener	V	0	M	P	0	0	O
...								

Figura 9.1: Extracto del corpus de entrenamiento. Los atributos 0 a 7 son los que se explican en la tabla 9.1. 'sal' es el atributo de salida.

Incorporación de grupos nominales

Realizamos otro experimento en base a una segmentación diferente del corpus. Se pre-procesó el corpus con el analizador sintáctico superficial de FreeLing, a partir del cual obtuvimos los grupos nominales. Con la finalidad de generar un corpus con distancias menores entre algunos elementos, como por ejemplo, los predicados de opinión y sus fuentes, se mantuvieron dichos grupos en una sola línea, como si fueran locuciones, tomándose para sus atributos los valores correspondientes al núcleo de cada grupo. La categoría N (nombre) fue sustituida por SN (sintagma nominal) y se descartó uno de los atributos, el género, ya que solo tiene sentido para la formación de grupos nominales. Llamamos a este conjunto de atributos AtrsBas+SN.

Al compactarse los elementos que componen cada grupo nominal en un solo elemento, la columna correspondiente a la palabra toma valores que consideramos poco representativos, ya que consisten en una concatenación de diferentes palabras unidas por un guion bajo. Por esta razón tuvimos en cuenta también un conjunto de atributos similar al descrito en el párrafo anterior pero excluyendo el atributo palabra, llamamos a esta variante AtrsBas+SN-Pal.

Incorporación de propiedades de los predicados de opinión

Incorporamos al conjunto AtrsBas nuevos atributos, provenientes del repertorio de predicados de opinión (capítulo 6). De toda la información que contiene el repertorio, tomada del modelo definido para los predicados de opinión (capítulo 5), solamente agregamos como atributo la que tiene relación con la identificación de las fuentes de las opiniones:

- para los verbos de opinión, el rol del sujeto (fuente o asunto);
- para los nombres de opinión, el tipo de complemento con *de* (fuente, asunto, ambiguo).

Llamamos AtrsBas+PrOp a este conjunto de atributos.

9.2.2 Tamaño del corpus

Se realizaron experimentos con corpus de entrenamiento de dos tamaños diferentes: 30.000 palabras y 15.000 palabras.

Se utilizó el mismo corpus de testeo de 10.000 palabras en ambos casos, de este modo los resultados son comparables.

Llamamos al corpus de 30.000 palabras C30 y al de 15.000 palabras C15.

El corpus C30 tiene un total de 486 fuentes y el corpus C15 tiene 227 fuentes.

El corpus de testeo tiene 158 fuentes.

9.2.3 Combinación de atributos de entrada

La operativa de CRF++ requiere que se definan familias de funciones de atributos (*templates*, según la terminología de la documentación).

Para nuestro trabajo especificamos un conjunto básico de *templates* que utilizamos en la mayoría de los experimentos. Para algunos experimentos particulares modificamos dicho conjunto, como se explica más adelante.

Conjunto básico de templates

En primer lugar, definimos templates que indican, para cada atributo, cuántos valores anteriores y cuántos posteriores queremos examinar:

- Para los atributos 'categoría', 'tipo de nombre' y 'tipo de verbo', para cada línea del corpus se examinan los valores tomados por el atributo en la línea actual, los valores tomados en las cuatro líneas anteriores y los valores tomados en las cuatro líneas siguientes.

- Para los atributos 'palabra', 'lema', 'número', 'género' e 'introducción de fuente', para cada línea del corpus se examinan los valores tomados por el atributo en la línea actual, los valores tomados en las dos líneas anteriores y los valores tomados en las dos líneas siguientes.

Como se observa, para los atributos que más inciden en la determinación de la fuente se definieron ventanas más grandes, hasta 4 líneas anteriores y siguientes.

Por otro lado, se definieron templates que generan atributos combinados, concatenando para cada línea el valor actual con el valor anterior y/o el siguiente. Para los atributos 'categoría', 'tipo de nombre' y 'tipo de verbo', se generan atributos combinados tomando dos, tres o cuatro valores, por ejemplo: [valor $i-1$ /valor i], [valor $i-1$ /valor i /valor $i+1$], [valor i /valor $i+1$ /valor $i+2$ /valor $i+3$]. Para los atributos restantes solamente se generan atributos combinando dos o tres valores.

Por último, existe un template particular que indica si se desea trabajar con bigramas o no. En el contexto de esta herramienta, un bigrama consiste en la combinación del atributo de salida actual con el atributo de salida anterior. Si se incluye esta indicación entre los templates, se generan automáticamente todos los bigramas.

Variantes en el tamaño de la ventana

En algunos experimentos, en lugar de definir las ventanas como se describió anteriormente, se utilizaron solamente dos valores anteriores y dos posteriores para todos los atributos. Más adelante, hablamos de ventana 2-4 para los casos en los cuales se configuran ventanas de 2 y 4 elementos anteriores y posteriores, según el atributo, y ventana 2-2 para los casos en los cuales la ventana siempre es de dos elementos.

Variantes en el uso de atributos combinados

Para algunos experimentos eliminamos los templates que generan atributos combinados.

Variantes en el uso de bigramas

También realizamos algunos experimentos sin bigramas.

9.3 Resultados de los experimentos

Para cada experimento, luego de obtenerse un modelo a partir del aprendizaje sobre el corpus de entrenamiento, se aplica dicho modelo al corpus de testeo. Se obtiene como salida el corpus con anotaciones, las cuales deben compararse con las anotaciones manuales para la evaluación de los resultados.

De dicha comparación, realizada gracias al script provisto en la web de CoNLL 2000 (Web CoNLL2000), surgen los valores de *precision* y *recall* que se muestran en la tabla 9.2. Los valores que mostramos corresponden todos a medidas exactas (sin solapamientos). Dado que el script utilizado solo permite calcular este tipo de medidas, no mostramos valores con solapamientos.

A continuación analizamos los datos presentados en la tabla.

Exp	Corpus	Atributos	Ventana	Combinados	Bigrama	<i>Precision</i>	<i>Recall</i>	F
1	C15	AtrsBas	2-4	sí	sí	81.19	50.93	62.60
2	C30	AtrsBas	2-4	sí	sí	92.04	64.60	75.91
3	C30	AtrsBas	2-2	sí	sí	91.67	61.49	73.61
4	C30	AtrsBas	2-4	no	sí	89.83	65.84	75.99
5	C30	AtrsBas	2-4	sí	no	51.19	53.75	52.44
6	C30	AtrsBas+SN	2-4	sí	sí	87.93	61.45	72.34
7	C30	AtrsBas+SN-Pal	2-4	sí	sí	87.29	62.05	72.54
8	C30	AtrsBas+PrOp	2-4	sí	sí	91.89	63.35	75.00

Tabla 9.2: Resultados de los diferentes experimentos

9.3.1 Análisis de los resultados de los experimentos realizados sobre el conjunto básico de atributos

La mayoría de los experimentos (experimentos 1 a 5) fueron realizados con el conjunto básico de atributos. De este modo se pudo evaluar el efecto de los demás factores: tamaño del corpus, tamaño de ventana, utilización de atributos combinados y utilización de bigramas.

Observamos que el tamaño del corpus de entrenamiento tiene gran incidencia en los resultados. Las mejoras obtenidas duplicando el corpus inicial (de 15.000 palabras) son muy significativas: el valor de *recall* aumentó 11 puntos y el de *precision* 14, aproximadamente, tomando los valores de los experimentos 1 y 2. Podemos pensar que agrandar aún más el corpus puede aportar mejoras importantes en los resultados. De ahora en más, las comparaciones se hacen siempre respecto al experimento 2, realizado con el corpus de entrenamiento de 30.000 palabras.

Por otro lado, si solo se trabaja con ventanas de dos elementos previos y dos posteriores (experimento 3), los resultados empeoran para el valor de *recall* que disminuye 3 puntos. La *precision* baja apenas unos 0.40 puntos, lo cual no es significativo. Creemos que la decisión de extenderse hasta 4 elementos previos y 4 posteriores para los atributos más relevantes es acertada. Podría ser interesante hacer más pruebas extendiendo la ventana a 4 elementos para todos los atributos, o extendiéndola aún más para los atributos relevantes.

El experimento 4 muestra los resultados de eliminar los atributos combinados. Se ve que este factor tiene incidencia diferente en los valores de *precision* y *recall*. El valor de *precision* disminuye en poco menos de 2 puntos mientras que el de *recall* aumenta un poco más de un punto. Si bien las diferencias en los resultados no son muy importantes, es interesante tener en cuenta este factor en otras pruebas ya que puede aumentar el valor de *recall*. Como se verá en el capítulo 9, para los últimos experimentos se realizaron pruebas con y sin atributos combinados.

El experimento 5 muestra que los resultados empeoran notoriamente cuando no se generan bigramas, principalmente en cuanto al valor de *precision* que baja más de 40 puntos. El valor de *recall* también empeora en una cifra muy significativa, baja unos 10 puntos. Uno de los errores que se generan cuando no se usan bigramas es la anotación de fuentes sin elemento inicial (B-

FU), ya que se pierde la noción de secuencia entre los valores B-FU e I-FU, al no tenerse en cuenta la relación entre el valor de salida anterior y el valor de salida actual, en cada línea.

9.3.2 Análisis de los resultados de los experimentos realizados sobre variantes del conjunto de atributos

Los experimentos 6 y 7 muestran los efectos de la compactación de grupos nominales en una sola línea. Teniendo en cuenta todos los atributos (experimento 6) los resultados sufrieron una disminución de 4 puntos en *precision* y 3 puntos en *recall*. Excluyendo la palabra del conjunto de atributos (experimento 7) la baja en *precision* alcanza unos 5 puntos y en *recall* alrededor de 2.5 puntos.

Contrariamente a lo que creíamos, compactar los grupos nominales no aporta mejoras al reconocimiento de fuentes sino que produce resultados peores. En particular, analizamos algunos casos de fuentes alejadas de sus predicados, por ser los casos que motivaron este experimento, comprobando que siguen sin ser reconocidas. Además, la exclusión del atributo palabra, experimento 7, no incide prácticamente en los resultados respecto al experimento 6.

El experimento 8 muestra los resultados de incorporar al corpus algunos atributos con información asociada a los predicados de opinión. Recordamos que la información incorporada indica: para los predicados verbales, si el sujeto tiene rol fuente o rol asunto; para los predicados nominales, si el complemento con *de* es la fuente, el asunto o cualquiera de los dos (ambiguo).

Observamos que los valores de *precision* y *recall* son casi iguales a los obtenidos en el experimento 2, ambos bajaron, pero menos de un punto. Analizamos los predicados de opinión contenidos en los corpus de entrenamiento y testeo para estudiar los valores que toman los atributos incorporados, de modo de explicar su influencia casi nula.

Por un lado, verificamos que la inmensa mayoría de verbos de opinión del corpus tienen sujeto con rol fuente. Hay muy pocas ocurrencias de verbos de opinión con rol asunto para el sujeto: solamente 4 en el corpus de entrenamiento (*preocupar, molestar*) y 3 en el de testeo (*satisfacer, gustar, conformar*). Además, esas pocas ocurrencias aparecen en estructuras sintácticas muy diferentes entre sí, lo cual no permite que se pueda generalizar la forma de encontrar su fuente. De hecho, para ninguno de los 3 verbos que tienen sujeto con rol asunto en el corpus de testeo se identificó la fuente, en ninguno de los experimentos realizados.

Por otra parte, estudiamos la ocurrencia de nombres de opinión con complementos con *de* en el corpus. En el corpus de entrenamiento encontramos 26 casos, de los cuales 9 están anotados con rol ambiguo para su complemento. Esto significa que dicho complemento puede ser el asunto (*reclamo de cumplimiento del convenio*) o la fuente (*declaraciones de un concejal*). De esos 9 casos, 4 son efectivamente una fuente y 5 contienen el asunto. Los restantes 17 nombres de opinión que tienen un complemento con *de* están anotados con rol fuente para su complemento. Solamente 1 de esos 17 casos no es fuente.

En el corpus de testeo hay 3 nombres de opinión cuyo atributo indica que el complemento con *de* es ambiguo, pero para ninguno de ellos se trata de la fuente. Los 3 elementos fueron bien analizados por los diferentes experimentos, ya que en ningún caso esos complementos con *de* fueron marcados como fuente. Por otra parte, hay 4 nombres de opinión para los cuales se indica

que el complemento tiene rol fuente y en los 4 casos se trata efectivamente de la fuente. De estas 4 fuentes, solo una es reconocida tanto por el experimento 2 como por el experimento 8.

Pensamos que la incorporación de esta información sobre las fuentes de los predicados de opinión podría ser útil, sobre todo para identificar las fuentes de los predicados de opinión nominales. Es necesario trabajar con un corpus mayor, en donde haya más ocurrencias de los diferentes casos posibles, para que el clasificador pueda hacer generalizaciones.

9.4 Síntesis

Aplicamos aprendizaje automático en base al modelo *Conditional Random Fields* (CRF) para el reconocimiento de la fuente. Utilizamos un corpus de entrenamiento de 30.000 *tokens* y un corpus de testeo de 10.000.

Utilizamos como atributos de entrada, por un lado, información proveniente del etiquetador morfo-sintáctico (palabra, lema, categoría gramatical, tipo de verbo, tipo de nombre, número y género), por otro lado, información sobre predicados de opinión. Esta información consiste en indicar, para cada verbo y nombre, si este pertenece al repertorio de predicados de opinión, y para las preposiciones o locuciones preposicionales, si se trata de un predicado de opinión preposicional (como *según*).

Sobre este conjunto básico de atributos aplicamos algunas variantes, como la compactación de grupos nominales y la incorporación de cierta información contenida en el repertorio para cada pieza léxica, en particular, información relacionada con la ubicación de las fuentes. Estas variantes no produjeron mejoras en los resultados, respecto a los experimentos realizados con el conjunto básico de atributos.

Por otro lado, realizamos diferentes experimentos variando la forma en que se combinan los atributos. Realizamos cambios en los tamaños de ventanas (cantidad de elementos anteriores y posteriores que se tienen en cuenta para cada posición), la utilización de atributos combinados y la utilización de bigramas para los valores de salida. La variante que más impactó es la utilización de bigramas para atributos de salida, que produce resultados notoriamente mejores que su no utilización.

Los mejores resultados obtenidos, entre todos los experimentos realizados, se muestran en la tabla 9.3 (que incluye solamente las líneas correspondientes a los experimentos 2 y 4 de la tabla 9.2, por ser los que dieron los mejores valores de *precision* y *recall*).

Exp	Corpus	Atributos	Ventana	Combinados	Bigrama	<i>Precision</i>	<i>Recall</i>	F
2	C30	AtrsBas	2-4	sí	sí	92.04	64.60	75.91
4	C30	AtrsBas	2-4	no	sí	89.83	65.84	75.99

Tabla 9.3: Mejores resultados obtenidos

Los dos experimentos tienen varias características en común:

- se realizaron sobre el corpus total de entrenamiento de 30.0000 *tokens* (esto se indica en la columna “Corpus”),

- se utilizó el conjunto básico de atributos (columna “Atributos”),
- se definieron ventanas de hasta 4 elementos anteriores y posteriores para los atributos con información sobre los nombres y verbos de opinión y para la categoría gramatical y ventanas de hasta 2 elementos anteriores y posteriores para el resto de los atributos (columna “Ventana”),
- se habilitó la opción de usar bigramas para los valores de salida (columna “Bigrama”).

La única diferencia entre los dos experimentos es la incorporación de atributos combinados. En el experimento 4 no usamos atributos de este tipo. En el experimento 2 definimos atributos que se generan a partir de la concatenación del valor actual (para cada línea) con valores anteriores y siguientes. Para el atributo categoría gramatical y los atributos con información sobre nombres y verbos de opinión se generan atributos combinados tomando dos, tres y cuatro valores anteriores y posteriores. Para los atributos restantes solamente se generan atributos combinando dos o tres valores.

Como se ve en la tabla, la utilización de atributos combinados afecta positivamente el valor de *precision* pero empeora el valor de *recall*.

10. Comparación y combinación de los dos sistemas

10.1 Comparación de los resultados de los dos sistemas

Para poder comparar los dos sistemas realizamos una nueva evaluación del sistema de reglas, esta vez solamente para el reconocimiento de fuentes, aplicándolo sobre el corpus de testeo utilizado en la evaluación del sistema basado en CRF.

Para esto, asumimos que todos los predicados de opinión son reconocidos, lo cual se resolvió aplicando el sistema de reglas con el repertorio de predicados de opinión enriquecido a partir del proceso de anotación manual del corpus. Esto pone en igualdad de condiciones a los dos sistemas ya que los corpus de entrenamiento y testeo tienen todos los predicados de opinión marcados como tales (como se aclaró en 9.2.1).

De esto se desprende una mejor recuperación de fuentes por parte del sistema de reglas en esta nueva evaluación en comparación con la evaluación del sistema completo (capítulo 8), en la cual los errores en el reconocimiento del predicado afectaron negativamente la recuperación de la fuente.

La tabla 10.1 muestra los resultados obtenidos para la identificación de fuentes por los dos sistemas (medidas exactas). Para el sistema basado en CRF retomamos dos experimentos diferentes, el experimento 2 que arroja el mejor valor de *precision* y el experimento 4 que da el mejor valor de *recall*.

