

## Utilización de corpórea textuales para la extracción de modificadores contextuales de valencia para tareas de Análisis de Sentimiento

Antonio Moreno Ortiz, Chantal Pérez Hernández, Rodrigo Hidalgo García

Facultad de Filosofía y Letras

Universidad de Málaga

### ABSTRACT

*El desarrollo de herramientas de Análisis de Sentimiento requiere de la codificación del lenguaje evaluativo en términos de polaridad. Una de las principales dificultades es la existencia de variables que modifican la interpretación positiva o negativa de lo expresado. Si bien algunos fenómenos discursivos, como la ironía, son muy difíciles de definir en términos formales, otros modificadores contextuales de la valencia sí muestran rasgos identificables. Un claro ejemplo es la negación, que invierte la polaridad de la unidad léxica a la que modifica. Sin embargo, hay formas de inversión de la valencia bastante más sutiles, cuya detección no es tan fácil y requieren igualmente ser codificadas en forma de reglas de contexto en las herramientas de análisis automatizado. En este trabajo describimos nuestra experiencia en el empleo de corpórea en la consecución de este objetivo dentro del proyecto Sentitext: una herramienta de Análisis de Sentimiento para el español.*

Keywords: análisis del sentimiento, minería de opinión, modificadores contextuales de valencia, expresiones multi-palabra, colocaciones.

### ABSTRACT

*The development of Sentiment Analysis tools requires the codification of evaluative language in terms of its polarity. One of the main difficulties is the existence of variables that modify the polarity, either positive or negative, of what is being said. Although some discourse phenomena, such as irony, are difficult to account for in formal terms, there are many other Contextual Valence Shifters (CVS) that do show certain formally identifiable traits. Negation is probably the most straightforward example of valence inversion, but there are certainly more subtle means in which valence inversion can be conveyed, which also need to be accounted for and defined in terms of Contextual Valence Shifters. In this paper, we describe our experience in the use of corpora in order to attain this objective for Sentitext: a Sentiment Analysis tool for Spanish.*

Keywords: sentiment analysis, opinion mining, contextual valence shifters, multiword expressions, collocations

### ***1. ANÁLISIS DE SENTIMIENTO***

El análisis de sentimiento es una vertiente de la minería de textos que ha recibido gran atención en los últimos años debido en gran parte a la aparición de la Web 2.0 y la explosión que ésta ha supuesto en cuanto a grandes cantidades de texto evaluativo generado por los usuarios de la Red en torno todo tipo de productos y servicios, cuyo análisis y escrutinio es de obvio interés para empresas y organizaciones de diversa naturaleza. En un sentido amplio, consiste en el empleo de herramientas computacionales para la detección y análisis del lenguaje valorativo, en términos de polaridad.

Pang y Lee (2008) repasan el avance de la investigación en el campo del análisis de sentimiento en la última década, donde el principal foco de atención ha sido el análisis de valoraciones de usuarios de películas (Pang, Lee & Vaithyanathan, 2002), productos electrónicos de consumo (Dave et al., 2003), pero también el análisis del mercado financiero (Das y Chen, 2001), de discursos políticos (Thomas et al., 2006) o las valoraciones de usuarios en webs de evaluación de productos de carácter general, como Epinions (Turney, 2002; Taboada y Grieve, 2004). En español la actividad en el campo es mucho más reducida, aunque hay notables trabajos, como los de Cruz et al. (2008) y Boldrini et al. (2009). Moreno Ortiz et al. (2010a) analizan las valoraciones que los usuarios de la web *tripadvisor.es* hacen de los hoteles que visitan a través de una nueva herramienta llamada Sentitext (Moreno Ortiz et al., 2010b), la cual utilizamos también para el estudio que llevamos a cabo en este artículo.

