

Textos subjetivos y objetivos: un estudio sobre reconocimiento de subjetividad en corpus

Antonio Reyes, Sonia Morett

Instituto Superior de Intérpretes y Traductores,
Laboratorio de Tecnologías Lingüísticas, México DF,
México

{antonioreyes,soniamorett}@isit.edu.mx

Resumen. En este artículo se describe un estudio exploratorio en el que se analiza cómo la subjetividad es expresada a través de elementos léxicos (subjetivemas) en un corpus de artículos periodísticos en español. En particular, dos tipos de textos constituyen el corpus: la nota informativa (texto expositivo) y la columna de opinión (texto argumentativo). El objetivo es identificar elementos que, en tanto subjetivos, son utilizados por los hablantes para persuadir de manera subyacente a los lectores, independientemente del tipo de género al que pertenece el texto. Los resultados hasta el momento han permitido reconstruir un panorama amplio de los subjetivemas que, sin importar el tipo de discurso, los hablantes emplean al emitir valoraciones subjetivas. El reconocimiento de estos elementos constituye un recurso relevante para tareas tales como análisis de sentimientos, minería de opiniones o reconocimiento automático de lenguaje persuasivo.

Palabras clave: Subjetividad, subjetivema, reconocimiento automático de subjetividad, lenguaje persuasivo.

1. Introducción

De acuerdo con [6], el término subjetividad no debe entenderse necesariamente como lo opuesto a objetividad, sino como una estrategia lingüística que permite a los hablantes involucrarse y verse reflejados en las acciones que codifican por medio del lenguaje. En este escenario, el presente trabajo describe la identificación de elementos léxicos subjetivos en español que dan cuenta de cómo los hablantes imprimen en sus enunciados sus posturas éticas, estéticas y, en definitiva, ideológicas, sobre diferentes tópicos a partir de un corpus integrado por dos tipos de textos: notas informativas, que tradicionalmente son consideradas como textos objetivos, y columnas de opinión, que a su vez, son consideradas como textos subjetivos.

Conceptualmente, este trabajo se cimienta en la noción de subjetivema introducida por [10] en la teoría de la enunciación, así como en los aportes realizados al estudio del léxico valorativo desde la teoría de la valoración ([17,13]) inserta en

el marco de la Lingüística Sistémico Funcional. Ahora bien, por subjetivema se entiende aquella forma o elemento léxico que porta una carga de subjetividad a los enunciados respecto a lo narrado o descrito. En este sentido, los subjetivemas se emplean para externar estados emocionales o valoraciones y apreciaciones de tipo cognitivo hacia personas, objetos, hechos o situaciones. Así, estos pueden corresponderse con *entidades* (parsimonia, apremio, inteligencia, torpeza), *cualidades* (grande, lejano, limpio, sucio, correcto, deshonesto), *procesos* (amar, odiar, susurrar, vociferar) u *opiniones* (juiciosamente, afortunadamente, lastimosamente). Dada esta heterogeneidad, los subjetivemas no son dependientes de categoría gramatical alguna; hecho que los vuelve difíciles de sistematizar y reconocer automáticamente. Aunado a lo anterior, los subjetivemas (ya sean adjetivos, adverbios, sustantivos o verbos) pueden ser o bien afectivos o bien evaluativos. En lo que respecta a los subjetivemas afectivos, estos dan cuenta de las reacciones emocionales tales como enfado, miedo, tristeza, alegría o sorpresa; en tanto que los subjetivemas evaluativos se relacionan con la transmisión de juicios de valor y apreciaciones de otra índole.

Nuestro objetivo, por tanto, consiste en i) identificar elementos concretos que son utilizados por los hablantes para codificar lingüísticamente la subjetividad en texto; y ii) determinar si este conjunto de elementos que, en tanto subjetivemas, son usados para transmitir valoraciones y apreciaciones subjetivas de manera directa e indirecta. Para lograrlo, se ha construido un corpus de artículos periodísticos integrado por dos tipos de textos: notas informativas (texto expositivo) y columnas de opinión (texto argumentativo). Los resultados han sido evaluados con una tarea de reconocimiento automático de subjetividad.

