

# La negación en español: análisis y tipología de patrones de negación\*

## *Negation in Spanish: analysis and typology of negation patterns*

**M. Antònia Martí, Mariona Taulé,  
Montserrat Nofre, Laia Marsó**  
CLiC- Universitat de Barcelona  
Gran Via 585, 08007, Barcelona  
{amarti, mtaule, mnofre}@ub.edu  
marso.laia@gmail.com

**M. Teresa Martín-Valdivia  
Salud María Jiménez-Zafra**  
Departamento de Informática  
Universidad de Jaén  
E-23071 – Jaén, España  
{maite, sjzafra}@ujaen.es

**Resumen:** En este artículo se presentan los criterios aplicados para la anotación del corpus *SFU Review<sub>SP</sub>-NEG* con negación y la tipología lingüística correspondiente. Esta tipología presenta la ventaja de ser fácilmente expresable en términos de un *tagset* para la anotación de corpus, de presentar tipos claramente delimitados, evitando así la ambigüedad en el proceso de anotación, y de presentar una amplia cobertura, es decir, que ha servido para resolver todos los casos que han aparecido. El corpus contiene 400 comentarios y 198.551 palabras. Actualmente está anotado en un 75% y, de un total de 6.331 oraciones revisadas, se han identificado 2.953 estructuras de negación.

**Palabras clave:** Negación, anotación de corpus, tipos de negación, análisis de opiniones, anotación de la polaridad

**Abstract:** In this paper we present the criteria applied for the annotation of the *SFU Review<sub>SP</sub>-NEG* corpus and the corresponding linguistic typology. This typology has the advantage that it is easy to express in terms of a *tagset* for corpus annotation: the types are clearly defined, which avoid the ambiguity in the annotation process, and they present a wide coverage (i.e. they covered/solved all the cases occurring in the corpus). The corpus consists of 400 reviews and 198,551 words. Currently, we have annotated 75% and from a total of 6,331 annotated sentences 2,953 contain at least one negation.

**Keywords:** Negation, scope, corpus annotation, sentiment analysis, polarity annotation

### *1 Introducción: Motivación*

En el marco del Procesamiento del Lenguaje Natural (PLN) el tratamiento de la negación ha cobrado un especial interés en la medida en que afecta directamente a la polaridad de los textos, en concreto los que expresan opiniones sobre artículos, productos, tendencias y servicios ((Pang et al., 2002), (Wiegand et al., 2010), (Polanyi y Zaenen, 2006), (Councill, McDonald y Velikovich, 2010) y (Morante and Sporleder, 2012)). El carácter idiosincrático de la expresión de la negación en cada lengua requiere un análisis lingüístico específico. Todo

proceso de anotación requiere una definición previa de los fenómenos que se van a anotar y una tipología de los mismos. Siendo la negación un fenómeno dependiente de la lengua, las tipologías sobre negación existentes para el inglés -con mucho la lengua en la que se han realizado más esfuerzos en el tratamiento de este fenómeno- no se pueden reutilizar para la anotación de corpus en otras lenguas.

En este artículo presentamos nuestra aproximación al tratamiento de la negación en un corpus del español, el *SFU Review<sub>SP</sub>* (Taboada et al., 2006). En concreto, se presenta una clasificación de las distintas maneras de expresar la negación en base a una tipología; se

\* Financiado por fondos FEDER, los proyectos: TIN2015-65136-C2-1-R y TIN2015-71147-C2-2 del MINECO y FPU014/00983 del MECD.

discuten los conceptos de foco, evento y alcance y se presenta el esquema general de anotación que estamos utilizando para la anotación del corpus antes mencionado.

El corpus *SFU Review<sub>SP</sub>* contiene 400 comentarios repartidos en 50 opiniones (la mitad positivas y la otra mitad negativas) de cada uno de los siguientes temas: coches, hoteles, lavadoras, móviles, ordenadores, música, libros y películas, extraídos de la página web Ciao.es.