	<i>Precision</i>	<i>Recall</i>	<i>F</i>
reglas	85%	73%	79%
CRF exp. 2	92%	65%	76%
CRF exp. 4	90%	66%	76%

Tabla 10.1: Comparación de los resultados de los dos sistemas.

Se observan diferencias importantes tanto en *precision*, medida para la cual los sistemas basados en CRF se desempeñan mejor, alcanzando 7 puntos más que el sistema de reglas, como en *recall*, para el cual el sistema de reglas logra 8 puntos de ventaja.

Estas diferencias motivaron los últimos experimentos, que mostramos en la sección siguiente, en donde combinamos los dos sistemas de modo de sacar provecho de las ventajas de cada uno. Intuitivamente, podemos pensar que combinando los dos sistemas es posible conservar las ventajas de cada uno: el buen desempeño del sistema basado en CRF en cuanto al valor de *precision* y la capacidad de lograr un mejor valor de *recall* del sistema basado en reglas.

10.2 Combinación de los dos sistemas

Realizamos dos experimentos que combinan la salida del sistema de reglas con la aplicación de *Conditinal Random Fields*.

Para esto, nos basamos en las variantes que arrojaron mejores resultados entre los experimentos 1 a 8 (ver tabla 8.2). Las variantes que dieron el mejor valor de *precision* son las del experimento 2: corpus de entrenamiento de 30.000 palabras, conjunto básico de atributos (AtrsBas), conjunto básico de templates y no compactación de grupos nominales. Las variantes que dieron mejores valores de *recall* son las del experimento 4, la diferencia con el experimento 2 es que no se utilizan combinaciones de atributos.

En los dos casos incorporamos información relacionada con la salida del sistema de reglas contextuales. Agregamos al conjunto básico de atributos un atributo que indica, para cada palabra del texto, si fue reconocida como fuente o como parte de una fuente por el sistema de reglas. Los valores de este atributo son: B-FU-R (*begin* fuente marcada por reglas), I-FU-R (*inside* fuente marcada por reglas) y O. Llamamos a este conjunto AtrsBas+FueReg (atributos básicos más fuentes marcadas por las reglas).

10.3 Resultados del sistema combinado

En la tabla 10.2 mostramos las características del experimento y los resultados.

Exp	Corpus	Atributos	Ventana	Combinados	Bigrama	<i>Precision</i>	<i>Recall</i>	F
9	C30	AtrsBas+FueReg	2-4	sí	sí	93	75	83
10	C30	AtrsBas+FueReg	2-4	no	sí	90	74	81

Tabla 10.2 Experimentos que incluyen como atributo la salida del sistema de reglas (sistema combinado)

Al incorporarse al corpus el atributo que refleja la salida del sistema de reglas, los resultados de *recall* mejoraron notoriamente. En el experimento 9 obtuvimos una mejora de unos 9 puntos en *recall* respecto al mejor valor anterior (experimento 4), y una mejora de 1 punto en *precision* respecto al mejor valor anterior (experimento 2).

Al contrario de lo que sucede con los experimentos 2 y 4, en este caso la eliminación de atributos combinados no aporta mejoras en ninguna de las dos medidas tomadas.

También mejoramos los resultados de las reglas, aumentando en 2 puntos el valor de *recall* y en 8 puntos el valor de *precision*.

Como suponíamos, logramos un sistema que combina las bondades de cada uno de los sistemas anteriores dando buenos resultados en *precision*, como CRF, y buenos resultados en *recall*, como las reglas. Incluso se alcanza uno o dos puntos más en cada medida respecto a los mejores valores de los sistemas originales.

En cuanto a la medida F, esta aumentó 4 puntos respecto al sistema de reglas y 7 puntos respecto al sistema CRF.

Mostramos en la tabla 10.3 algunos ejemplos en los cuales se observan las diferencias en las salidas de los tres sistemas. Como salida del sistema basado en CRF tomamos el experimento 4, que dio el mejor valor de *recall*.

En los ejemplos 1, 2 y 3 se comprueba que una de las ventajas del sistema basado en CRF respecto al sistema de reglas es la mayor flexibilidad para incluir en las fuentes diferentes elementos. Se trata de fuentes que las reglas reconocen en forma parcial por la forma en que se arman los grupos nominales candidatos a ser fuentes, que no permiten palabras como *por*, *para*, y dentro del grupo. El sistema basado en CRF logra armar los grupos completos en los ejemplos 1 y 2 y arma un grupo un poco más extenso en 3. El sistema combinado da el mismo resultado que el sistema basado solo en CRF, de modo que podemos pensar, al menos en estos casos, que la información errónea que se incorpora al incluir entre los atributos la salida de las reglas no afecta el resultado final.

En los ejemplos 4, 5 y 6, el sistema de reglas da resultados correctos que permiten que el sistema combinado pueda reconocer las fuentes correspondientes, a pesar de que el sistema basado solo en CRF no pudo hacerlo.

El ejemplo 7 muestra que, por más que el resultado del sistema combinado logró superar globalmente a los dos sistemas anteriores, algunos casos de reconocimiento exitoso por parte del sistema basado solo en CRF se pierden al aplicarse el sistema combinado.

Nº	texto original	salida reglas	salida CRF	salida CRF+reglas
1	Según [una denuncia efectuada por funcionarios del INAU], ...	Según [una denuncia efectuada] por funcionarios del INAU, ...	Según [una denuncia efectuada por funcionarios del INAU], ...	Según [una denuncia efectuada por funcionarios del INAU], ...
2	..., aseguró [el cardenal brasileño Claudio Hummes, prefecto de la Congregación para el Clero], al abrir ...	..., aseguró [el cardenal brasileño Claudio Hummes, prefecto de la Congregación para el Clero], al abrir ...	..., aseguró [el cardenal brasileño Claudio Hummes, prefecto de la Congregación para el Clero], al abrir ...	..., aseguró [el cardenal brasileño Claudio Hummes, prefecto de la Congregación para el Clero], al abrir ...
3	Según [el Instituto Brasileño de Geografía y Estadística (IBGE), que depende del Ministerio de Planificación], Brasil obtendrá ...	Según [el Instituto Brasileño de Geografía y Estadística] (IBGE), que depende del Ministerio de Planificación, Brasil obtendrá ...	Según [el Instituto Brasileño de Geografía y Estadística (IBGE)], que depende del Ministerio de Planificación, Brasil obtendrá ...	Según [el Instituto Brasileño de Geografía y Estadística (IBGE)], que depende del Ministerio de Planificación, Brasil obtendrá ...
4	En palabras del [economista Fernando Ribeiro] ...	En palabras del [economista Fernando Ribeiro] ...	En palabras del economista Fernando Ribeiro ...	En palabras del [economista Fernando Ribeiro] ...
5	Sobre este asunto [Mieres] dijo ...	Sobre este asunto [Mieres] dijo ...	Sobre este asunto Mieres dijo ...	Sobre este asunto [Mieres] dijo ...
6	el desinterés de [Brasil]	el desinterés de [Brasil]	el desinterés de Brasil	el desinterés de [Brasil]
7	[El ministro de Vivienda y Medio Ambiente, Carlos Colacce], dijo ...	El ministro de Vivienda y Medio Ambiente, Carlos Colacce, dijo ...	[El ministro de Vivienda y Medio Ambiente, Carlos Colacce], dijo ...	El ministro de Vivienda y Medio Ambiente, Carlos Colacce, dijo ...

Tabla 10.3: Ejemplos de fuentes marcadas en forma diferente por los tres sistemas (reglas, CRF, CRF+reglas)

10.4 Síntesis

Comparamos los dos sistemas en base a una nueva evaluación del sistema de reglas, esta vez solamente para el reconocimiento de fuentes, asumiendo que todos los predicados de opinión son reconocidos (como hicimos con el sistema CRF) y aplicándolo sobre el corpus de testeo utilizado para evaluar el sistema CRF. Observamos que el sistema de reglas da mejores resultados en *recall* (73% reglas, 66% CRF), mientras que el sistema CRF da mejores resultados en *precision* (92% CRF, 85% reglas).

Incorporamos la salida del sistema de reglas como atributo de entrada para el aprendizaje basado en CRF. Este atributo indica, para cada palabra, si esta es reconocida como fuente o como parte de una fuente por el sistema de reglas. De este modo logramos un sistema que combina las bondades de cada uno de los sistemas anteriores dando buenos resultados en *precision*, como CRF, y buenos resultados en *recall*, como las reglas. Incluso se alcanza uno o dos puntos más en cada medida respecto a los mejores valores de los sistemas originales. Esto redunda en una mejora en el valor de medida F (83%) respecto a los dos sistemas: 4 puntos respecto a las reglas y 7 puntos respecto al sistema CRF.

A partir de la observación de algunos ejemplos, comprobamos que una de las ventajas del sistema basado en CRF respecto al sistema de reglas es la mayor flexibilidad para incluir en las fuentes diferentes elementos. El sistema basado en CRF logra armar los grupos completos en algunos casos en los cuales las reglas dan resultados parciales. Por otro lado, el sistema de reglas da mejores resultados para los predicados nominales, que por su baja frecuencia en el corpus de entrenamiento no son bien tratados por el sistema CRF.

En general, el sistema combinado se comporta correctamente en los casos que son bien reconocidos por alguno de los sistemas originales. Sin embargo, existen algunos pocos casos en los cuales, si bien alguno de los sistemas originales acierta, el sistema combinado falla.

QUINTA PARTE: Conclusiones

11. Conclusiones

Nos propusimos estudiar y reconocer en forma automática las expresiones que transmiten opiniones de diferentes fuentes en los textos de prensa. En los análisis de corpus realizados observamos que estas expresiones son muy frecuentes en el tipo de texto con los que trabajamos. Un ejemplo claro de esto es el ejemplo (1.1), que reproducimos aquí como (11.1), en el cual se mencionan opiniones de 6 fuentes diferentes en 4 párrafos consecutivos de un texto.

(11.1) Tras el Consejo de Ministros de este lunes, el ministro de Economía Danilo Astori informó que el gobierno acatará el fallo de la Justicia, y que, una vez comunicados los fallos restantes, se estimará de qué manera obtener los recursos que se dejen de percibir.

Por su parte, el Partido Colorado presentará este martes en el Senado un proyecto de ley que deroga el tributo, y el Partido Nacional no descarta interpelar al titular de Economía.

El Directorio del Partido Nacional hizo público su rechazo a las manifestaciones del Partido Socialista sobre la ministra de la SCJ, Sara Bossio, a las que calificó de "un ataque absolutamente inusitado a la Justicia".

El diputado Javier García opinó que "es gravísimo. Cuando se pone en juego cierta sospecha sobre el honor y el trabajo de una jueza que tiene 40 años de actuación, se lo hace sin medir consecuencias, por el solo hecho de discrepar con un veredicto, nos parece de una gravedad muy importante".

Definimos la opinión como un objeto con 4 elementos: un predicado, una fuente, un asunto y un mensaje. En general, el predicado nuclea a los demás elementos, que suelen coincidir con algunos de sus argumentos.

Construimos un modelo que refleja las estructuras sintácticas en las que los predicados de opinión y sus argumentos de opinión pueden aparecer. El modelo da lugar a diferentes tipos, organizadas en una jerarquía, que contienen toda la información sintáctico-semántica de los predicados. El modelo se especificó por medio de Estructuras de Rasgos Tipificadas (*Typed Feature Structures*) por lo que puede ser integrado a una gramática basada en rasgos.

Creamos un repertorio de predicados de opinión que contiene 155 entradas. Cada entrada lleva asociado el tipo que le corresponde en el modelo definido. Además, para cada predicado se especifica información sobre las preposiciones que acepta, información que no puede incluirse en el modelo por ser específica de cada pieza léxica.

Desarrollamos tres sistemas informáticos. El primer sistema, basado en reglas contextuales, identifica las opiniones y, para cada opinión, su fuente, el asunto y el mensaje. Además, se determina la orientación semántica para los casos en los cuales el predicado, el asunto o la fuente no son neutros. Aunque no se trata de casos muy frecuentes, razón por la cual la mayor parte de las opiniones quedan con su orientación semántica sin determinar, es un primer paso en la resolución de este aspecto que es de una complejidad importante.

El sistema da buenos resultados, considerando una medida laxa de evaluación, es decir, contando como elementos bien identificados los segmentos que se solapan con los elementos buscados. Los valores de medida F obtenidos son: para el predicado 92%, para la fuente 81%, para el asunto 75%, para el mensaje 89% y para la opinión completa 85%. Una de las

conclusiones que se desprende de esta evaluación es la adecuación del repertorio de predicados, cuya cobertura es de 91%²⁵. Si bien este valor es mejorable, consideramos de todos modos que es muy satisfactorio.

En cuanto a la determinación de la orientación semántica, no obtuvimos resultados definitorios dado que, en el corpus de evaluación, encontramos pocas ocurrencias de los elementos que tuvimos en cuenta para este cálculo: predicados de opinión no neutros, introductores de asunto no neutros y palabras subjetivas particulares dentro de la fuente. De todos modos, consideramos que el procedimiento para el cálculo de la orientación semántica que proponemos, en función de la orientación de cada elemento de la opinión, puede resultar apropiado y debe ser evaluado en un corpus mayor.

El resultado del sistema de reglas fue re-calculado para el reconocimiento de las fuentes, considerando como reconocidos todos los predicados. Obtuvimos un 79% de medida F exacta, es decir, calculada sin incluir solapamientos. Esta segunda evaluación fue realizada sobre el mismo corpus de testeo utilizado para los dos sistemas que se describen a continuación, de este modo, pudimos comparar los diferentes resultados obtenidos para la fuente.

El segundo sistema desarrollado está basado en el modelo *Conditional Random Fields* (CRF) y se aplicó solamente para el reconocimiento de las fuentes, asumiendo como conocidos los predicados. Trabajamos con un corpus de entrenamiento anotado en forma manual. Para la evaluación de este sistema utilizamos una herramienta ya existente que solo calcula los valores exactos, es decir, sin considerar solapamientos. Obtuvimos un valor de medida F exacta de 76%. El desarrollo de este sistema deja como resultado, además del clasificador para el reconocimiento de fuentes, un corpus de 40.000 *tokens* anotado con las fuentes.

El tercer sistema combina la salida de las reglas, solo para el reconocimiento de fuentes, con el sistema basado en CRF. Se agrega a este último sistema un atributo que indica, para cada palabra, si esta pertenece o no a alguna fuente reconocida por las reglas. El sistema combinado mejora notoriamente los resultados de los dos sistemas anteriores, alcanzando un valor de medida F exacta de 83%. Respecto al sistema de reglas, se supera el resultado en 4 puntos y respecto al sistema CRF se mejora en 7 puntos.

Consideramos que los resultados obtenidos para la identificación de fuentes (83% de medida F exacta), en particular, son muy satisfactorios, tomando como referencia los valores de medida F reportados en trabajos aplicados a otros idiomas. En la tabla 11.1 (reproducción de la tabla 1.1) se muestran esos valores. Es importante aclarar que los trabajos mencionados en la tabla presentan varias diferencias respecto al nuestro, por lo que no son del todo comparables y solo se muestran a modo de referencia.

El único trabajo que logra un resultado mejor que el nuestro es el de Bethard et al. (2004), pero el tipo de expresiones que allí se consideran es más limitado, ya que solamente se trabaja con opiniones proposicionales (opiniones expresadas dentro una subordinada a un verbo de opinión, como *X dijo que ...*) y, además, se buscan las fuentes solamente para las opiniones de contenido subjetivo. Choi et al (2006) logran un resultado del mismo orden que el nuestro, pero se trata de una medida con solapamientos.

25 Medida calculada con una versión del repertorio anterior a la actual. Posteriormente a la evaluación que se presenta, el repertorio fue ampliado en un 30%.

Referencia	Descripción	medida F	Tipo de medida
Bethard et al., 2004	Inglés, opiniones proposicionales (<i>X dijo que ..</i>)	89,50%	no especificado
Choi et al., 2005	Inglés	69,00%	con solapamientos
Choi et al., 2006	Inglés	82,00%	con solapamientos
Wiegand et al., 2010	Inglés	62,60%	exacta
Lu, 2010	Chino	78,00%	con solapamientos
Nuestra propuesta	Español	83,00%	exacta

Tabla 11.1: Resultados obtenidos por diferentes sistemas para la identificación de fuentes.

Los resultados de nuestro tercer sistema son muy alentadores en cuanto a la conveniencia de combinar técnicas simbólicas y estadísticas. Por un lado, se mejora la *precision* respecto al sistema basado en reglas, para el cual habíamos obtenido un valor 7 puntos más bajo que para el sistema CRF. Por otro lado, se mejora el *recall* respecto al sistema CRF, que fue 7 puntos peor que el obtenido por las reglas. Podemos pensar que al combinar los dos sistemas logramos mantener las bondades de cada uno de ellos, mejorando así la medida F final.

Es importante destacar que, además de los sistemas para el reconocimiento de la opinión y sus elementos, se generó un conjunto de recursos importantes para el español: un repertorio de predicados de opinión con información sintáctico-semántica asociada, un corpus de 13.000 palabras anotado con las opiniones y sus elementos (predicado, fuente, asunto y mensaje) y un corpus de 40.000 palabras anotado con predicados de opinión y fuentes.

El estudio teórico realizado y las herramientas y recursos desarrollados constituyen un importante aporte para el desarrollo del Procesamiento de Lenguaje Natural aplicado a textos en español, en particular, para el área Minería de Opiniones en donde hay muy pocos trabajos orientados a este idioma.