El resto del artículo está organizado de la siguiente forma: en la Sección 2 se introduce el marco conceptual que sustenta este trabajo. En la Sección 3 se describe el proceso para la identificación de subjetivemas; en tanto que en la Sección 4 se presenta la evaluación que se realizó utilizando el corpus de artículos periodísticos. Los resultados se detallan en la Sección 5. Finalmente, en la Sección 6 se sintetizan las principales conclusiones y se delinea el trabajo futuro.

2. Marco conceptual

En términos léxicos, el Diccionario de la Real Academia Española (DRAE) define *objetivo* como aquello que es “perteneciente o relativo al objeto en sí mismo, con independencia de la propia manera de pensar o de sentir”; en tanto que lo *subjetivo* es aquello que es filtrado por el sujeto, es decir, lo “perteneciente o relativo a nuestro modo de pensar o de sentir, y no al objeto en sí mismo” (cf. DRAE [2001]). Dada esta división, lo objetivo siempre será distinto a lo subjetivo. Sin embargo, si concebimos el lenguaje no sólo como un sistema rígido y perfectamente definido (tal cual se infiere de las definiciones previas), sino como un sistema de opciones semánticas, donde el hablante elige, dentro de todas las posibilidades a su alcance, aquéllas que mejor satisfacen sus necesidades comunicativas, entonces la división que se propone en el DRAE no es funcional,

puesto que el proceso de producción de los enunciados se inicia con la elección de los significados para, posteriormente, desembocar en la selección de la estructura discursiva y el léxico más conveniente para la transmisión de aquello que desea comunicar, tanto explícita como implícitamente.

En el escenario que supone esta visión más funcional, la interacción social a través del lenguaje implica una constante toma de decisiones que dependen de varios factores lingüísticos y extralingüísticos. No obstante, caer en el relativismo de afirmar que todo en el lenguaje es subjetivo conduce a posiciones que niegan la posibilidad de acceder al análisis crítico del discurso o de un simple acto de habla. En tal sentido, [10] reconoce como simplista y reduccionista señalar que en el lenguaje, la subjetividad está en todas partes; antes bien, es más productivo y realista reconocer que todo acto comunicativo lleva la marca de su hablante en modos y grados diversos. Vista desde esta perspectiva, la subjetividad recae particularmente en el proceso de la enunciación y en el contexto en la que ésta se produce¹. Por lo tanto, hablar de subjetividad implica el desarrollo de una metodología que permita reconocer y sistematizar qué elementos formales están presentes al momento de la enunciación.

2.1. Teoría de la enunciación

Al situarnos en el plano de la enunciación, la descripción lingüística deja de centrarse en el sistema para también considerar quién produce el acto comunicativo, con qué recursos lingüísticos y con qué intención. Respecto a los recursos lingüísticos de la enunciación, [1] señala como elementos indicadores de subjetividad la deixis personal, la espacial y la temporal. Asimismo, se refiere a recursos modales que dotan de este tipo de contenido a los enunciados y apunta que el significado léxico prototípico de algunas palabras lo transporta de manera implícita. A partir de estos indicadores de subjetividad, [10] desarrolla de manera detallada la teoría de la enunciación que focaliza los siguientes tipos de subjetividad: deíctica, modal y léxica.

Por un lado, de acuerdo con la autora, la manifestación más notoria de la subjetividad se observa por la presencia de unidades deícticas, las cuales dependen absolutamente del contexto de la enunciación, pues hacen patente la existencia de un hablante que se refiere a otro u otros y elabora su discurso desde un lugar y un tiempo determinados. En lingüística se acepta por consenso la existencia de tres categorías deícticas: la personal, la espacial y la temporal. [5] se refirió, además, a la existencia de otra categoría deíctica, la deixis social, que se expresa mediante marcas en los enunciados que actualizan el contexto social de la enunciación.