En la sección 2, se presenta un estado de la cuestión en la anotación de corpus con negación y se sitúa el corpus que estamos anotando en este contexto; en las secciones 3 y 4 se define y delimita el concepto de negación que está en la base de nuestra anotación. En la sección 5 se describe la tipología desarrollada para la anotación del corpus *SFU Review<sub>SP</sub>*. En la sección 6 se presenta brevemente el esquema de anotación general y en la sección 7 se presentan las conclusiones y se apuntan las líneas futuras.

## **2 Antecedentes: corpus anotados con negación**

Los corpus anotados con negación disponibles, todos ellos del inglés, son de tamaño muy diverso (desde 20.000 a 1.000 oraciones anotadas) y difieren en el sistema de anotación empleado. Tienen en común que todos ellos anotan tanto las partículas negativas como el alcance; sólo uno de ellos, el de Blanco y Moldovan (2011), marca el foco y sólo el *ConanDoyle-neg* (Morante y Daelemans, 2012) marca el evento (véase la sección 4).

El primer corpus anotado con negación fue *BioInfer* (Pyysala et al., 2007), que incluye 1.100 oraciones extraídas de *abstracts* de artículos biomédicos. Se etiquetan los predicados con negación, pero no su alcance. Destacan por su tamaño los corpus *BioScope* (Vincze et al. 2008) y *SFU Review<sub>EN</sub>* (Konstantinova et al., 2012), que además de anotar la negación también incluyen la anotación de las expresiones especulativas y su alcance, información clave para identificar los enunciados subjetivos.

*BioScope* es un corpus formado por textos biomédicos en el que se anotaron por primera vez tanto las partículas negativas (y especulativas) como su alcance. El corpus contiene más de 20.000 oraciones anotadas, de las cuales el 13% incluye algún tipo de negación. *BioScope* está formado por textos

clínicos (6.383 oraciones), *abstracts* de artículos de biología (11.871 oraciones) y los 9 artículos completos de biología (2.670 oraciones) del corpus *Genia Event* (Kim et al., 2008).

Konstantinova et al., (2012) han anotado con negación (y especulación) el corpus *SFU Review<sub>EN</sub>*. Este corpus está formado por un total de 400 comentarios (17.263 oraciones de las cuales el 18% contienen negación) escritos en inglés y de distinta temática -comentarios sobre libros, coches, ordenadores, utensilios de cocina, hoteles, películas, música y teléfonos-extraídos de la página web Epinions.com. El corpus *SFU Review<sub>EN</sub>* contiene 50 documentos de cada una de las temáticas seleccionadas y cada uno de ellos tiene asignada una etiqueta que indica si se trata de un comentario positivo o negativo. Para la anotación de la negación siguen fundamentalmente los criterios utilizados en *BioScope* adaptados al dominio de los comentarios (Konstantinova y Sousa, (2011).

Entre los corpus de menor tamaño, en el mismo ámbito de los comentarios, cabe destacar el corpus *Product Review* (Councill, McDonald y Velikovich, 2010) formado por 268 comentarios de productos extraídos de *Google Product Search*. El corpus contiene 2.111 oraciones de las cuales 679 incluyen negación, es decir, el 32%. Los autores utilizan este corpus para desarrollar un sistema cuyo objetivo es identificar el alcance de la negación en el contexto del análisis de los sentimientos.

*ConanDoyle-neg* es el corpus de entrenamiento y evaluación desarrollado para la tarea 10 de SemEval-2010, *Linking events and their participants in discourse*<sup>1</sup> (Ruppenhofer et al. 2010). El corpus incluye textos literarios de dos obras de Arthur Conan Doyle<sup>2</sup>, anotados con las partículas negativas, su alcance y el evento o propiedad explícitamente negada. El corpus se encuentra en formato xml TIGER/SALSA (Erk y Padó, 2004)<sup>3</sup> y, además de la negación, también está anotado con correferencia, roles semánticos y argumentos implícitos. El corpus contiene 4.423 oraciones

<sup>1</sup>[http://www.coli.uni-saarland.de/projects/semEval2010\\_FG/](http://www.coli.uni-saarland.de/projects/semEval2010_FG/)

<sup>2</sup> Las obras son: *The Hound of the Baskervilles* y *The adventure of Wisteria Lodge*.