La visión más extendida en el ámbito del análisis de sentimiento es que éste es altamente dependiente del dominio (Aue & Gamon 2005). Es decir, determinadas palabras y unidades fraseológicas que en un dominio determinado serían contempladas como afectivamente neutras pueden adquirir una determinada *polaridad*, u *orientación semántica*, en ese dominio determinado. Por tanto, los esfuerzos han estado enfocados a construir recursos léxicos específicos de dominios determinados, empleando para ello técnicas de aprendizaje automático, con o sin supervisión (Turney & Littman, 2002). Por el contrario, la herramienta que hemos desarrollado y que usamos en este artículo, Sentitext, es independiente del dominio y no usa ningún algoritmo de aprendizaje como tal sino que cuenta con bases de datos léxicas de extensa cobertura cuyo contenido se ha adquirido de forma manual.

Sentitext se nutre de tres fuentes de datos fundamentales: (1) el léxico de palabras individuales, (2) el léxico de frases y (3) las reglas de contexto. Las tres han sido desarrolladas de forma manual, pero utilizando herramientas creadas al efecto para facilitar el proceso de adquisición. El lexicón de palabras individuales consta actualmente de unas 10.000 palabras con carga afectiva, marcadas con su polaridad con uno de estos 4 valores: -2 (muy negativa), -1 (negativa), 1 (positiva), 2 (muy positiva).

Una parte de especial importancia en nuestro trabajo y que determina en gran parte la efectividad del sistema de análisis de sentimiento ha sido la inclusión de locuciones o expresiones multi-palabra. Como señalan Wilson et al. (2009), a veces se usan palabras positivas en frases negativas (p. ej.: “abuso de confianza”) y viceversa. En nuestro caso, decidimos desde un principio adquirir un lexicón de locuciones de amplia cobertura y no

específico del dominio. Dicho lexicón contiene en la actualidad más de 17.000 entradas, aunque, a diferencia del lexicón de palabras individuales, no todas tienen carga afectiva.

El tercer componente de nuestro sistema, las *reglas de contexto*, está inspirado en gran medida en la propuesta de Polanyi y Zaenen (2006), en lo que ellas denominan *Context Valence Shifters (CVS)*, o modificadores contextuales de la valencia. Estas reglas sirven para formalizar la modificación de valencia que trae consigo una estructura determinada y codifican además el resultado de la aplicación de cada regla. De entre las 350 reglas de contexto que hemos compilado hasta la fecha, algunas son muy simples y generales (por ejemplo, la presencia del adverbio “muy” precediendo a un adjetivo con carga afectiva intensifica su valencia en el mismo sentido), aunque otras son mucho más complejas y específicas.

## 2. SELECCIÓN DE PALABRAS CLAVE O SEED SETS

Diversos autores han ensayado métodos diferentes de asignación automática de la polaridad a las palabras, haciendo uso de diferentes algoritmos de aprendizaje.

Hatzivassiloglou & McKeown (1997) usaron un sistema automatizado para asignar polaridad a las palabras basado en patrones de co-aparición estadísticamente frecuente. Wiebe (2000) emplea corpórea etiquetado con anotaciones manuales para obtener, al igual que en nuestro caso, no sólo polaridad sino también gradación. Otro enfoque ha consistido en la utilización de un conjunto inicial de palabras (*seed set*) a partir del cual se genera un léxico de gran cobertura mediante su utilización junto a un diccionario de sinónimos (Pitel & Grefenstette, 2008).

Nuestra aproximación particular estuvo motivada precisamente por esta última metodología. En primer lugar, adaptamos al español el *seed set* de Pitel & Grefenstette (2008), originalmente en francés, consistente en 44 pares de palabras antónimas con carga afectiva y, en segundo lugar, integramos este conjunto inicial en una aplicación junto a un diccionario de sinónimos, en concreto el Open Office Thesaurus, trabajando de forma cíclica con un proceso de adquisición semi-automático y supervisado.

La base de datos de palabras de Sentitext no sólo asigna polaridad a la palabra o expresión multi-palabra en cuestión sino que también distingue gradación (Hatzivassiloglou & McKeown 1997) en el rango -2 a +2. En una etapa posterior, las palabras de la base de datos fueron clasificadas y asignadas a una emoción básica, siguiendo la clasificación propuesta por Parrott (2001), de forma que es posible encontrar todas las palabras positivas/negativas pertenecientes a la emoción *rabia, dolor, orgullo, deseo*, etc.