Por otro lado, la subjetividad modal suele describirse como la actitud del hablante respecto al contenido de su discurso. Tal actitud se expresa de modo transversal en diferentes niveles de análisis de la lengua: i) el estilo de la

¹ Es importante señalar que este contexto es bastante amplio pues no sólo se concibe como un contexto exclusivamente lingüístico, sino como un contexto que, además, se compone de factores sociales, culturales, económicos, etc.

enunciación (directo o indirecto); ii) los modos gramaticales (indicativo, subjuntivo e imperativo); iii) la intención ilocutiva de los actos de habla (asertiva, directiva, comisiva, expresiva, performativa); iv) el modo de aserción (constativo, hipotético, obligatorio); v) la polaridad del acto de habla (atenuada, intensificada o neutra); etc.

Finalmente, la subjetividad léxica se define a partir de la presencia de palabras que en el plano discursivo funcionan como catalizadores de información valorativa y/o apreciativa. Particularmente, este tipo de subjetividad es la que le da sentido al concepto de subjetivema que se introdujo en la Sección 1. En relación con este tipo de subjetividad y con el concepto mismo de subjetivema, [10] señala que si bien “toda unidad léxica es, en un cierto sentido, subjetiva, dado que las *palabras* de la lengua no son jamás otra cosa que símbolos sustitutivos e interpretativos de las *cosas*”, existen palabras o expresiones que portan una mayor carga de subjetividad respecto a lo narrado o descrito, las cuales se emplean para externar estados emocionales o valoraciones y apreciaciones de tipo cognitivo hacia personas, objetos, hechos o situaciones.

A partir del concepto de subjetivema, expuesto en los párrafos previos, se aborda la problemática de la subjetividad en texto en este trabajo; puesto que el subjetivema, como entidad formal, permite delimitar la subjetividad en términos de elementos lingüísticos concretos. En este sentido, en la siguiente sección se describe la metodología que se siguió para identificar un conjunto de subjetivemas propios del español.

3. Identificación de subjetivemas

Como se describió en la Sección 1, no es posible relacionar los subjetivemas con una categoría gramatical en específico. Dada la información que los hablantes codifican a través de ellos, estos pueden pertenecer a la categorías sustantivo, adjetivo, adverbio y verbo. Aunque para el caso de los adverbios (en específico los terminados en *-mente*) y los adjetivos se han identificado ciertos patrones morfológicos, tales como los prefijos *anti-*, *contra-* y los sufijos *-ucho*, *-ete*, que al unirse a una base léxica neutra le aportan a su significado un componente evaluativo o afectivo, en general, no hay elementos formales que distingan a los subjetivemas de otras formas léxicas, lo que dificulta su reconocimiento tanto manual como automático. Así, [10] advierte que “los subjetivemas constituyen un conjunto de hechos demasiado heterogéneos para que se pueda esperar elaborar tan pronto un procedimiento de cálculo del porcentaje de subjetividad que implica un texto cualquiera”. No obstante, “su descripción permite ya, en cierta medida, evaluar comparativa, cualitativa y cuantitativamente, el funcionamiento enunciativo de dos ítems, secuencias, unidades o conjuntos textuales determinados”.

El desarrollo de una metodología que permita su identificación, por tanto, supone un acercamiento a la sistematización de aspectos de la lengua muy relevantes en el terreno del Procesamiento del Lenguaje Natural (PLN) en tareas tales como análisis de sentimientos y minería de opiniones (véanse los traba-

jos de [18,19,12]), procesamiento de lenguaje figurado ([15,3,8]), o persuasión automática ([9,7,16]). No obstante, desarrollar dicha metodología implica un proceso totalmente objetivo que permita el reconocimiento de patrones que puedan generalizarse, sobre todo, si se piensa en la automatización del proceso de identificación. En este sentido, a continuación se describe la metodología que se utilizó para la identificación y extracción de subjetivemas.