<sup>3</sup> El corpus está disponible en: <http://www.clips.ua.ac.be/BiographTA/corpora.html>

de las cuales el 22,49% incluyen al menos una partícula negativa.

Blanco y Moldovan (2013) seleccionaron 3.993 negaciones verbales del corpus *PropBank* (Palmer et al., 2005) para establecer el alcance y el foco de estas negaciones con el objetivo de representar su semántica. Siguiendo a Huddleston y Pullum, (2002), definen el foco como la parte del alcance que está más destacada y explícitamente negada.

Los corpus *SFU Review*, *Product Review<sub>EN</sub>* y *ConanDoyle-neg* se basan o inspiran en la guía de anotación (Vincze, 2010) utilizada para anotar *BioScope*. Las diferencias residen principalmente en la manera de anotar el alcance, en concreto, qué elementos quedan dentro o fuera del mismo.

En este artículo utilizaremos el corpus *SFU Review<sub>SP</sub>*, y lo anotaremos con negación, siguiendo parcialmente el sistema ABSA utilizado en la tarea 12 de SemEval<sup>4</sup>. *SFU Review<sub>SP-NEG</sub>* tiene un total de 198.551 palabras. El corpus está constituido por 400 comentarios, de los cuales ya se ha anotado un 75%<sup>5</sup>, lo que corresponde a un total de 6.331 oraciones, de las cuales 2.953 contienen al menos una estructura negativa. De éstas, 1.430 contienen una sola estructura negativa y 620 contienen más de una. Está organizado en ocho bloques de 50 ficheros cada uno. De estos 50 ficheros, 25 corresponden a opiniones positivas y 25 a opiniones negativas. Cada fichero contiene la opinión de un usuario acerca de un producto. Además, el corpus está anotado morfológicamente, con su categoría gramatical y lema correspondiente.

### 3 Definición y delimitación de la negación

La negación es un fenómeno lingüístico mediante el cual se invierte el valor de verdad de la unidad lingüística (proposición, sintagma o palabra) a la que se aplica. En las lenguas la negación se expresa mediante diversos mecanismos, siendo los más comunes el uso de partículas de negación sintácticamente independientes ('no', 'nunca', 'nadie', etc.), prefijos ('imposible', 'ilícito') y frases hechas ('en la vida'), entre otros.

En nuestra aproximación al tratamiento de la negación para la anotación del corpus en español nos hemos centrado, de momento, en la

<sup>4</sup> <http://alt.qcri.org/semeval2015/task12/>

<sup>5</sup> Faltan por anotar los comentarios de películas y ordenadores.

negación a nivel sintáctico, es decir, la que afecta a sintagmas y a la oración. Queda excluida de nuestra tipología la negación léxica ('dudar', 'ausencia de', 'falta de', etc.) y la morfológica, es decir, palabras con un afixo de negación ('descontento', 'incoherente').

Esta aproximación es acorde con la definición propuesta por la RAE (2009: 3631): "En sus múltiples manifestaciones gramaticales, la negación se considera un operador sintáctico en un sentido similar al de los cuantificadores y determinados adverbios, es decir, un elemento que condiciona (...) la referencia de otras unidades que se hallan en su ámbito de influencia". Las palabras que expresan negación pertenecen a diferentes categorías gramaticales: adverbios ('no', 'jamás', 'nunca', 'tampoco', 'nada'); pronombres ('nada', 'nadie', 'ninguno', 'nunca'); conjunciones ('ni', 'sino'); preposiciones ('sin', 'en vez de', etc.); determinantes indefinidos (ningún, ninguna, etc.). Como se puede observar, algunas palabras como 'nada' pueden pertenecer a más de una categoría.

### 4 Foco y alcance de la negación: <scope> y <event>

En los tratados gramaticales ((RAE, 2009) y (Bosque y Demonte, 1999))- se distingue entre el foco y el alcance de la negación. Según la gramática, el alcance de la negación corresponde a la totalidad de palabras afectadas por la misma, mientras que el foco corresponde a la palabra o sintagma dentro del alcance que se niega explícitamente.