12. Extensiones y trabajos futuros

Durante el año 2010 trabajamos con un grupo de estudiantes de grado en una extensión a la propuesta que aquí presentamos: la resolución de correferencias para fuentes (Acerenza et al, 2010). En este trabajo se detectan las fuentes que son correferenciales y se crean las cadenas de correferencias correspondientes. Se incluye, como caso particular de correferencias, la resolución de fuentes anafóricas y de fuentes omitidas, fenómeno muy común en español. Una evaluación primaria del sistema arroja los siguientes resultados: 83% de *precision* y 85% de *recall*.

Otra línea en la que estamos trabajando, con estudiantes de posgrado, es la determinación de la orientación semántica del mensaje de cada opinión. En este trabajo, que recién está en sus inicios, se trabajará sobre el corpus anotado con las opiniones y sus elementos, obtenido a partir de la corrección manual de la salida del sistema de reglas aplicado sobre el corpus de evaluación. Se explorarán métodos de aprendizaje automático. Una de las opciones que se están estudiando es la realización de una etapa previa de aprendizaje de palabras de contenido subjetivo para la creación de un diccionario. También se prevé analizar modificadores de subjetividad como inversores, intensificadores, etc.

Además de las extensiones a nuestro trabajo ya mencionadas, quedan diversos aspectos por resolver, que consideramos de gran importancia para completar nuestro estudio:

- Trabajar en la identificación del asunto de cada opinión, independientemente de que exista un segmento específico en el cual este es introducido (como *en lo que respecta a xxx*). Esto implica un estudio del contenido del mensaje para deducir cuál es el asunto sobre el cual se habla.
- Extender el repertorio de predicados de opinión. El conjunto de verbos y nombres debe seguir actualizándose mediante la incorporación de nuevas entradas y, quizá, la exclusión de algunas que resulten demasiado ambiguas. En este sentido, sería interesante trabajar en la desambiguación de los predicados de opinión para mejorar el 92% de *precision* alcanzado en su identificación. Además, consideramos de gran importancia incorporar adjetivos al repertorio, de modo de poder reconocer opiniones centradas en este tipo de palabras, apoyadas en verbos de estado. Se trata de expresiones como *estar/parecer/mostrarse/verse + feliz/molesto/satisfecho/contento/conforme/contrariado*.
- Analizar la incidencia del estudio de la factividad de los eventos sobre los predicados de opinión. Para este tipo de predicados, la negación puede implicar en muchos casos cambios en los valores de orientación semántica, por ejemplo, si decimos *X no rechaza la propuesta*, además de saber que se menciona un evento no realizado (el evento *rechazar la propuesta*), podemos deducir una orientación semántica positiva, quizá débil, de la fuente hacia el asunto. Esto nos lleva a la necesidad de contemplar diferentes valores de orientación semántica que tengan en consideración diferentes matices de apoyo o rechazo hacia el asunto.

BIBLIOGRAFÍA

- Acerenza, Fernando, Macarena Rabosto y Magdalena Zubizarreta. 2010. *Resolución de correferencias en expresiones de opinión*. Proyecto de grado. Facultad de Ingeniería, Universidad de la República. Montevideo.
- Albertuz, Francisco. 2007. *Sintaxis, semántica y clases de verbos: Clasificación verbal en el proyecto ADESSE*. En *Actas del VI Congreso de Lingüística General, Santiago de Compostela*, Vol. 2, Tomo 2, *Las lenguas y su estructura* (IIb), pp. 2015-2030.
- Asher, Nicholas, Farah Benamara e Yvette Y. Mathieu. 2008. *Distilling Opinion in Discourse: A Preliminary Study*. En *Coling 2008: Companion volume: Posters and Demonstrations*, pages 5--8, Manchester.
- Asher, Nicholas, Farah Benamara e Yvette Y. Mathieu. 2009. *Appraisal of Opinion Expressions in Discourse*. En *Linguisticae Investigationes*, 32:2. John Benjamins Publishing Company.
- Atserias, J., B. Casas, E. Comelles, M. González, L. Padró y M. Padró. 2006. *FreeLing 1.3: Syntactic and semantic services in an open-source NLP library*. En *Proceedings of the fifth international conference on Language Resources and Evaluation (LREC) ELRA*.
- Banea, Carmen, Rada Mihalcea, Janyce Wiebe, Samer Hassan. 2008. *Multilingual Subjectivity Analysis Using Machine Translation*. Conference on Empirical Methods in Natural Language Processing (EMNLP).
- Baker, Collin F., Charles J. Fillmore y John B. Lowe. 1998. *The Berkeley FrameNet project*. En *Proceedings of the COLING-ACL, Montreal, Canada*.
- Bethard, Steven, Hong Yu, Ashley Thornton, Vasileios Hatzivassiloglou y Dan Jurafsky. 2004. *Automatic Extraction of Opinion Propositions and their Holders*. *Proceedings of AAAI Spring Symposium on Exploring Attitude and Affect in Text*.
- Bethard, Steven, Hong Yu, Ashley Thornton, Vasileios Hatzivassiloglou y Dan Jurafsky. 2005. *Extracting opinion propositions and opinion holders using syntactic and lexical cues*. En James G. En Shanahan, Yan Qu y Janyce Wiebe, editores, *Computing Attitude and Affect in Text: Theory and Applications*. Springer 2005.
- Bosque, Ignacio y Violeta Demonte. 1999. *Gramática Descriptiva de la Lengua Española*. Real Academia Española, Colección Nebrija y Bello. Espasa. Madrid.
- Bradley, M. M. y P. J. Lang. 1999. *Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings*. Reporte técnico C-1. The Center for Research in Psychophysiology. University of Florida.
- Brooke, J., M. Tofiloski and M. Taboada. 2009. *Cross-Linguistic Sentiment Analysis: From English to Spanish*. RANLP 2009, Recent Advances in Natural Language Processing. Borovets, Bulgaria.
- Carpenter, Bob. 1992. *The Logic of Typed Feature Structure*. Cambridge Tracts in Theoretical Computer Science no. 32. Cambridge University Press. New York.
- Choi, Yejin y Claire Cardie, E. Riloff y S Patwardhan. 2005. *Identifying Sources of Opinions with Conditional Random Fields and Extraction Patterns*. En *HLT-EMNLP*.

- Choi, Yejin, Eric Breck y Claire Cardie. 2006. *Joint Extraction of Entities and Relations for Opinion Recognition*. En Empirical Methods in Natural Language Processing (EMNLP).
- Choi, Yejin y Claire Cardie. 2008. *Learning with Compositional Semantics as Structural Inference for Subsentential Sentiment Analysis*. En Empirical Methods in Natural Language Processing (EMNLP).
- Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton.
- Chomsky, Noam. 1964. *Aspects of the Theory of Syntax*. Cambridge: The MIT Press.
- Chomsky, Noam. 1981. *Lectures on Government and Binding*. Holland: Foris Publications.
- Copestake, Ann. 1992. *The ACQUILEX LKB: Representation Issues in the Semi-automatic Acquisition of Large Lexicons*. En Proceedings of the 3rd Conference on Applied Natural Language Processing, Trento, Italy.
- Copestake, Ann. 2002. *Implementing Typed Feature Structure Grammars*. CSLI Publications. Stanford.
- Davies, Mark. 2002. *Corpus del español (100 millones de palabras, siglo XIII - siglo XX)*. Disponible actualmente en <http://www.corpusdelespanol.org>.
- De Bruyne, Jacques. 1999. *Las preposiciones*. En Gramática Descriptiva de la Lengua Española (*Sintaxis básica de las clases de palabras*), 657-703.
- Esuli, Andrea y Fabrizio Sebastiani. 2006. *Determining term subjectivity and term orientation for opinion mining*. En Proceedings of the 11th Meeting of the European Chapter of the Association for Computational Linguistics (EACL-2006), pages 193–200. Trento.
- Fellbaum, Christiane. 1998. *WordNet: An Electronic Lexical Database*. MIT Press. Cambridge.
- Fellbaum, Christiane. 1990. *English verbs as a semantic net*. En *International Journal of Lexicography* 3 (4), pp. 278 - 301.
- Fillmore, Charles. 1968. *The Case for Case*. En Emmon Bach y R.T.Harms (eds.): *Universals in Linguistic Theory*; Chicago: Holt, Rinehart & Winston.
- García, David y Francesc Alías. 2008. *Identificación de emociones a partir de texto usando desambiguación semántica*. En Procesamiento del Lenguaje Natural, Revista nº 40, marzo de 2008, pp. 75-82.
- García-Miguel, José M. y Francisco J. Albertuz. 2005. *Verbs, Semantic Classes and Semantic Roles in the ADESSE project*. En Erk, Katrin, Alissa Melinger & Sabine Schulte im Walde (eds): *Proceedings of the Interdisciplinary Workshop on the Identification and Representation of Verb Features and Verb Classes*. Saarbrücken.
- García-Miguel, José M., Lourdes Costas y Susana Martínez. 2005. *Diátesis verbales y esquemas construccionales. Verbos, clases semánticas y esquemas sintáctico-semánticos en el proyecto ADESSE*. En Wotjak, Gerd, & Juan Cuartero Otal, eds. *Entre semántica léxica, teoría del léxico y sintaxis*. Frankfurt am Main: Peter Lang, pp. 373-384.
- Grassi, Mariela, Marisa Malcuori, Javier Couto, Juan José Prada y Dina Wonsever. 2001. *Corpus*

- informatizado: textos del español del Uruguay (CORIN)*. SLPLT-2 - Second International Workshop on Spanish Language Processing and Language Technologies. Jaén, España.
- Gross, Maurice. 1975. *Méthodes en syntaxe*. Paris: Hermann.
- Gruber, Jeffrey S. 1965. *Studies in Lexical Relations*. Tesis de Doctorado. MIT. Massachusetts.
- Halliday, M.A.K. 1985. *An introduction to functional grammar*. London; Baltimore, Md., USA : Edward Arnold.
- Hatzivassiloglou, Vasileios and Kathy McKeown. 1997. Predicting the semantic orientation of adjectives. In Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics (ACL-97), pages 174–181. Madrid.
- Indurkha, Nitin y Fred J. Damerau. 2010. *Handbook of Natural Language Processing*. Segunda edición. CRC Press.
- Kim, Soo-Min y Eduard Hovy. 2004. *Determining the Sentiment of Opinions*. En Proceedings of the COLING conference. Geneva.
- Kim, Soo-Min y Eduard Hovy. 2005. *Identifying Opinion Holders for Question Answering in Opinion Texts*. Proceedings of the Workshop on Question Answering in Restricted Domain at the Conference of the American Association of Artificial Intelligence (AAAI-05). Pittsburgh, PA.
- Kim, Soo-Min y Eduard Hovy. 2006. *Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Texts*. En Proceedings of the Workshop on Sentiment and Subjectivity in Text (SST'06), Association for Computational Linguistics. Stroudsburg, PA.
- Kipper-Schuler, Karin. 2005. *VerbNet: A broad-coverage, comprehensive verb lexicon*. PhD. Thesis. Computer and Information Science Dept., University of Pennsylvania. Philadelphia, PA.
- Kobayashi, Nozomi, Ryu Iida, Kentaro Inui y Yuji Matsumoto. 2005. *Opinion extraction using a learning-based anaphora resolution technique*. En Second International Joint Conference on Natural Language Processing (IJCNLP).
- Krestel, Ralf, Sabine Bergler y René Witte. 2008. *Minding the Source: Automatic Tagging of Reported Speech in Newspaper Articles*. Proceedings of the Sixth International Language Resources and Evaluation Conference (LREC 2008). Marruecos.
- Lafferty, J., A. McCallum y F. Pereira. 2001. *Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data*. En Proc. of ICML, pp.282-289.
- Levin, Beth. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. University of Chicago Press. Chicago.
- Liu, Bing. 2010. *Sentiment Analysis and Subjectivity*. En *Handbook of Natural Language Processing*, Second Edition (Indurkha y Damerau).
- Lu, Bin. 2010. *Identifying Opinion Holders and Targets with Dependency Parser in Chinese News Texts*. En Proceedings of the NAACL HLT 2010 Student Research Workshop, pages 46–51. Los Angeles, California.

- Lu, Bin, K. T. Benjamin y J. Tao. 2010. *Supervised Approaches and Dependency Parsing for Chinese Opinion Analysis at NTCIR-8*. En Proceedings of NTCIR-8 Workshop Meeting. Tokyo, Japón.
- Maldonado, Concepción. 1999. *Discurso directo y discurso indirecto*. En Gramática Descriptiva de la Lengua Española (*Entre la oración y el discurso*), 3549-3596.
- Martin, J. R. y P. R. R. White. 2005. *The Language of Evaluation, Appraisal in English*. Palgrave Macmillan. Londres y Nueva York.
- Mathieu, Yvette Yannick. 2000. *Les verbes de sentiments. De l'analyse linguistique au traitement automatique*. CNRS Editions. Paris.
- Mathieu, Yvette Yannick. 2009. *Sentiments et Opinions: deux filtres classifieurs*. En Filtrage sémantique: de l'annotation à la navigation textuelle, Jean-Luc Minel (ed). Hermès Sciences, Lavoisier. Paris.
- Mesa, Andrea. 2010. *Modelos markovianos para secuencias y aplicaciones a la predicción de genes*. Tesis de Maestría. Facultad de Ingeniería, Universidad de la República. Montevideo.
- Miller, George A., Richard Beckwith, Christiane Fellbaum, Derek Gross y Katherine Miller. 1990. *Introduction to WordNet: An On-line Lexical Database*. En *International Journal of Lexicography* 3 (4), pp. 235 - 244.
- Moilanen, Karo y Stephen Pulman. 2007. *Sentiment Composition*. En RANLP.
- Moreno Ortiz, Antonio, Francisco Pineda Castillo y Rodrigo Hidalgo García. 2010. *Análisis de Valoraciones de Usuario de Hoteles con Sentitext*: un sistema de análisis de sentimiento independiente del dominio*. En Procesamiento del Lenguaje Natural, Revista nº45, septiembre 2010, pp 31-39.
- Nakagawa, Tetsuji, Kentaro Inui y Sadao Kurohashi. 2010. *Dependency Tree-based Sentiment Classification using CRFs with Hidden Variables*. En Proceedings of the NAACL HLT 2010, pp 786–794. Los Angeles, California.
- Palmer, Martha, Daniel Gildea y Paul Kingsbury. 2005. *The Proposition Bank: An Annotated Corpus of Semantic Roles*. En Computational Linguistics, Vol. 31, 1.
- Palmer, Martha. 2009. *Semlink: Linking PropBank, VerbNet and FrameNet*. En Proceedings of the Generative Lexicon Conference (GenLex-09). Pisa, Italy.
- Pang, Bo, Lillian Lee, Shivakumar Vaithyanathan. 2002. *Thumbs up? Sentiment Classification using Machine Learning Techniques*. En Proceedings of EMNLP.
- Pang, Bo y Lillian Lee. 2008. *Opinion Mining and Sentiment Analysis*. En Foundations and Trends in Information Retrieval, vol. 2, nº 1-2 (2008), 1-135.
- Pavón Lucero, M. Victoria. 1999. *Clases de partículas: preposición, conjunción y adverbio*. En Gramática Descriptiva de la Lengua Española (*Sintaxis básica de las clases de palabras*), 565-653.
- Picallo, M. Carme. 1999. *La estructura del sintagma nominal: las nominalizaciones y otros*

- sustantivos con complementos argumentales*. En Gramática Descriptiva de la Lengua Española (*Sintaxis básica de las clases de palabras*), 363-393.
- Polanyi, Livia y Annie Zaenen. 2004. *Contextual Valence Shifters*. En AAAI spring Symposium on Attitude.
- Pollard, Carl e Iván a. Sag. 1987. *Information-Based Syntax and Semantics, Volume 1: Fundamentals*. CSLI Lecture Notes no. 13. Stanford: Center for the Study of Language and Information. University of Chicago Press.
- Pollard, Carl e Ivan a. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Stanford: Center for the Study of Language and Information. University of Chicago Press.
- Pouliquen, Bruno, Ralf Steinberger y Clive Best. 2007. *Automatic Detection of Quotations in Multilingual News*. En Proceedings of the International Conference Recent Advances in Natural Language Processing (RANLP 2007).
- Read, Jonathon, Davis Hope y John Carroll. 2007. *Annotating Expressions of Appraisal in English*. En Proceedings of the Linguistic Annotation Workshop, Association for Computational Linguistics.
- Redondo, Jaime, Isabel Fraga, Isabel Padrón y Montserrat Comesaña. 2007. *The Spanish Adaptation of ANEW (Affective Norms for English Words)*. En Behavior Research Methods.
- Rei, Marek y Ted Briscoe. 2010. Combining manual rules and supervised learning for hedge cue and scope detection. En CoNLL '10: Shared Task Proceedings of the Fourteenth Conference on Computational Natural Language Learning. Association for Computational Linguistics Stroudsburg, PA, USA.
- Riloff, Ellen, Janyce Wiebe y Theresa Wilson. 2003. *Learning subjective nouns using extraction pattern bootstrapping*. En Proceedings of the 7th Conference on Natural Language Learning (CoNLL-2003), pages 25–32, Edmonton, Canada.
- Rosá, Aiala. 2008. *Identification automatique de marques d'opinion dans des textes*, Actas de JEP-TALN-RECITAL '08. Avignon.
- Rosá, Aiala, Dina Wonsever y Jean-Luc Minel. 2010. *Opinion Identification in Spanish Texts*. En Proceedings of NAACL HLT 2010, Workshop on Computational Approaches to Languages of the Americas. Los Angeles, California.
- Rosá, Aiala, Dina Wonsever y Jean-Luc Minel. 2010. *Comparación de dos métodos para la extracción de opiniones en textos en español*. En Proceedings of IBERAMIA 2010, Workshop on Natural Language Processing and Web-based technologies. Bahía Blanca.
- Ruppenhofer, Josef, Swapna Somasundaran y Janyce Wiebe. 2008. *Finding the Sources and Targets of Subjective Expressions*. En Proceedings of LREC 2008.
- Sag, Ivan A., Thomas Wasow y Emily M. Bender. 2003. *Syntactic Theory. A Formal Introduction*. CSLI Publications, 2nd. ed. Stanford.
- Saurí, Roser. 2008. *A Factuality Profiler for Eventualities in Text*. Tesis de Doctorado. Brandeis University.