3.1. Extracción de subjetivemas

De acuerdo con lo descrito en las secciones previas, los subjetivemas pueden pertenecer a varias categorías gramaticales; por tanto, el primer paso para su extracción consistió en determinar una categoría de prueba que nos permitiera i) probar la eficacia de la metodología para identificar y extraer subjetivemas, y ii) evaluar de manera empírica que los subjetivemas extraídos funcionan realmente como catalizadores de información subjetiva. En este sentido, se optó por trabajar con la categoría **adjetivo**. Esta elección está en relación con el hecho de que los adjetivos son las piezas léxicas valorativas por excelencia y, en consecuencia, las que mayor carga de subjetividad pueden imprimir a los enunciados. Además, como apunta [2], se trata de una categoría semántica, puesto que hay significados exclusivos del dominio de los adjetivos, por ejemplo, relaciones de diferentes tipos, así como valoraciones y apreciaciones que precisan, explican, caracterizan, clasifican o subvierten el significado básico del sustantivo al que caracterizan; i.e. los adjetivos dan cuenta de propiedades semánticas que pueden ser aplicadas a todos los constituyentes de nuestra realidad. Ahora bien, de las diferentes clases de adjetivos que existen en nuestra lengua (calificativos, determinativos, subsectivos, intersectivos, etc.), se consideró solamente la de adjetivos calificativos puesto que esta clase contiene adjetivos subjetivos y no subjetivos relacionados con la caracterización de atributos humanos.

Una vez establecido que se trabajaría con adjetivos calificativos, el siguiente paso consistió en extraer un conjunto de adjetivos que, de acuerdo con información lexicográfica, pudieran ser considerados como subjetivemas². Para ello, se seleccionaron dos diccionarios como base de conocimiento: el DRAE y el Diccionario [11]. En cada diccionario se realizó una búsqueda a partir de lo que en lexicografía se conoce como marcas, es decir, etiquetas que declaran los atributos de la palabra desde diferentes puntos de vista (categoría gramatical, etimología, extensión geográfica de uso, etc.). Para este trabajo nos centramos

² De acuerdo con los registros del DRAE, el español cuenta con más de 170 mil acepciones correspondientes a adjetivos y sustantivos con usos adjetivales. Si consideramos que estos registros corresponden fundamentalmente al habla peninsular y que la lengua es un sistema dinámico, donde interviene el factor del tiempo y el desplazamiento entre categorías gramaticales, podemos suponer que este número es una referencia conservadora. Por tanto, el hecho de focalizar el concepto de subjetivema en un conjunto nuclear de adjetivos calificativos responde al hecho de que es simplista señalar que todas esas acepciones y, en consecuencia, el resto de elementos léxicos de una lengua, son subjetivos.

en las marcas que se refieren a particularidades sociales de uso. A continuación se sintetizan los pasos para la extracción:

- a) **DRAE**: Se realizó una búsqueda a partir de los criterios *adjetivo* (marca gramatical) y *despectivo* (marca pragmática). El resultado fue un conjunto de 222 adjetivos.
- b) **DIRAE**: Es un desarrollo de Gabriel Rodríguez que aprovecha toda la información del DRAE para realizar búsquedas inversas³. En esta versión se efectuó una búsqueda con los criterios *adjetivo* (marca gramatical) y *dicho de una persona* (marca de uso). El resultado fue un conjunto de 1644 adjetivos.
- c) **LEMA**: Para este diccionario se hizo una búsqueda utilizando los criterios de *adjetivo* (marca gramatical) y *despectivo, malsonante u ofensivo* (marca pragmática). El listado resultante fue de 51 adjetivos.

Efectuado este proceso de búsqueda, el total de candidatos a subjetivemas fue de 1,197 adjetivos. De este conjunto se hizo una depuración al desestimar adjetivos repetidos y adjetivos que se emplean para calificar situaciones u objetos y no a personas; así como a la supresión de los adjetivos conocidos como intencionales (verdadero, certero, falso, supuesto, principal) debido a que se refieren a la modalidad o toma de posición del hablante frente al referente de su discurso, lo que en la teoría de la enunciación se corresponde con la subjetividad modal (véase 2.1). El conjunto final de candidatos a subjetivemas fue de 752 adjetivos. Algunos ejemplos se enlistan en la Tabla 1.

Tabla 1. Ejemplo de candidatos a subjetivemas.