- (1) No pienso ir al concierto ni contigo ni con nadie. (RAE, 2009: 3638)

En la oración (1), el alcance sería la oración entera y el foco 'ni contigo ni con nadie'. Lo que se niega no es el hecho de ir al concierto sino el hecho de ir acompañado (foco).

El modo en que estos dos conceptos se han plasmado en los diferentes corpus anotados es muy diverso. En lo que se refiere al alcance, la RAE (2009: 3655) considera que si el sujeto es postverbal, queda incluido en el alcance, mientras que si es preverbal, queda fuera. De los corpus descritos en la sección 2, solo en el corpus de *ConanDoyle-neg* el sujeto se incluye en el alcance.

Respecto de la partícula negativa, la RAE no se pronuncia sobre su inclusión o no inclusión

en el alcance. De los corpus mencionados, solo *Bioscope* la incluye en el alcance.

La mayoría de corpus no anotan el foco, por ser un componente de la negación de carácter semántico-pragmático, que muchas veces resulta difícil de identificar. La resolución del foco requiere las más de las veces disponer de información contextual que no siempre se encuentra disponible. Entre los corpus revisados, sólo Blanco y Moldovan (2011) lo tratan, ya que su objetivo es la representación semántica de la negación. Como contrapartida, en algunos corpus anotados con negación, por ejemplo en *ConanDoyle-neg*, se anota un componente de la misma, el evento, que no aparece en los tratados gramaticales, y con el que se pretende marcar el elemento directamente afectado por la negación, siempre dentro del alcance.

En nuestra propuesta, el alcance siempre corresponde a un constituyente sintáctico, es decir un sintagma o una oración (2) y el sujeto queda incluido cuando la negación afecta al predicado verbal. En el sistema de anotación se marca con la etiqueta <scope><sup>6</sup>.

- (2) a. [Sin mirar el aceite.]<sub>sn</sub>  
 b. [Cero fiabilidad.]<sub>sn</sub>  
 e. [No llegaron a tiempo.]<sub>o</sub>

En lo que respecta al foco, no lo hemos tratado en la versión actual del corpus, pero sí que hemos considerado interesante marcar la palabra directamente negada por el operador negativo, es decir, el evento o núcleo del constituyente que se niega (el nombre, el adjetivo, el verbo y el adverbio). Utilizamos la etiqueta <event> para anotar este elemento. En el caso de los sintagmas preposicionales introducidos con la partícula negativa ‘sin’, el evento es el sintagma nominal o la oración afectados por la preposición. En el caso de los verbos copulativos, el evento de la negación es el verbo más el atributo. En el caso de los verbos con complemento predicativo, este último se incluye también en el evento. En el caso de las perífrasis (‘no acaba de salir’), las colocaciones (‘no da problemas’) y los verbos ‘light’ con complemento (‘no se dio por vencido’, ‘no decir mucho [a cerca de/sobre/...]’) el evento incluye a toda la forma verbal compleja.

<sup>6</sup> En los ejemplos, utilizamos los corchetes para marcar el alcance y subrayamos el evento.

Son casos especiales de evento y alcance los pronombres indefinidos de negación cuando se usan antepuestos al verbo, es decir, cuando no van acompañados de la partícula ‘no’ (3). En ‘Nadie [=‘ninguna persona’] vino’, el alcance y el evento coinciden en la forma ‘nadie’ (3a), del mismo modo que en ‘Ningún niño vino’ el alcance es ‘ningún niño’ y el evento ‘niño’ (3b). En estos casos no se niega el verbo, sino que se le asigna un sujeto que tiene como referente el conjunto vacío (RAE: 3646).

- (3) a. [Nadie] vino.  
 b. [Ningún niño] vino.

