

FSAL: Lexicón de dominio específico para los Mercados Financieros Argentinos

Juan Pablo Braña, Alejandra M.J. Litterio, Alejandro Fernández Casco

Centro de Altos Estudios en Tecnología Informática (CAETI)
juan.brana@uai.edu.ar, alejandra.litterio@uai.edu.ar, alejandro.fernandezc@uai.edu.ar

Resumen

Existen múltiples escenarios que resultan propicios para desarrollar e implementar técnicas de *Machine Learning* para detectar cómo las noticias del ámbito de las finanzas impactan en los mercados y en el proceso de toma de decisiones cuando se trata de comprar o vender una acción u otro título financiero. En las últimas décadas, se ha estudiado cómo el Análisis de Sentimiento basado en lexicones en combinación con técnicas de *Machine Learning* puede ser utilizado para implementar estrategias de *Trading Algorítmico*. El presente trabajo tiene como objetivo mostrar que un lexicón de dominio específico especialmente diseñado por expertos en el área de finanzas (FSAL) obtiene mejores resultados que un lexicón de propósitos generales (SDAL). En primer lugar presentamos un “lexicón a medida”, en segundo lugar, mostramos, en base al experimento llevado a cabo, que nuestro lexicón supera los resultados obtenidos en comparación a los resultados de un lexicón de propósitos generales aplicado sobre un corpus compuesto por tweets previamente clasificado de manera colaborativa por expertos en finanzas. Nuestro enfoque metodológico se basa en una *perspectiva híbrida* donde se combina el uso de un lexicón a medida con estrategias de aprendizaje supervisado. Finalmente, presentamos algunas conclusiones preliminares y consideraciones sobre trabajos futuros.

Palabras clave: Sentiment Analysis, Lexicon Based, Spanish, Financial Markets, Algorithmic Trading

Introducción

En la última década la tecnología ha revolucionado los mercados financieros, un ejemplo de ello es el Trading Algorítmico, comúnmente definido como el empleo de algoritmos computacionales que automatizan ciertas decisiones de trading, cargan ordenes y las administran luego de ser enviadas al Mercado. Esta tecnología representaba al 2009 el 73% de las operaciones de trading diario en el mercado de EE.UU. Existen diferentes tipos de algoritmos que codifican distintos tipos de estrategias, muchas de ellas basadas en análisis fundamental, análisis técnico e incluso metodologías más modernas como *High Frequency Trading* (HFT). En años recientes se ha mostrado un especial interés en el análisis de noticias financieras

[1, 2, 3, 5, 8, 15, 19, 21, 24, 29] con diversos propósitos. Así, surge una nueva aproximación para la interpretación de estas noticias, esto es comprender la forma en que el tono o “sentimiento” de una noticia financiera impacta en los mercados y en el proceso de toma de decisiones cuando se trata de comprar o vender una acción u otro título financiero, o en otras palabras, cómo un inversor reacciona ante las noticias y cómo esas noticias se correlacionan con los precios del instrumento.

Nuestro trabajo se centra en el Análisis de Sentimiento desde una *perspectiva híbrida*, combinando el uso de un lexicón a medida con el uso de una estrategia de aprendizaje supervisado. En este método se utilizan diccionarios que describen la orientación semántica o polaridad (positiva o negativa) de la palabra. Sobre este diccionario se realizan cálculos mediante algoritmos computacionales de clasificación utilizando técnicas de *Machine Learning* para obtener la orientación del texto completo en base a la polaridad de las palabras incluidas en el mencionado diccionario [22,32].

El propósito aquí es presentar un lexicón diseñado dentro del dominio específico de las finanzas para Análisis de Sentimiento en Español, a la vez que focalizarnos a nivel experimental y problematizar que un lexicón de terminología específica (FSAL) obtiene mejores resultados al compararlo con un lexicón de términos generales en español (SDAL) cuando son aplicados a tweets referidos al mercado financiero en Argentina tomando como metodología la perspectiva híbrida. En primer lugar, en este estudio, nos referiremos a una aproximación general sobre Análisis de Sentimiento y sus aplicaciones en diferentes áreas, principalmente discutiremos su rol en relación al Trading Algorítmico. En segundo lugar, proponemos y evaluamos un lexicón desarrollado a medida que apunta al mercado financiero de habla hispana en Argentina. Hacia el final dedicaremos una sección para describir trabajos futuros.