DRAE	DIRAE	LEMA
absuelto	santurrón	solterona
naco	paternal	subversivo
discapacitado	puritano	deschavetado
morboso	tacaño	incapaz
impúdico	caritativo	calenturiento
malhablado	humanitario	soplón

4. Evaluación

A partir del conjunto de 752 adjetivos, el siguiente paso consistió en evaluar qué tan bien funcionaban como subjetivemas. Esto implicó realizar una tarea de reconocimiento automático de subjetividad en texto. Para ello, se creó un corpus de artículos periodísticos que fueron representados en términos de presencia/ausencia de subjetivemas con el fin de recuperar los textos que más

³ Disponible en <http://dirae.es>.

probabilidad tuvieran de ser subjetivos. Posteriormente, los resultados fueron validados en una tarea de reconocimiento de subjetividad realizada por lectores humanos. A continuación se detallan las tareas realizadas.

4.1. Corpus

El corpus de trabajo se creó a partir de la recolección de artículos periodísticos en español. Se optó por considerar el lenguaje periodístico debido a que los comunicadores no pueden sustraerse de la visión y exposición parcial sobre los hechos que documentan, propiciando así la presencia de elementos subjetivos. En este sentido, siguiendo a [4], “el sujeto informa, y en este mismo acto interpreta y transforma desde su visión del mundo la realidad social y concreta que es objeto de la información; y es esa misma realidad y la relación entre los sujetos la que condiciona dicha visión del mundo”. Aunado a lo anterior, la prensa es muy apreciada como fuente de información léxica, en tanto que puede concebirse como poseedora de repertorios lingüísticos con actualización diaria que regulan las comunicaciones del sector de la sociedad a la que se destina cada publicación.

Ahora bien, en tanto que nuestro interés por el reconocimiento automático de subjetividad en texto tiene por propósito futuro confrontar el léxico subjetivo frente al objetivo, el corpus se conformó con textos que, en términos de géneros periodísticos, fueran representativos de ambos tipos de léxico: la **nota informativa** (como paradigma de texto objetivo puesto que pertenece al género expositivo) y la **columna de opinión** (que supone ejemplos prototípicos de textos subjetivos puesto que se clasifica dentro del género argumentativo).

Un total de 40 artículos constituyen el corpus: 20 corresponden a notas informativas y 20 a columnas de opinión. Los textos fueron recuperados automáticamente de los periódicos mexicanos El Universal y La Jornada en sus versiones online. Con el fin de asegurar una evaluación objetiva, se decidió que todos los artículos tuvieran como eje temático la misma noticia: el conflicto magisterial en México entre 2013 y 2014. Ningún otro criterio fue tomado en cuenta. En la Tabla 2 se presentan algunos datos estadísticos del corpus.

Tabla 2. Estadísticas generales del corpus.

	NOTA INFORMATIVA	COLUMNA DE OPINIÓN	TOTAL
Textos Universal	10	10	20
Textos Jornada	10	10	20
Types	3,621	3,810	6,103
Tokens	15,729	15,646	31,375
Oraciones	530	562	1,092

4.2. Recuperación automática de textos subjetivos

El proceso de recuperación automática de textos subjetivos consistió en identificar qué textos tenían mayor probabilidad de ser subjetivos, independientemente del género al que pertenecieran (nota informativa o columna de opinión). La tarea, en este sentido, se realizó por medio de la caracterización de los textos del corpus por medio de los subjetivemas extraídos del DRAE y del diccionario Lema. En primer lugar, tanto el corpus como los subjetivemas fueron lematizados con el fin de que las marcas morfológicas de género, número y apreciativo no incidieran ni sesgaran la caracterización. Para ello, se utilizó el algoritmo de [14] en su adaptación al español. En segundo lugar, se eliminaron todas las stopwords. Finalmente, cada texto del corpus se caracterizó de acuerdo con la Fórmula 1:

$$subjetividad_{(t)} = \frac{\sum s_1}{|t|} \quad (1)$$

donde s_1 hasta s_n es el conjunto de adjetivos que funcionan como subjetivemas y $|t|$ es la longitud de cada texto del corpus. Es importante destacar que en esta caracterización no importó si el subjetivema se repetía n veces o sólo aparecía una vez. Al final, todas las frecuencias se sumaron y normalizaron por la longitud del texto para asegurar que el tamaño no incidiera en el resultado. Para finalizar el proceso, los textos cuyo valor fue más alto se consideraron como subjetivos. Los resultados se describen en la Sección 5.