- Stoyanov, Vaseelin, Claire Cardie y Janyce Wiebe. 2005. *Multi-Perspective Question Answering Using the OpQA Corpus*. En Proceedings of HTL-EMNLP 2005.
- Stoyanov, Veselin y Claire Cardie. 2008. *Topic Identification for Fine-Grained Opinion Analysis*. En Proceedings of COLING 2008.
- Subirats, Carlos. 2009. *FrameNet Español: un análisis cognitivo del léxico del español*. En A. Alcina, E. Valero y E. Rambla, eds., *Terminología y Sociedad del conocimiento*. Frankfurt a. M.: Peter Lang, pp. 309-320.
- Sutton, Charles y Andrew McCallum. 2010. *An Introduction to Conditional Random Fields*. Cornell University Library.
- Taulé, Mariona. 1995. *Representación de las entradas verbales en una base de conocimiento léxico: diátesis y semántica léxica*. Tesis de Doctorado. Universidad de Barcelona.
- Turney, Peter. 2002. *Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews*. En Proceedings of the Association for Computational Linguistics 40th Anniversary Meeting. Association for Computational Linguistics, New Brunswick, NJ.
- Turney, Peter y Michael Littman. 2003. *Measuring praise and criticism: Inference of semantic orientation from association*. En ACM Transactions on Information Systems (TOIS).
- Vázquez, Gloria, Ana Fernández y Ma. Antonia Martí. 2000. *Clasificación verbal. Alternancias de diátesis*. Universitat de Lleida. Lérida.
- Wallach, H.M. 2004. *Conditional random fields: An introduction*. Technical report MS-CIS-04-21, University of Pennsylvania.
- Wiebe, Janyce, Theresa Wilson, Rebecca Bruce, Matthew Bell y Melanie Martin. 2004. *Learning subjective language*. En *Computational Linguistics* 30 (3).
- Wiebe, Janyce, Theresa Wilson y Claire Cardie. 2005. *Annotating expressions of opinions and emotions in language*. Language Resources and Evaluation, volume 39, issue 2-3, pp. 165-210.
- Wiegand, Michael y Dietrich Klakow. 2010. *Convolution Kernels for Opinion Holder Extraction*. En Proceedings of HLT-NAACL'10. Los Angeles, California.
- Wilson, Theresa, Janyce Wiebe y Paul Hoffmann. 2005. *Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis*. En HLT-EMNLP.
- Wilson, Theresa, Janyce Wiebe y Rebecca Hwa. 2006. *Recognizing strong and weak opinion clauses*. Computational Intelligence 22 (2): 73-99.
- Wonsever, D., J.-L. Minel J.-L. 2001. *Contextual Rules for Text Analysis*. En Lecture Notes in Computer Science.
- Wonsever, D., S. Caviglia, J. Couto y A. Rosá. 2006. Un sistema para la segmentación en proposiciones de textos en español. En Letras de hoje 144 (41).

Referencias en Internet

Web Aquilex: <http://www.cl.cam.ac.uk/research/nl/aquilex/>

Web ADESSE: <http://adesse.uvigo.es/>

Web Appinions: appinions.com

Web AppinionsPrism: www.appinions.com/apps/prism/index.vm

Web CoNLL2000: www.cnts.ua.ac.be/conll2000/chunking/output.html

Web CRF++: <http://crfpp.sourceforge.net/>

Web DRAE: www.rae.es

Web EAGLES: <http://www.ilc.cnr.it/EAGLES96/home.html>

Web FrameNet: <http://framenet.icsi.berkeley.edu/>

Web FreeLing1: <http://nlp.lsi.upc.edu/freeling/>

Web FreeLing2: <http://nlp.lsi.upc.edu/freeling/doc/userman/parole-es.html>

Web Gate: <http://gate.ac.uk/>

Web MPQA: <http://www.cs.pitt.edu/mpqa/>

Web NewsExplorer: <http://emm.newsexplorer.eu>

Web NTCIR-8: <http://research.nii.ac.jp/ntcir/ntcir-ws8/ws-en.html>

Web Spanish FrameNet: <http://gemini.uab.es:9080/SFNsite>

ANEXOS

Anexo A: Ejemplos seleccionados de corpus

Ejemplos seleccionados de los corpus de desarrollo.

- (E1) Consultado sobre la lentitud de los procesos judiciales uruguayos Carranza respondió: "Hay una situación de un muy alto número de presos sin condena, hay que agilizar los procesos".
- (E2) El vicepresidente del Instituto Nacional de Carnes de Uruguay, Fernando Pérez Abella, dijo que las conclusiones son muy alentadoras.
- (E3) El 56% de los estadounidenses cree que la invasión de Irak en 2003 fue un error.
- (E4) No existe más durante este gobierno el vocero presidencial como figura de estructura presidencial, afirmó Menem.
- (E5) La iglesia católica tomó partido durante el proceso revolucionario y apoyó a Victoriano Huerta.
- (E6) Mientras que otros sectores como el Partido Comunista, Asamblea Uruguay, el Partido Socialista y Alianza Progresista rechazaron la propuesta.
- (E7) Clinton formuló su declaración durante rueda de prensa gráfica junto con el primer ministro irlandés Bertie Ahern.
- (E8) Nosotros coordinamos esta relación con la Comisión Nacional Bancaria y tenemos mucho apoyo de la Comisión.
- (E9) Este sistema se utiliza en Estados Unidos desde 1982, según Roque Pifarré.
- (E10) De acuerdo a un informe de la Dirección de Prensa de Palacio de Gobierno, la reunión organizada por el Ministerio de la Presidencia se realizará en el Hotel París desde las primeras horas de hoy.
- (E11) Entrevistado por el semanario Crónicas, el abogado Gustavo Salle habló del pedido de extradición de Henry Kinssinger, sus discrepancias con el Gobierno del Frente Amplio, el caso Young y el trabajo del Secretario de Presidencia, Gonzalo Fernández durante la dictadura. "Este Frente Amplio vende una imagen, escenifica determinada postura ideológica y en la práctica edifica otra totalmente opuesta. Se vende como socialista y es ortodoxamente neoliberal. Si yo hubiera estado preso y hubiese sido torturado y hoy tuviera que enfrentar con esa pesada carga existencial y afectiva, la traición política e ideológica que está consumando la cúpula del Frente Amplio no sé cuál sería mi reacción".
- (E12) Mahuad dijo que su país no rechazaba la idea de crear un parque binacional en la zona en disputa.
- (E13) La Alianza Nacional de Agricultores Independientes de Cuba se pronuncia en contra de tales prácticas de corrupción.
- (E14) Los optimistas sostienen que la probabilidad de que esto suceda es muy baja pues el comportamiento que han mostrado los embalses hasta el momento es un indicador de que San Pedro será benévolo y los niveles de agua no llegarán a extremos.

Ejemplos seleccionados, con los elementos de la opinión marcados del siguiente modo: predicado en negrita, fuente subrayada, asunto en cursiva y mensaje con fondo gris.

- (E1) Consultado *sobre la lentitud de los procesos judiciales uruguayos* Carranza **respondió:** "Hay una situación de un muy alto número de presos sin condena, hay que agilizar los procesos".
- (E2) El vicepresidente del Instituto Nacional de Carnes de Uruguay, Fernando Pérez Abella, **dijo** que las conclusiones son muy alentadoras.
- (E3) El 56% de los estadounidenses **creo** que la invasión de Irak en 2003 fue un error.
- (E4) No existe más durante este gobierno el vocero presidencial como figura de estructura presidencial, **afirmó** Menem.
- (E5) La iglesia católica tomó partido durante el proceso revolucionario y **apoyó a** *Victoriano Huerta*.
- (E6) Mientras que otros sectores como el Partido Comunista, Asamblea Uruguay, el Partido Socialista y Alianza Progresista **rechazaron** *la propuesta*.
- (E7) Clinton formuló su **declaración** durante rueda de prensa gráfica junto con el primer ministro irlandés Bertie Ahern.
- (E8) Nosotros coordinamos esta relación con la Comisión Nacional Bancaria y tenemos mucho **apoyo de la Comisión**.
- (E9) Este sistema se utiliza en Estados Unidos desde 1982, **según** Roque Pifarré.
- (E10) **De acuerdo a un informe de la Dirección de Prensa de Palacio de Gobierno**, la reunión organizada por el Ministerio de la Presidencia se realizará en el Hotel París desde las primeras horas de hoy.
- (E11) Entrevistado por el semanario Crónicas, el abogado Gustavo Salle **habló del pedido de extradición de Henry Kinssinger, sus discrepancias con el Gobierno del Frente Amplio, el caso Young y el trabajo del Secretario de Presidencia, Gonzalo Fernández durante la dictadura**. "Este Frente Amplio vende una imagen, escenifica determinada postura ideológica y en la práctica edifica otra totalmente opuesta. Se vende como socialista y es ortodoxamente neoliberal. Si yo hubiera estado preso y hubiese sido torturado y hoy tuviera que enfrentar con esa pesada carga existencial y afectiva, la traición política e ideológica que está consumando la cúpula del Frente Amplio no sé cuál sería mi reacción".
- (E12) Mahud **dijo** que su país **no rechazaba** *la idea de crear un parque binacional en la zona en disputa*.
- (E13) La Alianza Nacional de Agricultores Independientes de Cuba **se pronuncia** *en contra de tales prácticas de corrupción*.
- (E14) Los optimistas **sostienen** que la probabilidad de que esto suceda es muy baja pues el comportamiento que han mostrado los embalses hasta el momento es un indicador de que San Pedro será benévolo y los niveles de agua no llegarán a extremos.

Anexo B: Repertorio de predicados de opinión

Mostramos el repertorio completo de predicados de opinión, en primer lugar los verbos y luego los nombres. Para simplificar la revisión del repertorio, ordenamos las piezas léxicas por tipo y luego por orden alfabético.

En cuanto a la información contenida en las tablas, tanto para los verbos como para los nombres se da una lista (representada por corchetes) de preposiciones aceptadas como introductoras de asunto, que puede ser una lista vacía. Si se incluye la preposición *sobre*, se asume que también se aceptan otras locuciones típicas introductoras de asunto como *en lo que respecta a*, *en cuanto a*, etc.

Para los verbos, en particular, se indica con el símbolo Φ la aceptación de asunto sin preposición (es decir, un grupo nominal con función complemento directo) y por medio de la palabra 'prop' se indica la aceptación de una proposición como asunto.

Para los nombres se indica, además de la lista de preposiciones que introducen el asunto, el rol semántico que debe asignarse al complemento con *de*. En algunos pocos casos este complemento corresponde siempre al asunto (*la difusión de material prohibido*). En la mayoría de los casos o bien se trata siempre de la fuente (*la afirmación del senador*), o bien el rol es ambiguo (*el planteo del senador / el planteo de la formación de una comisión*).

B.1 Verbos de opinión

B.1.1 Verbos de discurso referido

Todos los verbos de esta tabla tienen asociados los tipos: vo_sf_dref, vo_sf_cree_com y vo_sf.

verbo de opinión	forma asunto
citar	[Φ , a]
hablar	[de, sobre]
manifestarse	[sobre]
pronunciarse	[sobre]
referirse	[a]

B.1.2 Verbos de discurso reproducido y creencia

Todos los verbos de esta tabla tienen asociados los tipos: vo_sf_drep_cree, vo_sf_cree_com y vo_sf.

verbo de opinión	forma asunto
aclarar	[Φ]
adelantar	[Φ]
admitir	[Φ]
advertir	[sobre]
afirmar	[]
agregar	[]
alegar	[]
anunciar	[Φ]
añadir	[]
apuntar	[]
argumentar	[sobre]
asegurar	[]
aseverar	[]
atestiguar	[]
calcular	[]
comentar	[sobre]
comunicar	[Φ]
confesar	[Φ]
confirmar	[Φ]
considerar	[]
consignar	[]
contar	[de]
continuar	[]
creer	[]
decir	[Φ, de, sobre]
declarar	[sobre]
destacar	[Φ]
detallar	[Φ]
difundir	[Φ]
entender	[]
escribir	[sobre, de]
especular	[sobre]
estimar	[]

verbo de opinión	forma asunto
explicar	[Φ]
explicitar	[Φ]
exponer	[Φ]
expresar	[Φ]
indicar	[]
informar	[]
insistir	[]
manifestar	[Φ]
mencionar	[Φ]
observar	[]
opinar	[sobre]
plantear	[Φ]
precisar	[Φ]
presumir	[sobre]
prever	[Φ]
publicar	[Φ, sobre]
reafirmar	[Φ]
recalcar	[Φ]
reclamar	[Φ]
recordar	[Φ]
reiterar	[Φ]
relatar	[Φ]
remarcar	[Φ]
replicar	[]
responder	[Φ, sobre]
revelar	[Φ]
señalar	[Φ]
sospechar	[Φ]
sostener	[Φ]
subrayar	[Φ]
sugerir	[Φ]

B.1.3 Verbos positivos con sujeto fuente

Todos los verbos de esta tabla tienen asociados los tipos: vo_sf_pos, vo_sf_no_neu, vo_sf.

verbo de opinión	forma asunto
aceptar	[Φ, a, prop]
apoyar	[Φ, a, prop]
aprobar	[Φ, a, prop]
coincidir	[con, en]
concordar	[con]
defender	[Φ, a, prop]
felicitar	[Φ, a, prop]
festejar	[Φ, prop]
querer	[Φ, prop]
respaldar	[Φ, a]

B.1.4 Verbos negativos con sujeto fuente

Todos los verbos de esta tabla tienen asociados los tipos: vo_sf_neg, vo_sf_no_neu, vo_sf.

verbo de opinión	forma asunto
alertar	[sobre]
calumniar	[Φ, a]
condenar	[Φ, a, prop]
criticar	[Φ, a, prop]
denunciar	[Φ, a, prop]
desestimar	[Φ, a]
desmentir	[Φ, prop]
insultar	[Φ, a]
ironizar	[sobre]
lamentar	[Φ, prop]
quejarse	[de, por]
rechazar	[Φ, a, prop]
reprochar	[Φ, prop]
repudiar	[Φ, prop]

B.1.5 Verbos positivos con sujeto asunto

Todos los verbos de esta tabla tienen asociados los tipos: vo_sa_pos, vo_sa.

La columna 'forma asunto' no tiene sentido para estos verbos porque el asunto es sujeto, por lo tanto es un grupo nominal (que puede ser un pronombre dativo).

verbo de opinión	forma asunto
agradar	[]
conformar	[]
gustar	[]
satisfacer	[]

B.1.6 Verbos negativos con sujeto asunto

Todos los verbos de esta tabla tienen asociados los tipos: vo_sa_neg, vo_sa.