## 5 Tipología

Hemos construido la tipología de expresiones de negación teniendo en cuenta, por un lado, los principios básicos contenidos en las gramáticas descriptivas y normativas ((Bosque y Demonte, 1999) y (RAE, 2009)) y, por otro, la coherencia, la sistemática y la máxima sencillez en la metodología y el conjunto de etiquetas (*tagset*) para la anotación del corpus. Suele ocurrir que en los corpus aparecen estructuras, construcciones o expresiones que no están contempladas en las gramáticas, por lo que se plantean problemas a la hora de expresar el contenido de las mismas en términos de un *tagset*. Es por ello que nuestra tipología, si bien está basada en la gramática, garantiza que es consistente desde el punto de vista de la anotación y que los tipos definidos (o categorías) constituyen clases claramente disjuntas, lo que facilita el proceso de anotación. Todas las expresiones de negación que hemos hallado en el corpus *SFU Review<sub>SP</sub>*, pertenecen a una clase de nuestra tipología, por lo que queda probada suficientemente su validez y consistencia teniendo en cuenta que el corpus tiene un tamaño suficiente para garantizar que incluye una amplia gama de estructuras de negación.

Para definir nuestros tipos de expresiones de negación hemos tenido en cuenta tanto la estructura sintáctica como su interpretación semántica, es decir, si la estructura negativa expresa o no una negación. La tipología se estructura en torno a dos grandes bloques, la expresión de la negación simple (5.1) y compleja (5.2), ambas con la etiqueta ‘neg’ asociada. En (5.3) se presentan las estructuras negativas que no expresan negación.

## 5.1 Negación simple

Se considera ‘negación simple’ la expresión de la negación mediante una única partícula. Esta partícula va antepuesta al evento y puede ser un adverbio (‘no’, ‘jamás’, ‘apenas’, ‘nunca’) (4a-b), un pronombre antepuesto al verbo (‘nadie’, ‘nada’) (4c), o una preposición (‘sin’) (4d).

- (4) a. (...) para conductores que **apenas**<sub>adv</sub> tocan el coche.  
 b. **Nunca**<sub>adv</sub> tienen las piezas de recambio en el taller.  
 c. **Nadie**<sub>pr</sub> quedará decepcionado en este aspecto.  
 d. **Sin**<sub>p</sub> conexión.

Incluimos también en esta categoría la coordinación de oraciones negativas simples (5).

- (5) a. [Ni puedo desear más] [ni puedo contentarme con menos].  
 b. El aire acondicionado [ni enfría] [ni calienta].

## 5.2 Negación compleja

Dentro del tipo ‘negación compleja’ incluimos la expresión de la negación mediante dos o más partículas, continuas (6) o discontinuas (7)<sup>7</sup>, la primera de las cuales suele expresar negación, mientras que la segunda puede expresar también negación (7) reforzando así la primera (véase sección 5.2.1), o puede modular el valor de la negación (6) (véase sección 5.2.2).

- (6) **Casi no** llega.  
 (7) **No** vino **nunca**.

En nuestro sistema de anotación, las partículas de la negación compleja tienen asociada la etiqueta <discid=’1n/1c, 2n/2c,...’> (discontinua).

A continuación, describimos más detalladamente estas dos clases de negación.

### 5.2.1 Refuerzo de la negación

En español es frecuente que las expresiones de negación se refuercen mediante una segunda

<sup>7</sup> En el 75% del corpus que se ha anotado, se han identificado un total de 2.375 expresiones negativas simples y complejas continuas, -de las cuales 229 no expresan negación- y 449 complejas discontinuas.

partícula (8a-11a). Es lo que en nuestro sistema de anotación denominamos refuerzo de la negación. Estas expresiones siempre se pueden parafrasear anteponiendo al verbo la segunda partícula negativa, dando como resultado una negación simple (8b-11b):

- (8) a. Ustedes **no** pueden hacer **nada**.  
 b. Ustedes **nada** pueden hacer.  
 (9) a. En los Nokia que he utilizado **no** he tenido **nunca** este problema.  
 b. **Nunca** he tenido este problema en los Nokia que he utilizado.  
 (10) a. Allí **no** me esperaba **nadie**.  
 b. **Nadie** me esperaba allí.  
 (11) a. Puede que **ni** siquiera los hayan escuchado **jamás**.  
 b. Puede que **jamás** los hayan escuchado.

Cuando se da la coordinación de dos estructuras negativas en un mismo sintagma también lo consideramos dentro de esta categoría (12), ya que la repetición de partículas negativas (‘ni... ni...’) también da idea de refuerzo.