4.3. Validación humana

Con el fin de contrastar los resultados de la aproximación automática con las percepciones que los humanos tienen respecto a lo que se entiende por subjetividad, se realizó una validación piloto sobre los resultados. Para ello, se seleccionaron tres hablantes de español de México. Dos de ellos con formación académica de licenciatura y uno más de maestría. Todos hablantes nativos de español e informados sobre el conflicto magisterial. A cada uno se le dieron 2 textos: el más subjetivo y el menos subjetivo, según los resultados del proceso automático. Una columna de opinión y una nota informativa, respectivamente. Los títulos de cada artículo son *La cruzada moral contra el magisterio*, para el primero y *Nos tachan de revoltosos por defender la educación pública*, para el segundo.

La tarea para cada lector humano consistió en i) leer ambos textos y determinar cuál era subjetivo y cuál, objetivo; y ii) responder un cuestionario en el que aportaran los elementos que cada uno consideró para tomar su decisión respecto a lo que hicieron en i). A ninguno se le dijo qué debía entender por cada concepto ni se le informó acerca de la procedencia del artículo. Sólo se les dijo que decidieran en función de lo que ellos consideraban como subjetivo y objetivo.

Los resultados se presentan en la siguiente sección.

5. Resultados

A pesar de que el conjunto de subjetivemas que se extrajo parecía muy pequeño en tanto las muchas posibilidades que tiene una lengua para expresar información subjetiva (sólo 752 adjetivos), los resultados obtenidos con el proceso automático muestran datos interesantes. Así, de los 40 textos que integran el corpus de trabajo, todos ellos, según lo observado, contienen elementos con carga de información subjetiva. Obviamente, la presencia de subjetivemas no es constante en todos los textos. Por ejemplo, hay textos con alrededor de 900 tokens, de los cuales más de 80 corresponden a subjetivemas; i.e. casi el 10% del contenido es subjetivo. Por el contrario, hay otros textos que en ese mismo promedio de tokens contienen menos de 10 subjetivemas.

Como síntesis del proceso automático es importante destacar que alrededor del 42% de los textos del corpus (i.e. 17) tiene contenido suficiente para ser considerado como texto subjetivo. Esto se deduce del porcentaje de subjetividad que se obtiene al aplicar la Fórmula 1. Así, resulta que si definimos un umbral mínimo de subjetividad $\geq .3$, entonces 17 textos caen en la categoría de texto subjetivo dado su alto contenido de subjetivemas.⁴ Ahora bien, cabe destacar que de estos 17 textos, 5 pertenecen a la categoría nota informativa y los 12 restantes a la columna de opinión (claro está, que estos últimos ocupan los primeros lugares de subjetividad).

En relación con la validación humana, en la Tabla 3 se sintetizan los resultados que arrojó el análisis de los lectores humanos. De acuerdo con estos datos, se nota que entre los humanos no hay consenso respecto a lo que cada uno percibe como subjetivo u objetivo. Para la columna de opinión, que en esencia supone la presencia de valoraciones personales muy claras que lo convierten en subjetivo, dos de los lectores lo consideran como un texto objetivo; mientras que la nota informativa, que supone lo contrario, es categorizada por dos lectores como texto subjetivo.

Tabla 3. Resultados de la validación humana.

Texto	Modelo	Lector 1	Lector 2	Lector 3
Opinión	<i>Subjetivo</i>	Subjetivo	Objetivo	Objetivo
Nota	<i>Objetivo</i>	Subjetivo	Objetivo	Subjetivo

5.1. Discusión

Para entender este comportamiento, se analizaron las evidencias que aportó cada lector en torno al porqué de sus decisiones. Así, para la pregunta explícita

⁴ Este umbral se definió tomando en cuenta que lo único que reconoce el modelo, por ahora, son adjetivos. Por tanto, pensar en un umbral mayor supondría un texto incoherente que en su mayoría debería estar escrito con adjetivos.