La columna 'forma asunto' no tiene sentido para estos verbos porque el asunto es sujeto, por lo tanto es un grupo nominal (que puede ser un pronombre dativo).

verbo de opinión	forma asunto
desagradar	[]
molestar	[]
preocupar	[]

B.2 Nombres de opinión

B.2.1 Nombres neutros

Todos los nombres de esta tabla tienen asociados los tipos: nom_op_neu.

nombre de opinión	preposiciones asunto	rol del complemento con <i>de</i>
aclaración	[sobre]	ambiguo
afirmación	[sobre]	fuelle
anuncio	[sobre]	ambiguo

nombre de opinión	preposiciones asunto	rol del complemento con <i>de</i>
argumentación	[por]	ambiguo
argumento	[por]	ambiguo
aseveración	[]	ambiguo
cálculo	[sobre]	ambiguo
calificación	[]	ambiguo
cita	[sobre]	ambiguo
comentario	[sobre]	fuentes
comunicación	[sobre]	fuentes
comunicado	[sobre]	fuentes
confesión	[]	ambiguo
confirmación	[]	ambiguo
consideración	[sobre, hacia]	ambiguo
creencia	[sobre]	ambiguo
declaración	[sobre]	ambiguo
dicho	[sobre]	fuentes
difusión	[]	asunto
especulación	[sobre]	ambiguo
estimación	[]	ambiguo
explicación	[sobre]	ambiguo
expresión	[]	fuentes
indicación	[sobre]	ambiguo
manifestación	[sobre]	ambiguo
mención	[sobre, a]	ambiguo
opinión	[sobre]	fuentes
palabra	[sobre]	fuentes
planteo	[sobre]	ambiguo
postura	[sobre, hacia]	fuentes
precisión	[sobre]	ambiguo
presunción	[sobre]	ambiguo
publicación	[sobre]	ambiguo
reclamo	[sobre, por]	ambiguo
relato	[sobre]	fuentes
réplica	[]	fuentes
reporte	[sobre]	ambiguo
respuesta	[sobre, a]	fuentes
revelación	[]	ambiguo
sospecha	[hacia]	ambiguo

B.2.2 Nombres positivos

Todos los nombres de esta tabla tienen asociados los tipos: nom_op_no_neu, nom_op_pos.

nombre de opinión	preposiciones asunto	rol del complemento con <i>de</i>
aceptación	[a]	ambiguo
acuerdo	[con]	ambiguo
apoyo	[a, hacia]	fuentes
aprobación	[a]	ambiguo
interés	[hacia, por, en]	fuentes
respaldo	[a, hacia]	fuentes

B.2.3 Nombres negativos

Todos los nombres de esta tabla tienen asociados los tipos: nom_op_no_neu, nom_op_neg.

nombre de opinión	preposiciones asunto	rol del complemento con <i>de</i>
agravio	[hacia, a]	fuentes
condena	[a]	fuentes
crítica	[hacia, a]	fuentes
cuestionamiento	[hacia, a]	fuentes
denuncia	[]	ambiguo
desinterés	[por, en]	fuentes
protesta	[por, ante]	fuentes
rechazo	[hacia, a]	ambiguo
repudio	[hacia, a]	fuentes

Anexo C: Reglas Contextuales

MÓDULO PARA PREDICADO DE OPINIÓN Y VERBOS SOPORTE

[illegible]

```
modulo# predicadoOpinionVerSoporte.
```

```
prOp01a# [predOp,prOp01a,verbo,Modo,Tiempo,Per,Num,Pol,OrdArg,Asunto,Mensaje]
=>
>> nc(['V',_,Lema,_,Modo,Tiempo,Per,Num|_],[[perifInf|_]]) << //
verboOpinion(Lema,Pol,OrdArg,Asunto,Mensaje).
```

```
prOp01b# [predOp,prOp01b,verbo,Modo,Tiempo,Per,Num,Pol,OrdArg,Asunto,Mensaje]
=> >> [perifInf,_,Lema,Modo,Tiempo,Per,Num] << //
verboOpinion(Lema,Pol,OrdArg,Asunto,Mensaje).
```

```
prOp02# [predOp,prOp02,verbo,Modo,Tiempo,Per,Num,Pol,OrdArg,Asunto,Mensaje]
=> >> ['V',_, 'haber',_,Modo,Tiempo,Per,Num|_],
el([predOp,_,verbo,'P',_,_,_,Pol,OrdArg,Asunto,Mensaje]) << .
```

```
prOp03# [predOp,prOp03,nombre,nil,nil,nil,nil,Pol,Int,Asunto] => >>
['N', ,Lema| ] << // nombreOpinion(Lema,Pol,Int,Asunto).
```

```
veSo01# [verSop,veSo01,Modo,Tiempo,Per,Num] => >>
['V', ,Lema, ,Modo,Tiempo,Per,Num| ] << // verboSoporte(Lema).
```

% desambiguación de verbos

% se descartan <continuar + con> y <continuar + verbo>

```
prOp04a# [predOp,prOp04a,verbo,Modo,Tiempo,Per,Num,0,fs,[gn,de],si] => >>
['V',_,contar,_,Modo,Tiempo,Per,Num|_] << [CatSig,_,LemaSig|_] //
desambContar(CatSig,LemaSig).
```

```
prOp04b# [predOp,prOp04b,verbo,Modo,Tiempo,Per,Num,0,fs,[],no] => >>
['V', ,continuar, ,Modo,Tiempo,Per,Num| ] << ['D'| ] .
```

```
prOp04c# [predOp,prOp04c,verbo,Modo,Tiempo,Per,Num,0,fs,[],no] => >>
['V',_,continuar,_,Modo,Tiempo,Per,Num|_] << ['N'|_] .
```

```
prOp04d# [predOp,prOp04d,verbo,Modo,Tiempo,Per,Num,0,fs,[],no] => >>
['V', ,continuar, ,Modo,Tiempo,Per,Num] << ['F'| ] .
```

% verbos que son de opinión solo en su versión pronominal

```
prOp05a# [predOp,prOp05a,verbo,Modo,Tiempo,Per,Num,0,fs,[a],no] => >> ['P'|
], ['V', ,referir, ,Modo,Tiempo,Per,Num| ] << .
```

```

prOp05b# [predOp,prOp05b,verbo,Modo,Tiempo,Per,Num,0,fs,[],no] => >> ['P'|_],
['V',_,pronunciar,_,Modo,Tiempo,Per,Num|_] << .

% caer bien / caer mal
prOp06a# [predOp,prOp06a,verbo,Modo,Tiempo,Per,Num,3,as,[],no] => >>
['V',_,caer,_,Modo,Tiempo,Per,Num|_], ['R',bien|_] << .

prOp06b# [predOp,prOp06b,verbo,Modo,Tiempo,Per,Num,-3,as,[],no] => >>
['V',_,caer,_,Modo,Tiempo,Per,Num|_], ['R',mal|_] << .

% desambiguación de verbos que solo son de opinión en su versión no-
pronominal

prOp07a# [predOp,prOp07a,verbo,Modo,Tiempo,Per,Num,0,fs,[],si] =>
no(['P',_,_,_,Per|_]) >> ['V',_,afirmar,_,Modo,Tiempo,Per,Num|_] << .

prOp07b# [predOp,prOp07b,verbo,Modo,Tiempo,Per,Num,0,fs,[],si] =>
no(['P',_,_,_,Per|_]) >> ['V',_,confirmar,_,Modo,Tiempo,Per,Num|_] << .

% desambiguación de "decir": descarto "es decir"

prOp08a# [predOp,prOp08a,verbo,Modo,Tiempo,Per,Num,0,fs,[],si] => no(['V',es|
_]) >> ['V',decir,_,_,Modo,Tiempo,Per,Num|_] << .

prOp08b# [predOp,prOp08b,verbo,Modo,Tiempo,Per,Num,0,fs,[],si] => >>
['V',_,decir,_,Modo,Tiempo,Per,Num|_] << // decirNoInf(Modo).

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%           MÓDULO PARA ELEMS SUBJETIVOS FUENTE,
%           INVERSORES,
%           DESAMBIGUACIÓN DE COMILLAS
%           Y
%           DETERMINACIÓN DE ELEMENTOS DE ATRIBUCIÓN DE AUTORÍA
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

modulo# elSubOperComAtribAut.

% elemSubj fuente
elSubj01b# [elemSubjFue,elSubj01b,Cat,VS] => >> [Cat,_,Lema|_] << //
elemSubjFuente(Lema,VS).

% inversores

operad01# [operador,operad01,inversor] => >> ['R',_,no|_] << . % regla
particular para operador 'no'

operad02# [operador,operad02,inversor] => >> ['D','ningún'|_] << . % regla
particular para operador 'ningún'

operad03# [operador,operad03,inversor] => >> ['D','ninguna'|_] << . % regla
particular para operador 'ninguna'

```



```

% aplicación operadores: inversor + predicadoNoNeutro

ef01# [predOp,ef01,verbo,Modo,Tiempo,Per,Num,ValSubj,RolSuj,Asunto,Mensaje|_]
=> >> el([operador,_,Ef]),
el([predOp,_,verbo,Modo,Tiempo,Per,Num,Pol,RolSuj,Asunto,Mensaje|_]) << [Cat|
R] // aplicoEfecto(Pol,Ef,ValSubj).

ef02# [predOp,ef02,nombre,nil,nil,nil,nil,ValSubj,Int,Asunto|_] => >>
el([operador,_,Ef]), el([predOp,_,nombre,_,_,_,_,Pol,Int,Asunto|_]) << [Cat|
R] // aplicoEfecto(Pol,Ef,ValSubj).

% comillas (para reglas mensaje más adelante)

comilla01# ['F',comilla01,'"',Tipo] => >> el(['F','"',''',e]) << //
ocurrenciaParImpar(Tipo).

% elementos que atribuyen autoría, tipo según (posteriormente llamados
predicados preposicionales)

atrAu01a# [atribAut,atrAu01a,'según'] => ['F'|_] >> ['S',_, 'según'|_] << .

atrAu01b# [atribAut,atrAu01b,'para'] => ['F',_,_, 'p'|_] >> ['S',_, 'para'|_]
<< .

atrAu01c# [atribAut,atrAu01c,'de acuerdo a'] => >> ['S',_, 'de'|_],
['N', 'acuerdo'|_], ['S',_, 'a'|_] << .

atrAu01d# [atribAut,atrAu01d,'de acuerdo con'] => >> ['S',_, 'de acuerdo con'|
_] << .

atrAu01e# [atribAut,atrAu01e,'de acuerdo a'] => >> ['S',_, 'de acuerdo a'|_]
<< .

atrAu01f# [atribAut,atrAu01f,'de acuerdo al'] => >> ['S',_, 'de acuerdo al'|_]
<< .

atrAu01g# [atribAut,atrAu01g,'a juicio de'] => >> ['S',_, 'a juicio de'|_]
<< .

atrAu01h# [atribAut,atrAu01h,'a juicio del'] => >> ['S',_, 'a juicio del'|_]
<< .

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%                MÓDULO PARA GRUPOS NOMINALES QUE PUEDEN SER FUENTE
                (también para asunto)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

modulo# grupoNom.

```

```

gn01a# [gn,gn01a,Tipo,Lema,simple,Gen,Num,Sem1,Sem2] => >> op(['D'|_]),
nc(['N',_,Lema,Tipo,Gen,Num,Sem1,Sem2|_],[predOp,_,nombre|_]) << .

gn01b# [gn,gn01b,Tipo,Lema,simple,Gen,Num,Sem1,Sem2] => >> ['D'|_,
['V',_,_,_, 'P'|_], nc(['N',_,Lema,Tipo,Gen,Num,Sem1,Sem2|_],
[predOp,_,nombre|_]) << .
gn01c# [gn,gn01c,Tipo,Lema,simple,Gen,Num,Sem1,Sem2] => no(['V',_,haber|_])
>> ['V',_,_,_, 'P'|_], nc(['N',_,Lema,Tipo,Gen,Num,Sem1,Sem2|_],
[predOp,_,nombre|_]) << .

gn01d# [gn,gn01d,'A',Lema,simple,Gen,Num,'0','0'] => >> ['D'|_,
['A',_,Lema,_,_,Gen,Num|_] << .

gn02a# [gn,gn02a,Tipo,Lema,simple,Gen,Num,Sem1,Sem2] => >>
el([gn,_,Tipo,Lema,_,Gen,Num,Sem1,Sem2|_]), ['A'|_] << .

gn02b# [gn,gn02b,Tipo,Lema,simple,Gen,Num,Sem1,Sem2] => >>
el([gn,_,Tipo,Lema,_,Gen,Num,Sem1,Sem2|_]), ['V',_,LemaPart,_, 'P'|_] << //
noVerboOpinion(LemaPart). % chequeo que el participio no sea verbo de opinión
(las propuestas planteadas por XXX) (14/01/2010).

gn02c# [gn,gn02c,cant,Lema,simple,'0','0','0','0'] => >> op(['D'|_]),
['Z',_,Lema|_] << .

gn02d# [gn,gn02d,Tipo,Lema,simple,Gen,Num,Sem1,Sem2] => >> el([gn,_,cant|_]),
el([gn,_,Tipo,Lema,_,Gen,Num,Sem1,Sem2|_] << .

gn02e# [gn,gn02e,Tipo,Lema,simple,Gen,Num,Sem1,Sem2] => >> el([gn,_, 'A'|_]),
el([gn,_,Tipo,Lema,_,Gen,Num,Sem1,Sem2|_] << . % 11/3 para que "un nuevo
gobierno" no quede <un nuevo> <gobierno>

gn03a# [gn,gn03a,Tipo,Lema,simple,Gen,Num,Sem1,Sem2] => >>
el([gn,_,Tipo,Lema,_,Gen,Num,Sem1,Sem2|_]), ['S',_,de, 'P'|_], el([gn|_])
<< // noNombreOpinion(Lema). % chequeo que el primer gn NO sea pred de
opinión (las declaraciones de XXX)

gn03a2# [gn,gn03a2,Tipo,Lema,simple,Gen,Num,Sem1,Sem2] => >>
[gn,_,Tipo,Lema,_,Gen,Num,Sem1,Sem2|_], ['S',_,de, 'P'|_], [gn|_] << //
nombreOpinion(Lema,_,_,_). % chequeo que el primer gn SÍ sea pred de
opinión (las declaraciones de XXX) y mantengo los dos gn

gn03b# [gn,gn03b,Tipo,Lema,simple,Gen,Num,Sem1,Sem2] => >>
el([gn,_,Tipo,Lema,_,Gen,Num,Sem1,Sem2|_]), ['S',del,_, 'P'|_], el([gn|_]) <<
// noNombreOpinion(Lema).

gn03b2# [gn,gn03b2,Tipo,Lema,simple,Gen,Num,Sem1,Sem2] => >>
[gn,_,Tipo,Lema,_,Gen,Num,Sem1,Sem2|_], ['S',del,_, 'P'|_], [gn|_] << //
nombreOpinion(Lema,_,_,_).

% REGLAS PARA PRONOMBRES

% regla que busca "a mano" los pronombres sujeto
% agregar otros pronombres: "esto explica"
gn04a# [gn,gn04a,pronSuj,Lema,simple,Per,Num,'0','0'] => >>
['P',Pronombre,Lema, 'P',Per,_,Num|_] << // pronombreSujeto(Pronombre).

% regla para pronombre "se" (fuente genérica)

```

```

gn04b# [gn,gn04b,pronSe,se,simple,'3','S','0','0'] => >> ['P',se|_] <<
[predOp,_,verbo,_,_, '3','S'|_].

% REGLAS CON TÍTULOS:   el arqueólogo Leonel Cabrera
%                       el arqueólogo del Museo de Salto, Leonel Cabrera,
%                       queda por fuera "el Director P P P P P" porque Freeling junta
%                       Director_P P P P P

% en gn05a b c d cambio coma final por cualquier puntuación: "....., expresó
el director del instituto, Juan Pérez."

gn05a# [gn,gn05a,'P',titulo,compuesto,Gen,Num,Sem1,Sem2] => >>
el([gn,_,_, 'P',_,_,simple,_,_,Sem1,Sem2|_]), ['F',',','|_|],
el([gn,_,_,Lema,simple,Gen,Num|_]) << ['F'|_] // titulo(Lema).

gn05b# [gn,gn05b,'P',titulo,compuesto,Gen,Num,Sem1,Sem2] => >>
el([gn,_,_,Lema,simple,Gen,Num|_]), ['F',',','|_|],
el([gn,_,_, 'P',_,_,simple,_,_,Sem1,Sem2|_]) << ['F'|_] // titulo(Lema).

gn05c# [gn,gn05c,'P',titulo,simple,Gen,Num,Sem1,Sem2] => >>
el([gn,_,_,Lema,simple,Gen,Num|_]), el([gn,_,_, 'P',_,_,simple,_,_,Sem1,Sem2|_])
<< // titulo(Lema).

gn05d# [gn,gn05d,'P',titulo,simple,Gen,Num,Sem1,Sem2] => >>
el([gn,_,_,Lema,simple,Gen,Num|_]), el([gn,_,_, 'P',_,_,simple,_,_,Sem1,Sem2|_])
<< // institucion(Lema).

% reglas para varios gn
% se prueba regla para coordinaciones con coma, introducen muchos errores

gn06b# [gn,gn06b,Atr1,Atr2,compuesto,nil,'P',Sem1,Sem2] => >>
el([gn,_,_,Atr1,Atr2,_,_,_,Sem1,Sem2|_]), ['C','y'|_|], el([gn|_]) << . % me
quedo con valores del primero

modulo# grupoNomAjustes.

% sustituyo comas de cierre de incisos
gnAj01a# [comaCierreInciso,gnAj01a] => [gn,gn05a|_] >> el(['F',',','|_|]) << .
gnAj01b# [comaCierreInciso,gnAj01b] => [gn,gn05b|_] >> el(['F',',','|_|]) << .

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%                               MÓDULO PARA FUENTE
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

modulo# fuente.

%%% REGLAS PARA PREDICADO VERBAL, FORMA FINITA %%%
%%% reglas fue01

```

```

% exijo sujeto con rol fuente

% ojo con pronombre "se":      para Montevideo se propone ... ("se" es la
fuente genérica)
% Fulano se propone lograr ... ("proponerse tiene otro sentido, no
comunicación)

% para filtrar posibles errores, hay que chequear concordancia en persona y
número, se corrigen errores como los siguientes:
%      Por <fuente tipo="[fue01a, sujAnt]">el libro</fuente> no <predOp
tipo="[prOp01, verbo, I]">sabremos</predOp> nada de ...
%      quienes por <fuente tipo="[fue01a, sujAnt]">unanimidad</fuente> <predOp
tipo="[prOp01, verbo, I]">decidieron</predOp> que no eran votos por Sí

% prohíbo preposición antes del sujeto

% reglas sujAnt

fue01a# [fuente,fue01a,sujAnt,0] => no(['S'|_]) >> [gn,_,_,_,_,PerGN,NumGN|_]
<< *(zonal,3), [predOp,_,verbo,'I',_,_,PerV,NumV,_,fs|_] //
concordNumPer(NumGN,NumV,PerGN,PerV).

fue01b# [fuente,fue01b,sujAnt,0] => no(['S'|_]) >> [gn,_,_,_,_,PerGN,NumGN|_]
<< *(zonal,3), [predOp,_,verbo,'S',_,_,PerV,NumV,_,fs|_] //
concordNumPer(NumGN,NumV,PerGN,PerV).

% reglas sujPos

fue01c# [fuente,fue01c,sujPos,0] => ['F'|_],
[predOp,_,verbo,'I',_,_,PerV,NumV,_,fs|_], op(['R'|_]) >>
[gn,_,_,_,_,PerGN,NumGN|_] << // concordNumPer(NumGN,NumV,PerGN,PerV).

fue01d# [fuente,fue01d,sujPos,0] => ['F'|_],
[predOp,_,verbo,'S',_,_,PerV,NumV,_,fs|_], op(['R'|_]) >>
[gn,_,_,_,_,PerGN,NumGN|_] << // concordNumPer(NumGN,NumV,PerGN,PerV).

% permito cualquier preposición + gn antes de la fuente (principalmente para
abarcas cosas como "dijo sobre Pérez el senador")

fue01e# [fuente,fue01e,sujPos,0] => ['F'|_],
[predOp,_,verbo,'I',_,_,PerV,NumV,_,fs|_], *(zonal,3), ['S'|_], [gn|_] >>
[gn,_,_,_,_,PerGN,NumGN|_] << // concordNumPer(NumGN,NumV,PerGN,PerV). % dijo
a Brecha Fulano

fue01f# [fuente,fue01f,sujPos,0] => ['F'|_],
[predOp,_,verbo,'S',_,_,PerV,NumV,_,fs|_], *(zonal,3), ['S'|_], [gn|_] >>
[gn,_,_,_,_,PerGN,NumGN|_] << // concordNumPer(NumGN,NumV,PerGN,PerV). % dijo
a Brecha Fulano

% reglas según/como + verbo opinión
% según/como dijo XXX

fue01g# [fuente,fue01g,introdMens,0] => [_,_, 'según'|_], op([_,lo|_]),
[predOp,_,verbo,'I',_,_,PerV,NumV,_,fs|_], op(['R'|_]) >>
[gn,_,_,_,_,PerGN,NumGN|_] << // concordNumPer(NumGN,NumV,PerGN,PerV).