## 3. ADQUISICIÓN DE REGLAS DE CONTEXTO

El proceso de adquisición de las reglas de contexto fue la que desde el principio ha estado guiada por el análisis exhaustivo de corpórea. Para el análisis que aquí presentamos hemos seleccionado una serie de palabras clave, en su mayoría sustantivos, de nuestro *seed set* original. Estos sustantivos se han empleado como términos de búsqueda en la

identificación de patrones de co-aparición con un verbo o sustantivo deverbal, con el objetivo de analizar, en el corpus, cómo ciertos verbos pueden invertir, intensificar o modificar la polaridad de estos sustantivos. Todos ellos estaban incluidos en la base de datos de Sentitext con emoción etiquetada según Parrott (2001), tenían una valencia alta [+2 o -2] y una alta frecuencia en nuestro corpus de español:

**Tabla 1. Sustantivos analizados en corpus con indicación de valencia y clasificación según Parrott (2001)**

Tipo de Emoción	Sustantivos y valencias asignadas
rage	violencia [-2], ataque [-2]
neglect	ignorancia [-2], deterioro [-2]
nervousness	amenaza [-2], preocupación [-2]
pride	dignidad [2], éxito [2]
irritation	enfado [-2], queja [-2], provocación [-1]
cheerfulness	alegría [2], felicidad [2]
lust	ambición [-1]
suffering	dolor [-2], malestar [-1]
sadness	tristeza [-2], desolación [-2]
affection	solidaridad [2], empatía [1]
contentment	tranquilidad [2], satisfacción [1]
exasperation	aburrimiento [-1], cansancio [-1]
horror	pánico [-2], devastación [-2]
disgust	corrupción [-2], desprecio [-2]
shame	culpa [-2], equivocación [-1]
relief	alivio [2], seguridad [2]
optimism	confianza [1], esperanza [1]
envy	envidia [-2], celos [-2]
disappointed	decepción [-2], fracaso [-2]
sympathy	ayuda [1], compasión [1]
longing	ilusión [2], ensueño [1]

Aunque existen multitud de modificadores contextuales de la valencia dentro del contexto lingüístico más inmediato a la palabra que determinan cambios en la polaridad, como por ejemplo la introducción del operador “no” (*esto es bueno à esto no es bueno*), hay muchos otros que no son tan evidentes, ya que no todas las formas de inversión de la valencia son tan fácilmente identificables como una simple negación. Esto hace necesario el trabajo de detección y catalogación de modificadores contextuales de la valencia que hemos llevado a cabo, ya sea con las combinaciones de verbo + sustantivo (por ej., “carecer de dignidad”, “vulnerar las leyes”, “hacer frente a la crisis” o “superar el problema”), o las expresiones multipalabra y colocaciones en las que alguno de sus componentes tiene carga afectiva,

tales como “ser un rayo de luz para el enfermo”, “hacer un flaco favor al progreso”, etc. Si prestamos atención a la diversidad de construcciones gramaticales y léxicas implicadas, incluso en estos pocos ejemplos, parece obvio que la tarea de identificar, clasificar y definir estos modificadores contextuales de valencia, no es trivial.

#### 4. RESULTADOS DEL ANÁLISIS

Hemos estudiado las búsquedas de nuestras palabras clave (Tabla 1), ordenándolas alfabéticamente a partir de la segunda palabra precedente a la palabra clave (*Sort 2L*) para ayudar en nuestra labor de localización de patrones de co-aparición de verbos con sustantivos con más facilidad. A continuación, hemos analizado todos los casos en los que un verbo precede al sustantivo y modifica su valencia, clasificando estas combinaciones de palabras en alguna de las siguientes categorías:

**Tabla 2. Patrones de co-aparición (verbo + sustantivo) modificadores de valencia del sustantivo.**

<b><u>Modificación de valencia en sustantivos positivos</u></b>
CO-APARICIÓN CON VERBO O SUSTANTIVO DEVERBAL. INVERSIÓN [+ a -]
CO-APARICIÓN CON VERBO O SUSTANTIVO DEVERBAL. NEUTRALIZACIÓN [+ a 0]
CO-APARICIÓN CON VERBO O SUSTANTIVO DEVERBAL. INTENSIFICACIÓN [+ a ++]
<b><u>Modificación de valencia en sustantivos negativos</u></b>
CO-APARICIÓN CON VERBO O SUSTANTIVO DEVERBAL. INVERSIÓN [- a +]
CO-APARICIÓN CON VERBO O SUSTANTIVO DEVERBAL. NEUTRALIZACIÓN [- a 0]
CO-APARICIÓN CON VERBO O SUSTANTIVO DEVERBAL. INTENSIFICACIÓN [- a --]