sobre qué palabras del texto justifican la subjetividad u objetividad, estos indicaron algunos adjetivos que se correspondían con los que nosotros habíamos detectado como subjetivemas (disidente, revoltoso, indolente), además de otros que no se habían considerado (estandarizado, enraizado, gris); pero para la pregunta sobre lo que consideraban como recursos empleados por el periodista para mostrar dicha tendencia, los tres respondieron con argumentos que denotaban que su proceso de análisis no estuvo exento de valoraciones subjetivas. Por ejemplo, el lector 2 señaló que valoró la columna de opinión como objetiva porque *“el autor trata de explicar el origen y las razones de la protesta para mostrar su parte racional que es escondida en el vituperio y en el clima de crispación mediática que se ha creado interesadamente contra el magisterio y su protesta”*; en tanto, el lector 1 indicó que valoró la nota informativa como subjetiva porque *“hace ver a los maestros como mártires de la gente que está cansada de los inconvenientes que provocan sus marchas”*.

Esta falta de acuerdo entre los humanos, además de demostrar lo complejo de la tarea, pone en evidencia los siguientes puntos:

1. El proceso para la selección de subjetivemas responde de manera adecuada al proceso de representación de subjetividad en texto. No obstante, es importante explorar con otras categorías y con otro tipo de recursos que puedan integrar menos información léxica y más de tipo pragmática. Al respecto, hay un dato muy importante que sustenta esto: la información de la Tabla 2, sobre todo la relativa a los types, da indicios de que tanto notas como opiniones comparten bastante léxico. Así, lo que propicia que un texto se perciba como subjetivo está en relación con factores tales como la estructura sintáctica o la percepción, totalmente pragmática, que los humanos tienen sobre un hecho o periódico determinado.
2. La forma de representar la subjetividad en texto es operativa pero tiende a ser muy general. Si bien partimos de elementos cuantitativos, hay que perfeccionar los elementos cualitativos. Por ejemplo, la presencia de un solo subjetivema puede ser suficiente para hablar de un texto subjetivo.
3. Los textos que elegimos para probar los subjetivemas demuestran que para implementar un modelo de subjetivemas en escenarios reales es necesario partir de información formal, como los géneros periodísticos, que permita entender cómo funciona la transmisión explícita e implícita de valoraciones subjetivas. En este sentido, por ejemplo, nuestra intuición apuntaba a que serían los textos pertenecientes a la columna de opinión los que, dadas sus características, tenderían a ser los únicos subjetivos. No obstante, es interesante notar que, tanto en el proceso automático como en el humano, esta intuición no se corresponde totalmente con la realidad.
4. Aunado al punto anterior, es claro que las evidencias de subjetividad en los textos dependen del género discursivo de que se trate. Así pues, el discurso expositivo se vale de un léxico técnico o del uso de citas textuales para dar credibilidad, igualmente, recurre a la tercera persona como marca implícita de distanciamiento del hecho y, en consecuencia, de objetividad. En este

sentido, es importante considerar información tipográfica y sintáctica para complementar un modelo de reconocimiento de textos subjetivos.

5. Es indudable que los humanos tienen que participar en la evaluación de todo sistema automático; sin embargo, a partir de nuestros resultados nos queda claro que debemos explorar más opciones para que la valoración humana sea realmente efectiva. Por ejemplo, probar con lectores que desconozcan la noticia para no sesgar su análisis. Además, claro está, de tener un conjunto más amplio de evaluadores y de textos, posiblemente más cortos, que permitan perfeccionar el modelo.

Finalmente, es importante destacar que un modelo que permita reconocer documentos subjetivos basado en información fina, como es el caso de los subjetivemas, permitirá enfrentar problemas relacionados con el PLN desde un ángulo más complejo y abarcador.

6. Conclusiones y trabajo futuro

En este artículo se ha presentado una aproximación al procesamiento automático de subjetividad en texto. A partir del concepto de subjetivema se ha tratado de representar el proceso que los hablantes, a través de las palabras, realizan para emitir valoraciones que pueden clasificarse como subjetivas. En este sentido, se ha descrito una metodología para reconocer y extraer subjetivemas por medio de patrones lexicográficos. En particular, la metodología se probó con la categoría gramatical adjetivo. La relevancia del conjunto de subjetivemas fue evaluado en términos de una tarea de recuperación automática de textos subjetivos. Para ello, se creó un corpus de artículos periodísticos constituido por textos considerados como subjetivos y objetivos. Los resultados de la tarea, además, fueron valorados por tres lectores humanos.