```

```

fue01h# [fuente,fue01h,introdMens,0] => [_,'según'|_], op([_,lo|_]),
[predOp,_,verbo,'S',_,PerV,NumV,_,fs|_], op(['R'|_]) >>
[gn,_,_,_,PerGN,NumGN|_] << // concordNumPer (NumGN,NumV,PerGN,PerV) .

fue01i# [fuente,fue01i,introdMens,0] => [_,'según'|_], op([_,lo|_]),
[predOp,_,verbo,'I',_,PerV,NumV,_,fs|_], *(zonal,3), ['S',_,a,'P'|_], [gn|_]
>> [gn,_,_,_,PerGN,NumGN|_] << // concordNumPer (NumGN,NumV,PerGN,PerV) .

fue01j# [fuente,fue01j,introdMens,0] => [_,'según'|_], op([_,lo|_]),
[predOp,_,verbo,'S',_,PerV,NumV,_,fs|_], *(zonal,3), ['S',_,a,'P'|_], [gn|_]
>> [gn,_,_,_,PerGN,NumGN|_] << // concordNumPer (NumGN,NumV,PerGN,PerV) .

fue01k# [fuente,fue01k,introdMens,0] => [_,'como'|_], op([_,lo|_]),
[predOp,_,verbo,'I',_,PerV,NumV,_,fs|_], op(['R'|_]) >>
[gn,_,_,_,PerGN,NumGN|_] << // concordNumPer (NumGN,NumV,PerGN,PerV) .

fue01l# [fuente,fue01l,introdMens,0] => [_,'como'|_], op([_,lo|_]),
[predOp,_,verbo,'S',_,PerV,NumV,_,fs|_], op(['R'|_]) >>
[gn,_,_,_,PerGN,NumGN|_] << // concordNumPer (NumGN,NumV,PerGN,PerV) .

fue01m# [fuente,fue01m,introdMens,0] => [_,'como'|_], op([_,lo|_]),
[predOp,_,verbo,'I',_,PerV,NumV,_,fs|_], *(zonal,3), ['S',_,a,'P'|_], [gn|_]
>> [gn,_,_,_,PerGN,NumGN|_] << // concordNumPer (NumGN,NumV,PerGN,PerV) .

fue01n# [fuente,fue01n,introdMens,0] => [_,'como'|_], op([_,lo|_]),
[predOp,_,verbo,'S',_,PerV,NumV,_,fs|_], *(zonal,3), ['S',_,a,'P'|_], [gn|_]
>> [gn,_,_,_,PerGN,NumGN|_] << // concordNumPer (NumGN,NumV,PerGN,PerV) .

% "lo que se proponen los investigadores"

fue01o# [fuente,fue01o,sujPos,0] => ['D',lo|_], [_,'que'|_], [_,'que'|_],
[predOp,_,verbo,'I',_,PerV,NumV,_,fs|_], op(['R'|_]) >>
[gn,_,_,_,PerGN,NumGN|_] << // concordNumPer (NumGN,NumV,PerGN,PerV) .

fue01p# [fuente,fue01p,sujPos,0] => ['D',lo|_], [_,'que'|_],
[predOp,_,verbo,'S',_,PerV,NumV,_,fs|_], op(['R'|_]) >>
[gn,_,_,_,PerGN,NumGN|_] << // concordNumPer (NumGN,NumV,PerGN,PerV) .

fue01q# [fuente,fue01q,sujPos,0] => ['D',lo|_], [_,'que'|_],
[predOp,_,verbo,'I',_,PerV,NumV,_,fs|_], *(zonal,3), ['S'|_], [gn|_] >>
[gn,_,_,_,PerGN,NumGN|_] << // concordNumPer (NumGN,NumV,PerGN,PerV) .

fue01r# [fuente,fue01r,sujPos,0] => ['D',lo|_], [_,'que'|_],
[predOp,_,verbo,'S',_,PerV,NumV,_,fs|_], *(zonal,3), ['S'|_], [gn|_] >>
[gn,_,_,_,PerGN,NumGN|_] << // concordNumPer (NumGN,NumV,PerGN,PerV) .

%%% REGLAS PARA PREDICADO VERBAL, FORMA PARTICIPIO %%%
%%% reglas fue02

% exijo sujeto con rol fuente

fue02# [fuente,fue02,vozP,0] => [predOp,_,verbo,'P',_,_,_,_,fs|_],
*(zonal,3), ['S',_,por,'P'|_] >> [gn|_] << .

```

```

%% REGLAS PARA PREDICADO NOMINAL
%% reglas fue03

```

```

% complemento con "de" con rol fuente

```

```

fue03a# [fuente,fue03a,nomDe,0] => [predOp,_,nombre,_,_,_,_,_,[subj|_]],
['S',_,de,'P'|_] >> [gn|_] << .

fue03aa# [fuente,fue03aa,nomDe,0] => [predOp,_,nombre,_,_,_,_,_,[subj|_]],
['S',del,_,,'P'|_] >> [gn|_] << .

fue03a4# [fuente,fue03a4,nomDe,0] => [predOp,_,nombre,_,_,_,_,_,[amb|_]],
['S',_,de,'P'|_] >> [gn,_,,'P'|_] << .

fue03a1# [fuente,fue03a1,nomDe,0] => [predOp,_,nombre,_,_,_,_,_,[amb|_]],
['S',_,de,'P'|_] >> [gn,_,_,titulo|_] << .

fue03a2# [fuente,fue03a2,nomDe,0] => [predOp,_,nombre,_,_,_,_,_,[amb|_]],
['S',_,de,'P'|_] >> [gn,_,_,Lema|_] << // titulo(Lema).

fue03a3# [fuente,fue03a3,nomDe,0] => [predOp,_,nombre,_,_,_,_,_,[amb|_]],
['S',_,de,'P'|_] >> [gn,_,_,Lema|_] << // institucion(Lema).

fue03a42# [fuente,fue03a42,nomDe,0] => [predOp,_,nombre,_,_,_,_,_,[amb|_]],
['S',del,_,,'P'|_] >> [gn,_,,'P'|_] << .

fue03a12# [fuente,fue03a12,nomDe,0] => [predOp,_,nombre,_,_,_,_,_,[amb|_]],
['S',del,_,,'P'|_] >> [gn,_,_,titulo|_] << .

fue03a22# [fuente,fue03a22,nomDe,0] => [predOp,_,nombre,_,_,_,_,_,[amb|_]],
['S',del,_,,'P'|_] >> [gn,_,_,Lema|_] << // titulo(Lema).
fue03a32# [fuente,fue03a32,nomDe,0] => [predOp,_,nombre,_,_,_,_,_,[amb|_]],
['S',del,_,,'P'|_] >> [gn,_,_,Lema|_] << // institucion(Lema).

```

```

% verbo soporte: el ministro brindó declaraciones / presentará ante ... un
informe / una opinión arriesga Trindade / los argumentos presentados por
Uruguay

```

```

% otros verbos sop: lanzar una propuesta, ...
% acá "plantear" es verbo soporte de "interrogantes etc"
% <fuente tipo="[fue01a, sujAnt]">las interrogantes y las perplejidades de
naturaleza política</fuente> que se le <predOp tipo="[prOp01, verbo,
I]">plantean</predOp> al intelectual

```

```

fue03b# [fuente,fue03b,nomVS,0] => [predOp,_,nombre|_],
[verSop,_, 'I',_,PerV,NumV] >> [gn,_,_,_,PerGN,NumGN|_] << //
concordNumPer (NumGN,NumV,PerGN,PerV) .

```

```

fue03c# [fuente,fue03c,nomVS,0] => >> [gn,_,_,_,PerGN,NumGN|_] <<
[verSop,_, 'I',_,PerV,NumV], *(zona2,4), [predOp,_,nombre|_] //
concordNumPer (NumGN,NumV,PerGN,PerV) .

```

```

fue03d# [fuente,fue03d,nomVS,0] => [predOp,_,nombre|_], [verSop,_, 'P'|_],
['S',_,por,'P'|_] >> [gn|_] << .

```

```

%%% REGLAS PARA según y similares %%%
%%% reglas fue04

```

```

% EN MÓDULO DE DESAMBIGUACIÓN DE COMILLAS FUERON MARCADOS LOS ELEMENTOS DE
ATRIBUCIÓN DE AUTORÍA tipo según (posteriormente llamados predicados
preposicionales)

```

```

fue04a# [fuente,fue04a,introdMens,0] => [atribAut,_,_] >> [gn|_] << .

```

```

% absorbo la coma de los segmentos con según y similares (si hay)
fue04c# [fuente,fue04c,introdMens,0] => >> el([fuente,_,introdMens|_]),
['F',',','|_|] << .

```

```

% REGLAS PARA PREDICADOS CON SUJETO ROL ASUNTO

```

```

fue05a# [fuente,fue05a,fueDat,0] => >> ['S',_,a|_], [gn|_] << *(zona1,3),
[predOp,_,verbo,'I',_,_,_,_,as|_] .

```

```

fue05b# [fuente,fue05b,fueDat,0] => >> ['S',_,a|_], [gn|_] << *(zona1,3),
[predOp,_,verbo,'S',_,_,_,_,as|_] .

```

```

fue05c# [fuente,fue05c,fueDat,0] => [predOp,_,verbo,'I',_,_,_,_,as|_] >>
['S',a|_], [gn|_] << .

```

```

fue05d# [fuente,fue05d,fueDat,0] => [predOp,_,verbo,'S',_,_,_,_,as|_] >>
['S',a|_], [gn|_] << .

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% MÓDULO PARA ASUNTO
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

```

modulo# asunto.

```

```

%% asunto introducido por preposición característica (sobre, respecto a, en
contra de, a favor de, ...)

```

```

asu01a# [asunto,asu01a,prepAsunto,Pol] => [predOp|_], *(zona3,4) >>
['S',_,Lema|_], [gn|_] << // prepAsunto(Lema,Pol).

```

```

asu01b# [asunto,asu01b,prepAsunto,-3] => [predOp|_], *(zona3,4) >>
['A',_,contrario|_], ['S',a|_], [gn|_] << .

```

```

asu01c# [asunto,asu01c,prepAsunto,0] => [predOp|_], *(zona3,4) >> ['S',en|_],
['N',_,relación|_], ['S',a|_], [gn|_] << . % agregarlas al archivo de
locuciones de freeling

```

```
%%% regla para asunto introducido por preposiciones regidas por el predOp
verbal
```

```
% exijo sujeto con rol fuente
```

```
asu02a# [asunto,asu02a,Lema,0] => [predOp,_,verbo,_,_,_,_,_,fs,ListaAsunto|_]
>> ['S',_,Lema|_], [gn|_] << // member(Lema,ListaAsunto).
```

```
asu02b# [asunto,asu02b,de,0] => [predOp,_,verbo,_,_,_,_,_,fs,ListaAsunto|_]
>> ['S',del|_], [gn|_] << // member(de,ListaAsunto).
```

```
asu02c# [asunto,asu02c,a,0] => [predOp,_,verbo,_,_,_,_,_,fs,ListaAsunto|_] >>
['S',al|_], [gn|_] << // member(a,ListaAsunto).
```

```
asu02d# [asunto,asu02d,prop,0] => [predOp,_,verbo,_,_,_,_,_,fs,ListaAsunto|_]
>> [prop|_] << // member(prop,ListaAsunto).
```

```
%%% reglas especiales para asunto introducido por "de", con predicado nominal
% complemento con "de" rol ambiguo o asunto
```

```
asu03a# [asunto,asu03a,de,0] => [predOp,_,nombre,_,_,_,_,_,[amb|_]] >>
['S',del|_], nc([gn|_],[[fuente|_],[gn|_]]) << .
```

```
asu03b# [asunto,asu03b,de,0] => [predOp,_,nombre,_,_,_,_,_,[amb|_]] >>
['S',del|_], nc([gn|_],[[fuente|_],[gn|_]]) << .
```

```
asu03c# [asunto,asu03c,de,0] => [predOp,_,nombre,_,_,_,_,_,[obj|_]] >>
['S',del|_], nc([gn|_],[[fuente|_],[gn|_]]) << .
```

```
asu03d# [asunto,asu03d,de,0] => [predOp,_,nombre,_,_,_,_,_,[obj|_]] >>
['S',del|_], nc([gn|_],[[fuente|_],[gn|_]]) << .
```

```
%%% reglas especiales para asunto introducido por "a", con predicado nominal
```

```
asu03e# [asunto,asu03e,a,0] => [predOp,_,nombre,_,_,_,_,_,[_a]] >> ['S',a|_],
[gn|_] << .
```

```
asu03f# [asunto,asu03f,a,0] => [predOp,_,nombre,_,_,_,_,_,[_a]] >>
['S',al|_], [gn|_] << .
```

```
%%% reglas especiales para asunto GN sin preposición
```

```
% exijo sujeto con rol fuente
```

```
asu04# [asunto,asu04,gn,0] => [predOp,_,verbo,_,_,_,_,_,fs,ListaAsunto|_] >>
nc([gn|_],[[fuente|_],[gn|_]]) << // member(gn,ListaAsunto).
```

```
%%% reglas para asunto antepuesto
```

```
% consultado sobre ...,
```

```
% en cuanto a, respecto a, respecto de, en lo que refiere a, en lo referente
a (pocos casos en CdE), con respecto a
```



```

% hay casos en que el asunto termina cuando aparece el predOp, sin coma (ni
fuente)
% igualmente por ahora se exige siempre coma.
% se hace una sola regla usando el predicado prepAsunto.
% se pide loc. prep. neutra porque son las que pueden aparecer antepuestas

%asu05a# [asunto,asu05a,Lema,0] => ['F'|_] >> ['S',_,Lema|_], *(zona7,40) <<
['F',' ','|_|], *(zona8,15), [predOp|_] // prepAsunto(Lema,0).

%asu05a# [asunto,asu05a,Lema,0] => ['F'|_] >> ['S',_,Lema|_], *(zona7,40) <<
['F',' ','|_|], [predOp|_] // prepAsunto(Lema,0).

% regla similar a anterior con coma opcional, con zona7b, sin puntuación al
principio
asu05a# [asunto,asu05a,Lema,0] => >> ['S',_,Lema|_], *(zona7b,40) <<
op(['F',' ','|_|], [predOp,_,verbo,'I'|_] // prepAsunto(Lema,0). % 20100121

% repito la regla sin coma separando el asunto, pero con fuente
asu05b# [asunto,asu05b,Lema,0] => >> ['S',_,Lema|_], *(zona7b,40) <<
op(['F',' ','|_|], [fuente,_,subjAnt|_], *(zona1,3), [predOp,_,verbo,'I'|_] //
prepAsunto(Lema,0).

% regla para asunto como sujeto pasivo (verbo opinión en participio)

asu06a# [asunto,asu06a,vozP,0] => >> [gn|_] <<
[predOp,_,verbo,'P',_,_,_,_,fs|_] .