La primera dificultad con la que nos encontramos es que hay palabras clave cuya clasificación como palabras positivas o negativas es incierta y altamente dependiente del contexto. Consideremos los siguientes ejemplos:

- a) *El nuevo presidente tiene una enorme ambición.*
- b) *una asquerosa ambición por el dinero.*

En el ejemplo a), la palabra *ambición* está ligada a la emoción *optimism* mientras que en b) tiene un significado más próximo a la codicia (*lust*). El hecho de que Sentitext clasifique esta palabra como negativa se rige por la mayor frecuencia de casos negativos, pero habrá que observar con detenimiento cómo el contexto modifica su valencia. Tras analizar la totalidad de las palabras clave seleccionadas, hemos obtenido extensas listas de combinaciones de *verbo + sustantivo* que modifican la valencia del sustantivo.

Resumimos a continuación las categorías más destacadas.

#### 4.1 Verbos modificadores de la valencia exclusivos de sustantivos positivos/negativos

Una de las primeras conclusiones que podemos extraer es que, si bien algunos modificadores contextuales de la valencia son comunes a los sustantivos positivos y negativos, existe un numeroso grupo de verbos que modifican la valencia de sustantivos exclusivamente positivos, así como otro grupo modificador de sustantivos exclusivamente negativos. En general, el verbo que modifica la valencia de un sustantivo positivo indica *oposición*, de forma que invierte la polaridad del sustantivo a negativo, por ejemplo, *romper el acuerdo*. El verbo *romper* pertenece a la categoría de verbos que invierten de positivo a negativo, pero nunca al revés. De hecho, los verbos que invierten de negativo a positivo suelen ser verbos de *superación*, por ej. *resolver el conflicto*.

Otro de los verbos más comunes de inversión de positivo a negativo (INV [+ a -]) es *perder*. Este verbo precede en el corpus a sustantivos positivos, como por ejemplo: *puestos de empleo o confianza*. Sin embargo, si revisamos la lista de modificadores INV [- a +], no encontraremos *perder*, pero sí *vencer (violencia, pánico)*. Es decir, existen parejas de verbos modificadores de la valencia que apuntan en sentido opuesto, unos hacia *oposición* y otros hacia *superación*. A continuación se muestran algunos ejemplos de verbos comunes de *superación* (Tabla 3) y *oposición* (Tabla 4) que invierten la valencia del sustantivo con el que aparecen:

**Tabla 3. Verbos modificadores de valencia (INV [- a + ]).**

Modificación de la valencia del sustantivo por verbos específicos de inversión [- a +] (Indican superación, esfuerzo por cambiar o invertir la situación):	
VERBO	SUSTANTIVO
acabar con	crisis, problema, conflicto, violencia, dolor, corrupción
afrontar	crisis, problema, conflicto, violencia, ataque
combatir	inflación, delincuencia, crisis, enfermedad, violencia, amenaza, ataque, ignorancia, dolor, aburrimiento, corrupción
evitar	inflación, delincuencia, crisis, problema, violencia, amenaza, deterioro, provocación, dolor, cansancio, pánico, devastación, corrupción, fracaso
superar	crisis, problema, malestar, dolor, tristeza, pánico, corrupción, decepción
resolver	crisis, problema, conflicto, paradoja, violencia, amenaza, queja

**Tabla 4. Verbos modificadores de valencia (INV [+ a -]).**

Modificación de la valencia del sustantivo por verbos específicos de inversión [+ a -] (Indican oposición, trabar o dificultar):	
VERBO	SUSTANTIVO
dañar	alianza, paz social, dignidad, confianza
destruir	puestos de empleo, dignidad, felicidad, confianza
romper	acuerdo, tregua, paz, dignidad, éxito, solidaridad, tranquilidad, seguridad, confianza
perder	Protagonismo, paz, dignidad, alegría, seguridad, confianza, esperanza, ilusión