- (12) a. No comió **ni** pan **ni** vino.  
 b. No me sentí **ni** libre **ni** poderoso.  
 c. **Sin** agua **ni** comida.

### 5.2.2 Negación con modificadores

La negación, al igual que muchos otros fenómenos lingüísticos, no es categorial, sino que puede presentar gradación. Existen diferentes mecanismos para expresar esta gradación, que en nuestro sistema de anotación denominamos modificadores y que pueden ser incrementadores, cuando potencian la negación (13) y decrementadores, cuando la atenúan (14).

- (13) a. Mi coche no frena **en absoluto**.  
 b. No te molesta **nada**<sup>8</sup>.  
 (14) a. No estoy **muuy** segura.  
 b. No tiene **mucho** sentido.  
 c. No da **demasiadas** opciones de idioma.

En nuestro sistema de anotación anotamos los incrementadores con la etiqueta

<sup>8</sup> Nótese que ‘nada’ se ha interpretado como un adverbio, en el sentido de ‘en absoluto’, pero podría ser también un pronombre. Solo el contexto ha permitido desambiguarlo.

<increment> y los decrementadores con la etiqueta <reduction>. En estos casos, la partícula negativa aparece en primer lugar, y solo en casos de dislocación, la partícula negativa va en segundo lugar, precedida por el modificador (15).

(15) **Más** equivocado **no** pude estar.

### 5.2.3 Comparativas con negación

Un tipo particular de negación es el que se da en una estructura que expresa comparación, anotadas con la etiqueta 'comp'. Son siempre estructuras discontinuas (16).

- (16) a. **No** me gusta **tanto como** lo otro.  
b. Mi amor **no** iba a ser **más** pequeño **que** yo.  
c. El ambiente de este local es agradable pero **no** (verbo elidido) **tanto como** el del otro.  
d. El motor **no** es **todo lo** potente **que** debería.

### 5.2.4 Frases hechas que expresan negación

Existen construcciones complejas lexicalizadas que expresan negación (17). En nuestro sistema de anotación las consideramos como una sola unidad, de manera que formarían parte del léxico de partículas de negación.

- (17) a. En la vida.  
b. En toda mi vida.  
c. Ni lo sueñes.

Cabe destacar que los casos que presentan variables (17a y 17b) se tratan como unidades diferentes. Estas expresiones complejas pueden incluir (17c) o no (17a y 17b) una partícula negativa.

### 5.3 Estructuras negativas que no expresan negación

Existen expresiones que aunque contienen partículas de negación, semánticamente o bien no expresan negación o bien expresan un contraste o contraposición entre dos o más opciones o posibilidades. Dentro de este tipo distinguimos las estructuras simples de las complejas.

#### 5.3.1. Estructuras simples que no expresan negación

Dentro de esta clase se incluyen las partículas negativas en oraciones interrogativas (18), las partículas negativas en contextos que no expresan negación (19) y las partículas negativas con valor expletivo (20).

- (18) El coche lo compré para viajar, **no**?  
(19) **Nada\_más** darle al contacto<sup>9</sup>.  
(20) No pienso irme hasta que **no** vengas.

Las frases hechas o expresiones lexicalizadas con partícula negativa que no expresan negación (21) como se tratan como una única expresión léxica multipalabra, las incluimos en esta clase.

- (21) a. Visto y no visto.  
b. Sin pena ni gloria.  
c. No hace más que.  
d. No hay más que.

Todas estas estructuras se anotan con la etiqueta 'noneg'.

#### 5.3.2. Estructuras complejas que no expresan negación

Dentro de esta categoría se incluyen las estructuras de contraste entre dos o más elementos que se contraponen bien para introducir una corrección (22a) o para añadir información nueva (22b). En otros casos se expresa una contraposición respecto de un límite o cota que se explicita (23).

- (22) a. **No** vinieron 2 soldados, **sino** 6.  
b. **No\_sólo** lleva rueda de recambio **sino\_también** caja de herramientas.  
(23) a. BMW **no** suele poner **más\_que** lo que considera necesario.  
b. **No** veo otra salida **que** pedirle otra lavadora.