% regla para asunto con función sujeto
% exijo sujeto con rol asunto
% chequeo concordancia sujeto verbo

asu07a# [asunto,asu07a,asuSuj,0] => no(['S',_,a|_]) >>
[gn,_,_,_,_,PerGN,NumGN|_] << *(zona1,3),
[predOp,_,verbo,'I',_,_,PerV,NumV,_,as|_] //
concordNumPer(NumGN,NumV,PerGN,PerV).

asu07b# [asunto,asu07b,asuSuj,0] => no(['S',_,a|_]) >>
[gn,_,_,_,_,PerGN,NumGN|_] << *(zona1,3),
[predOp,_,verbo,'S',_,_,PerV,NumV,_,as|_] //
concordNumPer(NumGN,NumV,PerGN,PerV).

asu07c# [asunto,asu07c,asuSuj,0] => [predOp,_,verbo,'I',_,_,PerV,NumV,_,as|_],
*(zona1,3) >> [gn,_,_,_,_,PerGN,NumGN|_] << //
concordNumPer(NumGN,NumV,PerGN,PerV).

asu07d# [asunto,asu07d,asuSuj,0] => [predOp,_,verbo,'S',_,_,PerV,NumV,_,as|_],
*(zona1,3) >> [gn,_,_,_,_,PerGN,NumGN|_] << //
concordNumPer(NumGN,NumV,PerGN,PerV).

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% MÓDULO PARA MENSAJE
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
```

```
modulo# mensaje.
```

```
% regla para marcar una cita (acá entran también las comillas de otro tipo)
```

```
cita01# [cita,cita01] => >> ['F',_, "'", abre], *(zona4,200),
['F',_, "'", cierra] << .
```

```
% reglas para las citas que son mensaje de una opinión
```

```
men01a# [mensaje,men01a,cita,polMens] => >> [cita,_] << .
```

```
men01b# [mensaje,men01b,cita,PolMens] => >> el([mensaje,_,cita,PolMens1]),
op(['F',_, '.'|_]), el([mensaje,_,cita,PolMens2]) << ['F', '.'|_] //
combinoValoresSubj([PolMens1,PolMens2],PolMens).
```

```
% reglas para marcar el mensaje proposicional
```

```
% exijo que el verbo acepte un mensaje proposicional
```

```
%men02a# [mensaje,men02a,prop,polMens] => [predOp,_,verbo,_,_,_,_,fs,_,si|_],
op(['R',_,_, 'G'|_]) >> [prop|_] << .
```

```
%men02b# [mensaje,men02b,prop,polMens] => [predOp,_,verbo,_,_,_,_,fs,_,si|_],
['S',_,a|_], [gn|_] >> [prop|_] << .
```

```
men02a# [mensaje,men02a,prop,polMens] => [predOp,_,verbo,_,_,_,_,fs,_,si|_],
*(zona10,5) >> [prop|_] << .
```

```
% incorporo cita siguiente, si hay
```

```
men02c# [mensaje,men02c,prop,PolMens] => >> el([mensaje,_,prop,PolMens1]),
['F',_, '.'|_], el([mensaje,_,cita,PolMens2]) << ['F', '.'|_] //
combinoValoresSubj([PolMens1,PolMens2],PolMens).
```

```
% reglas para mensaje cuando hay introductor de fuente tipo 'según'
```

```
men03a# [mensaje,men03a,introdMensPos,polMens] => [fuente,_,introdMens,_] >>
*(zona5,40) << ['F',_,_,p|_] . % esta regla debe marcar hasta un fin de prop
que no se abra dentro del mensaje, no es posible hacerlo.
```

```
men03b# [mensaje,men03b,introdMensAnt,polMens] => ['F', '.'|_] >> *(zona6,40)
<< ['F', ', '|_], [atribAut|_], [fuente,_,introdMens,_] .
```

```
% reglas para mensaje antepuesto sin comilla
```

```
men04b# [mensaje,men04b,citaSinComilla,polMens] => ['F', '.'|_] >> *(zona9,40)
<< ['F', ', '|_], [predOp,_,verbo, 'I',_,_,_,_,fs|_] .
```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%                                MÓDULO PARA PREDICADOS DE OPINIÓN EXPRESADOS MEDIANTE
%                                LOCUCIONES O EXPRESIONES
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

```

modulo# otrosPredicados.

```

```

% x está (totalmente) de acuerdo con Y
% x está (relativamente) a favor de Y
% x está (absolutamente) en contra de Y

```

```

% ver que cuando se niega se expresan de manera diferente los operadores:
% x no está para nada de acuerdo
% x no está en absoluto a favor de Y

```

```

fueOtro01a# [fuente,fueOtro01a,'estar+prepNoNeutra',0] => >>
[gn,_,_,_,PerGN,NumGN|_] << *(zonal,3), ['V',_,estar,_,_,_,PerV,NumV|_],
op([operador|_]), ['S',_, 'de acuerdo con'|_] //
concordNumPer(NumGN,NumV,PerGN,PerV).

```

```

predOtro01a#
[predOp,predOtro01a,'estar+prepNoNeutra',nil,nil,nil,nil,3,0,nil,nil] => >>
['V',_,estar|_], ['S',_, 'de acuerdo con'|_] << .

```

```

predOtro01b#
[predOp,predOtro01b,'estar+prepNoNeutra',nil,nil,nil,nil,VS,0,nil,nil] => >>
['V',_,estar|_], [operador,_,Ef], ['S',_, 'de acuerdo con'|_] << //
aplicoEfecto(3,Ef,VS).

```

```

fueOtro02a# [fuente,fueOtro02a,'estar+prepNoNeutra',0] => >>
[gn,_,_,_,PerGN,NumGN|_] << *(zonal,3), ['V',_,estar,_,_,_,PerV,NumV|_],
['S',_, 'a favor de'|_] // concordNumPer(NumGN,NumV,PerGN,PerV).
fueOtro02b# [fuente,fueOtro02b,'estar+prepNoNeutra',0] => >>
[gn,_,_,_,PerGN,NumGN|_] << *(zonal,3), ['V',_,estar,_,_,_,PerV,NumV|_],
['S',_, 'en contra de'|_] // concordNumPer(NumGN,NumV,PerGN,PerV).

```

```

fueOtro02c# [fuente,fueOtro02c,'estar+prepNoNeutra',0] => >>
[gn,_,_,_,PerGN,NumGN|_] << *(zonal,3), ['V',_,estar,_,_,_,PerV,NumV|_],
['S',_, 'contra'|_] // concordNumPer(NumGN,NumV,PerGN,PerV).

```

```

predOtro02a#
[predOp,predOtro02a,'estar+prepNoNeutra',nil,nil,nil,nil,VS,0,nil,nil] => >>
['V',_,estar|_], ['S',_,Lema|_] << // prepAsunto(Lema,VS).

```

```

predOtro02b#
[predOp,predOtro02b,'estar+prepNoNeutra',nil,nil,nil,nil,VS,0,nil,nil] => >>
['V',_,estar|_], [operador,_,Ef], ['S',_, 'a favor de'|_] << //
aplicoEfecto(3,Ef,VS).

```

```

predOtro02c#
[predOp,predOtro02c,'estar+prepNoNeutra',nil,nil,nil,nil,VS,0,nil,nil] => >>
['V',_,estar|_], [operador,_,Ef], ['S',_, 'en contra de'|_] << //
aplicoEfecto(-3,Ef,VS).

```

```
predOtro02d#
[predOp,predOtro02d,'estar+prepNoNeutra',nil,nil,nil,nil,VS,0,nil,nil] => >>
['V',_,estar|_], [operador,_,Ef], ['S',_, 'contra'|_] << // aplicoEfecto(-
3,Ef,VS).
```

```
asuOtro# [asunto,asuOtro,'estar+prepNoNeutra',0] =>
[predOp,_, 'estar+prepNoNeutra'|_] >> [gn|_] << .
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%                                MÓDULO PARA OPINIONES
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
modulo# opinion.
```

```
% esquema < fuente pred asunto : mensaje>
```

```
op00a# [opinion,op00a,Pol] => >> [fuente,_,sujAnt,PolFue], *(zona1,3),
[predOp,_,_,_,_,_,PolPre|_], *(zona3,4), [asunto,_,_,PolAsu], ['F',':'|_],
[mensaje,_,cita,PolMens] << //
calculoPolOp(PolFue,PolPre,PolAsu,PolMens,Pol).
```

```
% esquema < asunto fuente pred mensaje>
```

```
op08a# [opinion,op08a,Pol] => >> [asunto,_,_,PolAsu], op(['F',',',_|_]),
[fuente,_,sujAnt,PolFue|_], *(zona1,3), [predOp,_,_,_,_,_,PolPre,_,_,si|_],
*(zona10,5), [mensaje,_,prop,PolMens] << //
calculoPolOp(PolFue,PolPre,PolAsu,PolMens,Pol).
```

```
op08c# [opinion,op08c,Pol] => >> [asunto,_,_,PolAsu], op(['F',',',_|_]),
[fuente,_,_,PolFue|_], *(zona1,3), [predOp,_,_,_,_,_,PolPre|_],
*(zona10,5), ['F',_,':'|_], [mensaje,_,cita,PolMens] << //
calculoPolOp(PolFue,PolPre,PolAsu,PolMens,Pol).
```

```
% esquema < asunto pred mensaje>
```

```
op09a# [opinion,op09a,Pol] => >> [asunto,_,_,PolAsu], op(['F',',',_|_]),
*(zona1,3), [predOp,_,verbo,'I',_,_,_,PolPre,_,_,si|_], *(zona10,5),
[mensaje,_,prop,PolMens] << // calculoPolOp(0,PolPre,PolAsu,PolMens,Pol).
```

```
op09c# [opinion,op09c,Pol] => >> [asunto,_,_,PolAsu], op(['F',',',_|_]),
*(zona1,3), [predOp,_,verbo,'I',_,_,_,PolPre|_], *(zona10,5), ['F',_,':'|_],
[mensaje,_,cita,PolMens] << // calculoPolOp(0,PolPre,PolAsu,PolMens,Pol).
```

```
% esquema < fuente pred mensaje >
```

```
op01a# [opinion,op01a,Pol] => >> [fuente,_,sujAnt,PolFue], *(zona1,3),
[predOp,_,_,_,_,_,PolPre,_,_,si|_], *(zona10,5), [mensaje,_,prop,PolMens]
<< // calculoPolOp(PolFue,PolPre,0,PolMens,Pol).
```

```

op01c# [opinion,op01c,Pol] => >> [fuente,_,sujAnt,PolFue], *(zona1,3),
[predOp,_,_,_,_,_,_,PolPre|_], *(zona10,5), ['F',':'|_],
[mensaje,_,cita,PolMens] << // calculoPolOp(PolFue,PolPre,0,PolMens,Pol).

% esquema < fuente pred asunto >

op02a# [opinion,op02a,Pol] => >> [fuente,_,sujAnt,PolFue], *(zona1,3),
[predOp,_,_,_,_,_,_,PolPre|_], *(zona3,4), [asunto,_,_,PolAsu] << //
calculoPolOp(PolFue,PolPre,PolAsu,0,Pol).

op02b# [opinion,op02b,Pol] => >> [fuente,_, 'estar+prepNoNeutra',PolFue],
*(zona1,3), [predOp,_,_,_,_,_,_,PolPre|_], [asunto,_,_,PolAsu] << //
calculoPolOp(PolFue,PolPre,PolAsu,0,Pol).

% esquema < fuente pred>
op03a# [opinion,op03a,Pol] => >> [fuente,_,sujAnt,PolFue], *(zona1,3),
[predOp,_,_,_,_,_,_,PolPre|_] << // calculoPolOp(PolFue,PolPre,0,0,Pol).

% esquema < pred asunto mensaje>

op04a# [opinion,op04a,Pol] => >> [predOp,_,_,_,_,_,_,PolPre|_], *(zona3op,4),
[asunto,_,_,PolAsu], ['F',':'|_], [mensaje,_,cita,PolMens] << //
calculoPolOp(0,PolPre,PolAsu,PolMens,Pol).

% esquema < pred asunto >

op05a# [opinion,op05a,Pol] => no([fuente|_]) >> [predOp,_,_,_,_,_,_,PolPre|
_], *(zona3op,4), [asunto,_,_,PolAsu] << //
calculoPolOp(0,PolPre,PolAsu,0,Pol).

% esquema < pred mensaje >

op06a# [opinion,op06a,Pol] => >> [predOp,_,_,_,_,_,_,PolPre,_,_,si|_],
*(zona10,5), [mensaje,_,prop,PolMens] << //
calculoPolOp(0,PolPre,0,PolMens,Pol).

op06c# [opinion,op06c,Pol] => >> [predOp,_,_,_,_,_,_,PolPre|_], *(zona10,5),
['F',':'|_], [mensaje,_,cita,PolMens] << //
calculoPolOp(0,PolPre,0,PolMens,Pol).

% esquema < mensaje > solo si el mensaje entero constituye una oración
completa.

op07a# [opinion,op07a,PolMens] => ['F',_, '.'|_] >> [mensaje,_,cita,PolMens]
<< ['F',_, '.'|_] .

% esquema < mensaje pred fuente asunto>

op10a# [opinion,op10a,Pol] => >> [mensaje,_,cita,PolMens], ['F',':'|_],
[predOp,_,_,_,_,_,_,PolPre|_], *(zona1op,5), [fuente,_,sujPos,PolFue],
op(['F',_, '.'|_]), [asunto,_,_,PolAsu] << //
calculoPolOp(PolFue,PolPre,PolAsu,PolMens,Pol).

```

```

op10b# [opinion,op10b,Pol] => >> [mensaje,_,citaSinComilla,PolMens],
['F',' ','|_], [predOp,_,_,_,_,_,_,PolPre|_], *(zonalop,5),
[fuente,_,subjPos,PolFue], op(['F',_,',','|_]), [asunto,_,_,PolAsu] << //
calculoPolOp(PolFue,PolPre,PolAsu,PolMens,Pol).

% esquema < mensaje pred fuente>

op11a# [opinion,op11a,Pol] => >> [mensaje,_,cita,PolMens], ['F',' ','|_],
[predOp,_,_,_,_,_,_,PolPre|_], *(zonalop,5), [fuente,_,subjPos,PolFue] << //
calculoPolOp(PolFue,PolPre,0,PolMens,Pol).

op11b# [opinion,op11b,Pol] => >> [mensaje,_,citaSinComilla,PolMens],
['F',' ','|_], [predOp,_,_,_,_,_,_,PolPre|_], *(zonalop,5),
[fuente,_,subjPos,PolFue] << // calculoPolOp(PolFue,PolPre,0,PolMens,Pol).

% esquema < mensaje pred asunto>

op12a# [opinion,op12a,Pol] => >> [mensaje,_,cita,PolMens], ['F',' ','|_],
[predOp,_,_,_,_,_,_,PolPre|_], [asunto,_,_,PolAsu] << //
calculoPolOp(0,PolPre,PolAsu,PolMens,Pol).

op12b# [opinion,op12b,Pol] => >> [mensaje,_,citaSinComilla,PolMens],
['F',' ','|_], [predOp,_,_,_,_,_,_,PolPre|_], [asunto,_,_,PolAsu] << //
calculoPolOp(0,PolPre,PolAsu,PolMens,Pol).

% esquema < mensaje pred>

op13a# [opinion,op13a,Pol] => >> [mensaje,_,cita,PolMens], ['F',' ','|_],
[predOp,_,verbo,'I',_,_,_,_,PolPre|_] << //
calculoPolOp(0,PolPre,0,PolMens,Pol).

% esquema < pred fuente asunto mensaje>

op14a# [opinion,op14a,Pol] => >> [predOp,_,_,_,_,_,_,PolPre|_],
[fuente,_,_,PolFue], [asunto,_,_,PolAsu], ['F',' ':'|_],
[mensaje,_,cita,PolMens] << //
calculoPolOp(PolFue,PolPre,PolAsu,PolMens,Pol).

% esquema < pred fuente mensaje>

op15a# [opinion,op15a,Pol] => >> [predOp,_,_,_,_,_,_,PolPre|_],
[fuente,_,_,PolFue], ['F',' ':'|_], [mensaje,_,cita,PolMens] << //
calculoPolOp(PolFue,PolPre,0,PolMens,Pol).

% esquema < 'según' fuente mensaje>

op16a# [opinion,op16a,Pol] => >> [atribAut|_], [fuente,_,introdMens,PolFue],
[mensaje,_,introdMensPos,PolMens] << // calculoPolOp(PolFue,0,0,PolMens,Pol).

% esquema < mensaje 'según' fuente>

op17a# [opinion,op16a,Pol] => >> [mensaje,_,introdMensAnt,PolMens], ['F',' ','|_],
[atribAut|_], [fuente,_,introdMens,PolFue] << //
calculoPolOp(PolFue,0,0,PolMens,Pol).

```

```

% esquema < pred_participio 'por' fuente asunto>

op18a# [opinion,op18a,Pol] => >> [predOp,_,_,_,_,_,_,PolPre|_], *(zona1,3),
['S',por|_], [fuente,_,vozP,PolFue], [asunto,_,_,PolAsu] << //
calculoPolOp(PolFue,PolPre,PolAsu,0,Pol).

% esquema < pred_participio 'por' fuente>

op19a# [opinion,op19a,Pol] => >> [predOp,_,_,_,_,_,_,PolPre|_], *(zona1,3),
['S',por|_], [fuente,_,vozP,PolFue] << //
calculoPolOp(PolFue,PolPre,0,0,Pol).