#### 4.2 Verbos modificadores de la valencia comunes a sustantivos positivos/negativos

La categoría semántica de los verbos que configuran este grupo puede considerarse diferente de los grupos semánticos de *oposición/superación* mencionados en 3.1. Se trata, por un lado, de verbos que expresan *eliminación* y por otro lado *disminución*. En ambos casos nos encontramos con estos modificadores de la valencia comunes a nuestro *seed set* de palabras positivas y negativas (véase sección 2). El motivo es que *la eliminación, la desaparición o el fin* de algo bueno, automáticamente convierte este momento en negativo y viceversa. De esta misma forma, *la reducción o disminución* de algo malo nos da como resultado una inversión de la valencia a positivo. Estos son sólo algunos ejemplos de entre muchos otros.

**Tabla 5. Verbos de inversión bidireccional con sentido de eliminar.**

Modificación de la valencia del sustantivo por verbos de inversión bidireccional [+ a -] y [- a +] (Indican eliminación):	
VERBO / SUSTANTIVO DEVERBAL	SUSTANTIVO
desaparecer / desaparición de	puestos de empleo, solidaridad, confianza, esperanza [+ a -] barreras, violencia, amenaza, dolor [- a +]
eliminar / eliminación de	puestos de empleo, ilusión [+ a -] dolor, cansancio, pánico, corrupción [- a +]
finalizar	acuerdo, tregua, paz [+ a -] violencia, dolor [- a +]

Modificación de la valencia del sustantivo por verbos de inversión bidireccional [+ a -] y [- a +] (Indican disminución):	
VERBO	SUSTANTIVO
disminuir	salario, ayuda, seguridad, confianza [+ a -] violencia, dolor, aburrimiento, fracaso [- a +]
reducir	puestos de empleo, salario, éxito, alegría, esperanza [+ a -] dolor, cansancio, riesgo, ineficiencia, amenaza, corrupción [- a +]
finalizar	acuerdo, tregua, paz [+ a -] violencia, dolor [- a +]

**Tabla 6. Verbos de inversión bidireccional con sentido de disminución.**

Por último, deberíamos detenernos en una categoría de verbos que tienen diferente efecto al anteponerse a sustantivos positivos o negativos. En concreto, se trata de *verbos aumentativos*. Consideremos los siguientes ejemplos:

- 1) exceso de (éxito, alegría, felicidad, ambición, solidaridad, tranquilidad, seguridad, confianza)
- 2) exceso de (dolor, corrupción, celos)

Mientras el exceso de algo bueno, como en 1, puede llegar a poner la situación en contra (INV [+ a -]), el exceso de algo malo, si bien menos frecuente como observamos en el ejemplo 2, va a producir un efecto intensificador, en este caso INT [- a -]. Los límites de este artículo no nos permiten describir la intensificación que determinados verbos producen en el componente valorativo de los sustantivos con los que co-aparecen, sin embargo, en general podemos entender que, irónicamente, el exceso de algo bueno nunca se entiende como algo positivo: “exceso de confianza” o “cegarse de felicidad”.

## 5. CONCLUSIONES

Hemos de tener en cuenta que este trabajo se centra fundamentalmente en un tipo particular de modificación de la valencia, concretamente la inversión de [+ a -] y [- a +], y aun así, hemos podido producir y clasificar multitud de patrones de co-aparición verbo-sustantivo del que sólo hemos podido mostrar unos pocos. Además, nos hemos centrado únicamente en la modificación de valencia por medio de un verbo o sustantivo deverbal, dejando de lado una multitud de posibilidades de modificación postpositiva, no sólo del sustantivo sino aplicable también a otras categorías de palabras. Parece evidente, por tanto, que la integración de la reglas de contexto a Sentitext suponen una ingente cantidad de trabajo adicional, si bien es cierto que ello implicaría, a la vista de nuestros resultados,



una mejora muy sustancial del rendimiento de la herramienta. Por otro lado, Sentitext pasaría a convertirse en la primera herramienta de análisis de sentimiento del español que hace uso de los CVS como reglas de contexto, un trabajo que podría hacerse extensible en el futuro a otras lenguas.