La oración de (23a) se parafrasea como 'BMW suele poner sólo lo que considera necesario', de manera que se especifica el límite en las inversiones de BMW. La oración de (23b) se parafrasea como 'La única salida es pedirle otra lavadora', por lo tanto lo que se

<sup>9</sup> Nótese que 'nada\_más' lo tratamos como un único elemento léxico.

expresa es la única opción posible, el límite. En ningún caso se expresa una negación.

Todas estas estructuras se anotan con la etiqueta ‘contrast’.

## 6 Esquema de anotación

En esta sección se describen brevemente los atributos utilizados en la anotación de la negación del corpus *SFU Review<sub>SP</sub>-NEG* recogidos en el esquema general de anotación de la Figura 1.

La etiqueta `<review_polarity>` indica la polaridad de todo el comentario, que puede ser positiva o negativa. En *SFU Review<sub>SP</sub>-NEG* solo se anotan las oraciones (`<sentence>`) que contengan al menos una negación. Cuando la oración contiene más de una estructura negativa (`<neg_structure>`) se asigna el valor ‘yes’ al atributo `<sentence_complex>` y cuando solo incluye una única estructura negativa el valor ‘no’.

```

<review_polarity= ‘positive/negative’
<sentence_complex=‘yes/no’>
<neg_structure
polarity=‘positive/negative/neutral’
change=‘yes/no’

polarity_modifier=‘increment/reduction’
value=‘neg/contrast/comp/noneg’
<scope>
<negexp discid=‘1n/1c’>
</negexp>
<event>
</event>
</scope>
</neg_structure>
</sentence>

```

Figura 1: Esquema general de anotación.

La etiqueta `<neg_structure>` tiene asociados cuatro atributos:

- `<polarity>`: indica la orientación positiva, negativa o neutra de la estructura negativa (p.e.: ‘no es un chico malo’, ‘no es un chico bueno’, ‘no es un chico alto’).
- `<change>`: indica si, debido a la negación, la estructura negativa ha visto modificada o no totalmente su polaridad (p.e.: ‘chico bueno’ vs. ‘chico no bueno’) o su significado (‘chico alto’ vs. ‘chico no alto’).
- `<polarity_modifier>`: indica si en la estructura negativa hay algún elemento que

modifica o matiza su polaridad (p.e.: ‘chico bueno’ vs. ‘chico no muy bueno’). Este atributo tiene dos valores posibles: ‘increment’ para indicar que se incrementa la polaridad (p.e.: ‘no me arrepiento para nada’) y ‘reduction’ para cuando se reduce (p.e.: ‘no lo he utilizado mucho’).

- `<value>`: indica el significado expresado por la estructura negativa. Tiene cuatro valores posibles: ‘neg’ cuando indica negación; ‘contrast’ cuando expresa contraste u oposición ente términos; ‘comp’ cuando expresa comparación o desigualdad entre términos; y ‘noneg’ para indicar las estructuras que contienen una partícula negativa pero que no niegan.

La etiqueta `<scope>` se usa para anotar el alcance de la negación, incluyendo la propia partícula negativa y `<negexp>` para delimitar la palabra o palabras que expresan negación. `<negexp>` puede llevar asociado el atributo `<discid>`, que se aplica en aquellas estructuras negativas donde hay más de un elemento y los casos de estructuras negativas discontinuas. La etiqueta `<event>` sirve para marcar la palabra o palabras directamente negadas por el operador negativo.

## 7 Conclusiones y líneas futuras

En este artículo hemos presentado los diferentes tipos de negación en español y el sistema de etiquetas utilizado para la anotación del corpus *SFU Review<sub>SP</sub>-NEG*, el primer corpus del español anotado con esta información. Aunque se ha anotado sólo un 75% del corpus, el número de casos observados y anotados (2.050) permite suponer que nuestra tipología es completa y abarca el fenómeno en su totalidad. El corpus es de libre disposición<sup>10</sup>.

Tenemos previsto como líneas futuras, por un lado, el tratamiento del foco y de la negación léxica y morfológica y, por otro, terminar la anotación del corpus.