Las conclusiones a las que llegamos apuntan a que el método de representación de subjetividad en texto a partir de subjetivemas es operativo. No obstante, para crear un modelo más robusto, es necesario considerar sustantivos y verbos, expandir la base de adjetivos con sinónimos y, posiblemente, antónimos, así como elementos tipográficos e información sintáctica con el fin de asegurar que el modelo no recaiga exclusivamente en información léxica sino que, además, incorpore información pragmática.

Estos aspectos serán abordados a detalle en trabajos futuros. Asimismo, dado nuestro interés por tareas y aplicaciones a las que subyazca el tratamiento de contenido subjetivo, será interesante estudiar el reconocimiento automático de estrategias pragmáticas complejas que se manifiestan en el nivel discursivo; tales como la identificación de usos tendenciosos del lenguaje o el reconocimiento automático de léxico persuasivo.

Referencias

1. Benveniste, E.: De la subjetividad en el lenguaje. In: Problemas de lingüística general I, pp. 179–187 (1982)

2. Catena, A.: Contribución a la formalización del adjetivo para la traducción automática español-francés. Ph.D. thesis, Universitat Autònoma de Barcelona (2006)
3. Davidov, D., Tsur, O., Rappoport, A.: Semi-supervised recognition of sarcastic sentences in Twitter and Amazon. In: Proceedings of the Fourteenth Conference on Computational Natural Language Learning. pp. 107–116. CoNLL '10, Association for Computational Linguistics, Stroudsburg, PA, USA (2010)
4. Espeche, C.: Periodismo objetivo o subjetivo. una falsa dicotomía. In: Congreso de periodismo y medios de comunicación. Buenos Aires, Argentina (2012)
5. Fillmore, C.: Lectures on deixis (1997)
6. Fournet, A.: Michel Bréal (1832–1915), a forgotten precursor of enunciation and subjectivity. *ReVEL* 9(16) (2011)
7. Gatti, L., Guerini, M., Stock, O., Strapparava, C.: Sentiment variations in text for persuasion technology. In: Proceedings of the 9th International Conference on Persuasive Technology (PERSUASIVE 2014). Padua, Italy (2014)
8. González-Ibáñez, R., Muresan, S., Wacholder, N.: Identifying sarcasm in Twitter: A closer look. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers - Vol. 2. pp. 581–586. The Association for Computer Linguistics (2011)
9. Guerini, M., Ozbal, G., Strapparava, C.: Echoes of persuasion: The effect of euphony in persuasive communication. In: Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-2015). Denver, Colorado (2015)
10. Kerbrat-Orecchioni, C.: L'énonciation. De la subjectivité dans le langage. Armand Colin, Paris (1980)
11. Lema: Diccionario de la lengua española. versión en cd room (2001)
12. Liu, B.: Sentiment analysis and subjectivity. In: Indurkha, N., Damerau, F.J. (eds.) Handbook of Natural Language Processing, Second Edition. CRC Press, Taylor and Francis Group (2010)
13. Martin, J., White, P.: The Language of Evaluation. Appraisal in English. Palgrave Macmillan, London (2005)
14. Porter, M.: An algorithm for suffix stripping. *Program* 14(3), 130–137 (1980)
15. Reyes, A., Rosso, P., Veale, T.: A multidimensional approach for detecting irony in Twitter. *Language Resources and Evaluation* 47(1), 239–268 (2013)
16. Strapparava, C., Guerini, M., Stock, O.: Predicting persuasiveness in political discourses. In: Proceedings of the 7th conference on International Language Resources and Evaluation (LREC'10). Valletta, Malta (2010)
17. White, P.: Un recorrido por la teoría de la valoración (teoría de la valoración) (2000)
18. Wilson, T., Wiebe, J., Hoffmann, P.: Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational Linguistics* 35(3), 1–35 (2009)
19. Zhang, L., Barnden, J., Hendley, R., Wallington, A.: Exploitation in affect detection in open-ended improvisational text. In: SST '06: Proceedings of the Workshop on Sentiment and Subjectivity in Text. pp. 47–54. Association for Computational Linguistics (2006)