% esquema < asunto pred_participio 'por' fuente>

op20a# [opinion,op20a,Pol] => >> [asunto,_,vozP,PolAsu],
[predOp,_,_,_,_,_,_,PolPre|_], *(zona1,3), ['S',por|_],
[fuente,_,vozP,PolFue] << // calculoPolOp(PolFue,PolPre,PolAsu,0,Pol).

% esquema < pred_nom 'de' fuente>

op21a# [opinion,op21a,Pol] => >> [predOp,_,nombre,_,_,_,_,PolPre|_], ['S',de|
_], [fuente,_,nomDe,PolFue] << // calculoPolOp(PolFue,PolPre,0,0,Pol).

% esquema < pred_nom 'de' fuente asunto>

op22a# [opinion,op22a,Pol] => >> [predOp,_,nombre,_,_,_,_,PolPre|_], ['S',de|
_], [fuente,_,nomDe,PolFue], [asunto,_,_,PolAsu] << //
calculoPolOp(PolFue,PolPre,PolAsu,0,Pol).

% esquema < pred_nom verSop fuente>

op23a# [opinion,op23a,Pol] => >> [predOp,_,nombre,_,_,_,_,PolPre|_], [verSop|
_], op(['S'|_]), [fuente,_,nomVS,PolFue] << //
calculoPolOp(PolFue,PolPre,0,0,Pol).

% esquema < fuente verSop_part pred_nom>

op24a# [opinion,op24a,Pol] => >> [fuente,_,nomVS,PolFue], [verSop|_],
*(zona2,4), [predOp,_,nombre,_,_,_,_,PolPre|_] << //
calculoPolOp(PolFue,PolPre,0,0,Pol).

% esquema < pred_nom verSop fuente asunto>

op25a# [opinion,op25a,Pol] => >> [predOp,_,nombre,_,_,_,_,PolPre|_], [verSop|
_], op(['S'|_]), [fuente,_,nomVS,PolFue], [asunto,_,_,PolAsu] << //
calculoPolOp(PolFue,PolPre,PolAsu,0,Pol).

% esquema < fuente verSop_part pred_nom asunto>

```

```

op26a# [opinion,op26a,Pol] => >> [fuente,_,nomVS,PolFue], [verSop|_],
*(zona2,4), [predOp,_,nombre,_,_,_,PolPre|_], [asunto,_,_,PolAsu] << //
calculoPolOp(PolFue,PolPre,PolAsu,0,Pol).

% esquema < fuente predOp_as asunto >

op27a# [opinion,op27a,Pol] => >> [fuente,_,fueDat,PolFue], *(zona2,3),
[predOp,_,verbo,_,_,_,PolPre,as|_], *(zona3,3), [asunto,_,asuSuj,PolAsu] <<
// calculoPolOp(PolFue,PolPre,PolAsu,0,Pol).

% esquema < asunto predOp_as fuente >

op28a# [opinion,op28a,Pol] => >> [asunto,_,asuSuj,PolAsu], *(zona3,3),
[predOp,_,verbo,_,_,_,PolPre,as|_], [fuente,_,fueDat,PolFue] << //
calculoPolOp(PolFue,PolPre,PolAsu,0,Pol).

% esquema <opinión + mensaje>

op07a# [opinion,op07a,Pol] => >> [opinion,_,PolOp], ['F',_, '.'|_],
[mensaje,_,cita,PolMens], ['F',_, '.'|_] << //
incorporacionMensaje(PolOp,PolMens,Pol).

% esquema < mensaje >, solo si el mensaje entero constituye una oración
completa y no fue incorporado a una opinión por regla anterior.

op07b# [opinion,op07b,PolMens] => ['F',_, '.'|_] >>
nc([mensaje,_,cita,PolMens],[[opinion|_]]) << ['F',_, '.'|_] .

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%                                zonas de exclusión
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

exclusion# zona1 => ['V'|_], ['F'|_], ['C'|_], ['S'|_], [gn|_], [predOp|_],
[fuente|_] . % puntuación, conjunciones

exclusion# zona1op => ['V'|_], ['F'|_], ['C'|_], [predOp|_], [fuente|_] . %
permiso gn por "dijo al semanario el senador"

exclusion# zona2 => ['V'|_], ['F'|_], ['C'|_], ['P','R'|_], [predOp|_],
[fuente|_], [verSop|_] .

exclusion# zona3 => ['V'|_], ['F'|_], ['C'|_], ['P','R'|_], [predOp|_],
[verSop|_], [asunto|_] .

exclusion# zona3op => ['V'|_], ['F'|_], ['C'|_], ['P','R'|_], [predOp|_],
[verSop|_], [asunto|_], [fuente|_], [mensaje|_] .

exclusion# zona4 => ['F',_, ""|_] .

exclusion# zona5 => ['F', '.'|_], [fuente,_,introdMens|_] .

```



```

exclusion# zona6 => ['F', '.', '|_], ['F', ', ', '|_] .

exclusion# zona7 => ['F'|_], ['S', _, 'sobre'|_], ['S', _, 'en cuanto a'|_],
['S', _, 'en cuanto al'|_], ['S', _, 'respecto a'|_], ['S', _, 'respecto al'|_],
['S', _, 'respecto de'|_], ['S', _, 'respecto del'|_], ['S', _, 'en lo que refiere
a'|_], ['S', _, 'en lo que refiere al'|_], ['S', _, 'en lo que respecta a'|_],
['S', _, 'en lo que respecta al'|_], ['S', _, 'en lo referente a'|_], ['S', _, 'en
lo referente al'|_], ['S', _, 'en lo que concierne a'|_], ['S', _, 'en lo que
concierne al'|_], ['S', _, 'en lo concerniente a'|_], ['S', _, 'en lo
concerniente al'|_], ['S', _, 'con respecto a'|_], ['S', _, 'con respecto al'|_],
['S', _, 'con respecto de'|_], ['S', _, 'con respecto del'|_], ['S', _, 'con
relación a'|_], ['S', _, 'con relación al'|_], ['S', _, 'en relación a'|_],
['S', _, 'en relación al'|_] .

exclusion# zona7b => ['F'|_], [predOp, _, verbo|_], [fuente|_], ['S', _, 'sobre'|
_|, ['S', _, 'en cuanto a'|_], ['S', _, 'en cuanto al'|_], ['S', _, 'respecto a'|
_|, ['S', _, 'respecto al'|_], ['S', _, 'respecto de'|_], ['S', _, 'respecto del'|
_|, ['S', _, 'en lo que refiere a'|_], ['S', _, 'en lo que refiere al'|_],
['S', _, 'en lo que respecta a'|_], ['S', _, 'en lo que respecta al'|_],
['S', _, 'en lo referente a'|_], ['S', _, 'en lo referente al'|_], ['S', _, 'en lo
que concierne a'|_], ['S', _, 'en lo que concierne al'|_], ['S', _, 'en lo
concerniente a'|_], ['S', _, 'en lo concerniente al'|_], ['S', _, 'con respecto
a'|_], ['S', _, 'con respecto al'|_], ['S', _, 'con respecto de'|_], ['S', _, 'con
respecto del'|_], ['S', _, 'con relación a'|_], ['S', _, 'con relación al'|_],
['S', _, 'en relación a'|_], ['S', _, 'en relación al'|_] .

exclusion# zona8 => ['F', ', ', '|_], [predOp|_] .

exclusion# zona8b => [fuente|_], [predOp|_] .

exclusion# zona8c => [fuente|_], [predOp|_], ['F', ', ', '|_] .

exclusion# zona9 => ['F', '.', '|_], ['F', ', ', '|_], ['F', _, '""|_], [asunto|_] .

exclusion# zona10 => ['F'|_], [predOp|_], [mensaje|_], ['V'|_], [prop|_],
[fuente|_] .

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%          PREDICADOS AUXILIARES
%          UTILIZADOS POR LAS REGLAS
%          (después del símbolo // )
%
%          +
%
%          LISTAS DE PALABRAS Y EXPRESIONES:
%          VERBOS SOPORTE
%          TÍTULOS PERSONA
%          INDICADOR INSTITUCIÓN
%          LOCUCIONES INTRODUCTORAS DE ASUNTO
%          (CON VALOR DE POLARIDAD)
%          OTRAS
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

```
% desambiguación de verbos op
```

```

%"contar" no seguido de "con"
desambContar('S',con):-
    !,
    fail.
desambContar(_,_).

% para "es decir"
decirNoInf('N'):-
    !,
    fail.
decirNoInf(_).

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% títulos para reglas que arman gn posible fuente
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

titulo('señor').
titulo(doctor).
titulo(presidente).
titulo(vicepresidente).
titulo(ingeniero).
titulo('arqueólogo').
titulo(abogado).
titulo(economista).
titulo(senador).
titulo(intendente).
titulo(diputado).
titulo(director).
titulo(consejero).
titulo(embajador).
titulo(canciller).
titulo(ministro).
titulo('líder').
titulo('capitán').
titulo(juez).
titulo(secretario).
titulo(fundador).
titulo(tupamaro).
titulo(frenteamplista).
titulo(estudiante).
titulo(alumno).
titulo(fundador).
titulo(secretario).
titulo(directivo).
titulo(maestro).
titulo(padre).
titulo(orador).
titulo(investigador).
titulo(jefe).
titulo('técnico').
titulo(colega). % no tiene femenino
titulo('responsable').
titulo(magistrado).
titulo(candidato).
titulo('psicólogo').
titulo('portavoz').
titulo('fiscal').
titulo('dirigente').
titulo('zaguero').
titulo('correligionario').

```

% en femenino, porque los nombres no tienen flexión de género, es un problema porque yo miro los lemas

```
titulo('señora').
titulo(doctora).
titulo(presidenta).
titulo(ingeniera).
titulo('arqueóloga').
titulo(abogada).
titulo(senadora).
titulo(intendenta).
titulo(diputada).
titulo(directora).
titulo(consejera).
titulo(embajadora).
titulo(ministra).
titulo(capitana).
titulo(jueza).
titulo(secretaria).
titulo(periodista).
titulo(fundadora).
titulo(tupamara). %léxico uruguayo
titulo(secretaria).
% titulo(directiva). (muy ambiguo)
titulo(maestra).
titulo(alumna).
titulo(madre).
titulo(oradora).
titulo(investigadora).
titulo(jefa).
titulo('técnica'). % sólo cuando es N
titulo(activista).
titulo(magistrada).
titulo(candidata).
titulo('psicóloga').
titulo('correligionaria').
```

% gentilicios (si son adjetivos no necesito versión femenina, si no, sí)

```
titulo(argentino).
titulo(uruguayo).
titulo(latinoamericano).
```

% lemas para fuente no persona: grupos, instituciones, publicaciones

```
institucion(municipalidad).
institucion(colectividad).
institucion(partido). % ambiguo
institucion(sector). % ambiguo
institucion(escuadra). %ambiguo
institucion('institución').
institucion(instituto).
institucion(diario).
institucion(revista).
institucion('publicación').
institucion(organismo).
institucion(parlamento).
institucion(sindicato).
institucion('hospital').
institucion('semanario').
```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% predicado noNombreOpinion para regla que arma <gn de gn>
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

noNombreOpinion(Lema):-
    \+ nombreOpinion(Lema,_,_,_).

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% predicado noVerboOpinion para regla que arma <N + Part>
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

noVerboOpinion(Lema):-
    \+ verboOpinion(Lema,_,_,_,_).

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% pronombres sujeto
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

pronombreSujeto(yo) :- !.
pronombreSujeto('tú') :- !.
pronombreSujeto(usted) :- !.
pronombreSujeto('él') :- !.
pronombreSujeto(ella) :- !.
pronombreSujeto(nosotros) :- !.
pronombreSujeto(vosotros) :- !.
pronombreSujeto(ustedes) :- !.
pronombreSujeto(ellos) :- !.
pronombreSujeto(ellas) :- !.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% para desambiguación de comillas
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

ocurrenciaParImpar(TipoActual):-
    comilla(TipoAnterior),
    calculoActual(TipoAnterior,TipoActual),
    retract(comilla(_)),
    assert(comilla(TipoActual)).
calculoActual(abre,cierra):-!.
calculoActual(cierra,abre):-!.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% lista de verbos soporte
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

verboSoporte(elevar) :- !.
verboSoporte(realizar) :- !.
verboSoporte(efectuar) :- !.
verboSoporte(fabricar) :- !.
verboSoporte(arriesgar) :- !. % una opinión
verboSoporte(presentar) :- !. % un argumento
verboSoporte(brindar) :- !. % brindar un informe
verboSoporte(pronunciar) :- !.
verboSoporte(emitir) :- !.
verboSoporte(formular) :- !.

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% reglas de concordancias para nombres propios (vienen de FreeLing con el
% valor '0' en el atributo persona).
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

concordNumPer(NumGN,NumV,PerGN,PerV):-
    concordNum(NumGN,NumV),
    concordPer(PerGN,PerV,NumV). % paso el num del verbo por el caso especial

concordNum('0',_) :- !.
concordNum(N,N) :- !.
concordNum('N',_) :-!. % gn con número invariable

concordPer(P,P,_) :- !.
% caso especial: "yo quería" / "él quería" 1 sing - 3 sing

concordPer('1','3','S') :- !.
% acá entran todos los gn que no son pronombres y tienen en este atributo el
% género y no la persona
concordPer(PerGN,'3',_) :-
    PerGN \= '1',
    PerGN \= '2'.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% reglas para preposiciones que introducen el asunto
% se incluye las que terminan en contracción (al, del)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% polaridad neutra
prepAsunto(sobre,neu) :- !.
prepAsunto('respecto a',neu) :- !.
prepAsunto('respecto al',neu) :- !.
prepAsunto('respecto de',neu) :- !.
prepAsunto('respecto del',neu) :- !.
prepAsunto('con respecto a',neu) :- !.
prepAsunto('con respecto al',neu) :- !.
prepAsunto('con respecto de',neu) :- !.
prepAsunto('con respecto del',neu) :- !.
prepAsunto('con relación a',neu) :- !.
prepAsunto('con relación al',neu) :- !.
prepAsunto('en cuanto a',neu) :- !.
prepAsunto('en cuanto al',neu) :- !.

% las siguientes se agregaron como locuciones en el archivo de FreeLing

prepAsunto('en lo concerniente a',neu) :- !.
prepAsunto('en lo concerniente al',neu) :- !.
prepAsunto('en lo que concierne a',neu) :- !.
prepAsunto('en lo que concierne al',neu) :- !.
prepAsunto('en lo referente a',neu) :- !.
prepAsunto('en lo referente al',neu) :- !.
prepAsunto('en lo que refiere a',neu) :- !.
prepAsunto('en lo que refiere al',neu) :- !.
prepAsunto('en lo que respecta a',neu) :- !.
prepAsunto('en lo que respecta al',neu) :- !.

```

```

prepAsunto('en relación a',neu) :- !.
prepAsunto('en relación al',neu) :- !.
prepAsunto('a propósito de',neu) :- !.
prepAsunto('a propósito del',neu) :- !.

% polaridad positiva
prepAsunto('a favor de',pos) :- !.
prepAsunto('en favor de',pos) :- !.
prepAsunto('a favor del',pos) :- !.
prepAsunto('en favor del',pos) :- !.
%prepAsunto(por,pos) :- !. MUY ambigua se elimina de acá, pensar reglas
especiales

% polaridad negativa
prepAsunto(contra,neg) :- !.
prepAsunto('en contra de',neg) :- !.
prepAsunto('en contra del',neg) :- !.
% prepAsunto('contrario a',neg) :- !. No es locución de freeling, se hace
regla aparte

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% preposiciones que introducen la fuente
% (elementos de atribución de autoría o predicados preposicionales)
% locuciones incluidas en 'locucionsPLN.dat'
% se incluye las que terminan en contracción (al, del)
% no llevan orient. sem., son todas neutras
% no se incluye "para" porque requiere punto final previo, se hace con una
regla aparte
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

prepAutor('según') :- !.
prepAutor('de acuerdo con') :- !.
prepAutor('de acuerdo a') :- !.
prepAutor('de acuerdo al') :- !.
prepAutor('a juicio de') :- !.
prepAutor('a juicio del') :- !.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%% Polaridad Opinion
%%% parámetros: PolFue, PolPred, PolAsu, PolMens, PolOp
%%% el valor que se calcula es PolOp

calculoPolOp(_,PolPred,_,_,(PolPred)):-
    PolPred \= neu,
    !.
calculoPolOp(_,_,PolAsu,_,(PolAsu)):-
    PolAsu \= neu,
    !.
calculoPolOp(PolFue,_,_,_,(PolFue)):-
    PolFue \= neu,
    !.
calculoPolOp(_,_,_,PolMens,(PolMens)):-!.

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% incorporación de mensaje a
% opinión anterior
% calcular polaridad

incorporacionMensaje(pos,_,pos):-!.
incorporacionMensaje(neg,_,neg):-!.
incorporacionMensaje(_,_,nd):-!.           % si la opinión es neutra, al
agregarle un mensaje, el valor final es nd

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%% elementos subjetivos especiales que determinan la orientación subjetiva
de la fuente

elemSubjFuente('optimista',pos):-!.
elemSubjFuente('entusiasta',pos):-!.
elemSubjFuente('seguidor',pos):-!.
elemSubjFuente('partidario',pos):-!.
elemSubjFuente('pesimista',neg):-!.
elemSubjFuente('decepcionado',neg):-!.
elemSubjFuente('detractor',neg):-!.

```