## Bibliografía

- Blanco E. y D. Moldovan. 2013. Retrieving implicit positive meaning from negated statements. *Natural Language Engineering*, 20 (4): 501-535. Cambridge University Press.

<sup>10</sup> <http://sinai.ujaen.es/sfu-review-sp-neg/>

- Bosque I. y V. Demonte. 1999. Gramática Descriptiva de la Lengua Española, Vol. 2. Espasa Calpe, España.
- Councill, I. G., R. McDonald, y L. Velikovich, L. 2010. What's great and what's not: learning to classify the scope of negation for improved sentiment analysis. *Proceedings of the workshop on negation and speculation in natural language processing*, páginas 51-59, Uppsala, ACL.
- Erk K., y S. Padó. 2004. A powerful and versatile XML format for representing role-semantic annotation. *Proceedings of 4th International Conference on Language Resources and Evaluation (LREC'04)*, Lisboa, Portugal.
- Huddleston, R.D. y G. K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge University Press, Cambridge, UK.
- Kim J.D., T. Ohta y J. Tsujii. 2008. Corpus annotation for mining biomedical events from literature. *BMC Bioinformatics*, 9:10.
- Konstantinova, N., S. C de Sousa, N. P. Díaz, N. P. Cruz, M. J. Maña, M. Taboada y R. Mitkov. 2012. A review corpus annotated for negation, speculation and their scope. *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC'12)*, páginas 3190-3195, Turkey.
- Konstantinova, N. y S. C de Sousa. 2011. Annotating Negation and Speculation: the Case of the Review Domain. *Proceedings of the Student Research Workshop associated with RANLP 2011*, páginas 139-144, Bulgaria.
- Morante, R. y W. Daelemans. 2012. ConanDoyle-neg: Annotation of negation in Conan Doyle stories. *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC'12)*, páginas 1563-1568, Turkey.
- Morante, R. y C. Sporleder. 2012. Modality and negation: An introduction to the special issue. *Computational linguistics*, 38(2), 223-260.
- Palmer, M., P. Kingsbury y D. Gildea. 2005. The Proposition Bank: An Annotated Corpus of Semantic Roles, *Computational Linguistics*, 21 (1).
- Pang, B., L. Lee, y S. Vaithyanathan, S. 2002. Thumbs up?: sentiment classification using machine learning techniques. *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*: 79-86. ACL.
- Polanyi L., Zaenen, A. 2006. Contextual Valence Shifters. *Computing affect and attitude in text: Theory and applications*, 20: 1-10. *The Information Retrieval Series*.
- Pyysala S., F. Ginter, J. Heimonen, J. Björne, J. Boberg, J. Järvinen y T. Salakosk. 2007. BioInfer: a corpus for information extraction in the biomedical domain. *BMC Bioinformatics*, 8: 50.
- RAE. 2009. Nueva Gramática de la Lengua Española. Vol. 2. Espasa Libros, España.
- Ruppenhofer J., C. Sporleder, R. Morante, C. Baker y M. Palmer. 2010. Semeval-2010 task 10: Linking events and their participants in discourse. *Proceedings of the 5th Workshop on Semantic Evaluations (ACL 2010)*, páginas 45-50, Suecia.
- Taboada, M., C. Anthony y K. Voll. 2006. Methods for creating semantic orientation dictionaries. *Proceedings of the 5th Conference on Language Resources and Evaluation (LREC'06)*, páginas 427-432.
- Vincze, V., Szarvas G., Farkas R., Móra G. y Csirik J. 2008. The BioScope corpus: biomedical texts annotated for uncertainty, negation and their scopes. *BMC Bioinformatics*, 9:1-9.
- Vincze, V. 2010. Speculation and negation annotation in natural language texts: what the case of bioscope might (not) reveal. *Proceedings of the workshop on negation and speculation in natural language processing*, páginas 51-59, Uppsala, ACL.
- Wiegand, M., A. Balahur, B. Roth, D. Klakow, y A. Montoyo. 2010. A survey on the role of negation in sentiment analysis. In *Proceedings of the workshop on negation and speculation in natural language processing*, páginas 60-68, ACL.