#### REFERENCIAS

- AUE, A., & GAMON, M. (2005). Customizing Sentiment Classifiers to New Domains: A Case Study. Presented at the Recent Advances in Natural Language Processing (RANLP), Borovets, Bulgaria.
- CRUZ, F., TROYANO, J. A., ENRIQUEZ, F., & ORTEGA, J. (2008). Clasificación de documentos basada en la opinión: experimentos con un corpus de críticas de cine en español. *Procesamiento del Lenguaje Natural*, (41), 73-80.
- DAS, S. R., & CHEN, M. (2001). Yahoo! for Amazon: Sentiment Parsing from Small Talk on the Web. *Proceedings of the 8th Asia Pacific Finance Association Annual Conference*.
- DAVE, K., LAWRENCE, S., & PENNOCK, D. M. (2003). Mining the peanut gallery: opinion extraction and semantic classification of product reviews. *Proceedings of the 12th international conference on World Wide Web* (págs 519-528). Budapest, Hungary: ACM.
- HATZIVASSILOGLOU, V., & McKEOWN, K. R. (1997). Predicting the semantic orientation of adjectives. *Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics* (págs 174-181). Madrid, Spain: Association for Computational Linguistics.
- BOLDRINI, E., & BALAHUR, A., PATRICIO MARTÍNEZ-BARCO & MONTOYO, A. (2009). EmotiBlog: an Annotation Scheme for Emotion Detection and Analysis in Non-traditional Textual Genres. *Proceedings of The 2009 International Conference on Data Mining* (págs 491-497). Presented at the DMIN 2009, Las Vegas, USA: CSREA Press.
- MORENO ORTIZ, A., PÉREZ POZO, Á., & TORRES SÁNCHEZ, S. (2010a). Sentitext: sistema de análisis de sentimiento para el español. *Procesamiento de Lenguaje Natural*, 45, 297-298.
- MORENO ORTIZ, A., PINEDA CASTILLO, F., & HIDALGO GARCÍA, R. (2010b). Análisis de Valoraciones de Usuario de Hoteles con Sentitext: un sistema de análisis de sentimiento independiente del dominio. *Procesamiento de Lenguaje Natural*, 45, 31-39.
- PANG, B., LEE, L., & 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* (págs 79-86). Association for Computational Linguistics.
- PARROTT, W. (2001). *Emotions in Social Psychology*. Philadelphia: Psychology Press.

- PITEL, G., & GREFFENSTETTE, G. (2008). Semi-automatic Building Method for a Multidimensional Affect Dictionary for a New Language. *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2008*.
- POLANYI, L., & ZAENEN, A. (2006). Contextual Valence Shifters. *Computing Attitude and Affect in Text: Theory and Applications* (págs 1-10). Dordrecht, The Netherlands: Springer.
- TABOADA, M., & GRIEVE, J. (2004). Analyzing Appraisal Automatically. *AAAI Technical Report SS-04-07* (págs 158-161). Presented at the American Association for Artificial Intelligence Spring Symposium on Exploring Attitude and Affect in Text, Stanford.
- THOMAS, M., PANG, B., & LEE, L. (2006). Get out the vote: Determining support or opposition from Congressional floor-debate transcripts. *Proceedings of EMNLP* (págs 327-335). Presented at the EMNLP.
- TURNER, P. D. (2002). Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)* (págs 417-424). Presented at the ACL 2002, Philadelphia, USA.
- TURNER, P. D., & LITTMAN, M. L. (2003). Measuring praise and criticism: Inference of semantic orientation from association. *ACM Trans. Inf. Syst.*, 21(4), 315-346. doi:10.1145/944012.944013
- WIEBE, J. M. (2000). Learning subjective adjectives from corpora. *Proceedings of the 17th National Conference on Artificial Intelligence* (pág 268-275). Menlo Park, CA: AAAI Press.
- WILSON, T., WIEBE, J., & HOFFMANN, P. (2009). Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational Linguistics*, 35(3), 399-433.