1. Consideraciones generales sobre Análisis de Sentimiento

Análisis de Sentimiento (SA) también conocido como *Minería de Opinión* ha sido considerado como un caso especial del Procesamiento del Lenguaje Natural (NLP) ya que está estrechamente relacionado con el análisis semántico, semántica léxica, desambiguación de palabras, extracción de información y análisis del discurso [3,14]. Dada la importancia de la opinión pública y su impacto en la esfera política, social y económica, en la actualidad, tanto organizaciones gubernamentales, entidades privadas, financieras como el área de negocios se han centrado en las aplicaciones de SA. Así, identificar el tono “positivo” o “negativo” de un enunciado mediante un algoritmo se ha convertido en un tema crucial no sólo en el área de investigación sino también ha contribuido a rediseñar los negocios o modificar las tendencias en la vida social y política. Un ejemplo son las mediciones de intención de voto en EE.UU. e Inglaterra, que analizaban el sentimiento sobre la opinión de los candidatos por parte de los usuarios de las redes sociales, pudieron predecir con buenos niveles de certeza el resultado de las elecciones [25]. Una aplicación que nos interesa, en especial, y que ha motivado nuestra investigación refiere al mercado de valores, como el caso de la identificación de opiniones negativas y positivas en Twitter para predecir el movimiento de los índices en Wall Street, como DJ, S&P y NASDAQ [30, 31].

Ahora bien, es preciso mencionar que, por razones metodológicas que hacen al presente estudio, existen dos grandes enfoques de Análisis de Sentimiento. Por un lado, aquellos basados en lexicones que calculan la orientación semántica de un texto a partir de palabras o frases presentes en el mismo, las cuales han sido clasificadas de antemano. Por el otro, los centrados en aprendizaje supervisado que utilizan clasificadores entrenados a partir de textos previamente clasificados [22, 32]. De modo que, a efectos de desarrollar nuestro experimento consideramos fundamental integrar los enfoques anteriores en lo que se denomina *estrategia híbrida*. Así, hemos adoptado la aproximación de combinar un lexicón a medida, dependiente del dominio de las finanzas, cuyas palabras son indicadores de sentimiento extraídas de las opiniones de usuarios en las redes sociales, en este caso Twitter, y que su vez expresan la polaridad (positiva/negativa) contextual de estos enunciados, y técnicas de Machine Learning con algoritmos de clasificación para crear un modelo supervisado en el dominio específico.

Además, es importante destacar, con respecto a los lexicones, que si bien hay varias opciones en Inglés, siendo SentiWordNet [7] y LIWC [23] los más populares, para otros idiomas— el Español en

particular— las opciones son pocas. Como ejemplo de este último se cuenta con el Spanish DAL [20], un lexicón de propósito general para el análisis de sentimiento de textos en Español, que consiste en la réplica del Modelo de Whissell [26, 27] para la clasificación de palabras y textos en tres dimensiones: nivel de agrado, activación (activo/pasivo), imaginabilidad (pleasantness, activation, imagination). Por esto se ha decidido diseñar un lexicón de términos especializados—a medida— para el área de finanzas y en particular para el Mercado de Valores de Argentina y que constituye el tema central de nuestra investigación.

2. Análisis de Sentimiento y su relación con el Trading Algorítmico

El *trading algorítmico*, esto es, la carga de órdenes de compra/venta de títulos financieros en los mercados de manera automática mediante algoritmos computacionales, representa un alto volumen del trading diario, calculándose en aproximadamente el 75% en los mercados más importantes del mundo. La velocidad con la que un algoritmo reacciona frente a un indicador técnico, ante un patrón encontrado, y el gran volumen de datos y cálculos, que este tipo de programas son capaces de procesar en tiempo real, superan de manera significativa la capacidad de reacción de un humano frente a estas variables. Una de las ventajas más notables se presenta cuando un algoritmo es capaz de detectar e interpretar una noticia o acontecimiento en el mismo instante que se está produciendo y así conocer su correlación con los movimientos de precios con el mercado de valores.

El ejemplo más reciente del potencial de Análisis de Sentimiento en trading algorítmico se presentó cuando un grupo de piratas informáticos se apoderó de la cuenta de Twitter de Associated Press (@AP) donde se informaba de la explosión de dos bombas en la Casa Blanca. En ese preciso momento diversos algoritmos pudieron interpretar la semántica de dicho enunciado y comenzaron a vender diferentes títulos financieros provocando pérdidas considerables¹.

Ahora bien, para que esto suceda, los algoritmos basados en Análisis de Sentimiento deben poder monitorear noticias en tiempo real, determinar si tienen una polaridad positiva o negativa y establecer cómo repercuten las mismas en el Mercado de Valores. Con toda esa información deben tener la capacidad de tomar una decisión de compra/venta de un título y enviarla automáticamente al mercado.

¹Moore, H. and D. Roberts (2013). *AP Twitterhack causes panic in Wall Street and sends Dow plunging*. The Guardian Online, UK. <http://www.theguardian.com/business/2013/apr/23/ap-tweet-hack-wall-street-freefall> - último acceso 2015/11/1

En relación con esta temática, los mercados argentinos ya han comenzado a mostrar un especial interés en materia de Trading Algorítmico, particularmente en el Mercado de Valores de Buenos Aires (MERVAL) y otros mercados como el Mercado Argentino de Valores (MAV) o el Mercado de Futuros (ROFEX). Razón por la cual, resulta fundamental combinar estas estrategias para el análisis de noticias financieras utilizando, como se ha mencionado técnicas de SA y TA para poder predecir las reacciones de las opiniones de los usuarios de las cuentas de Twitter y su impacto en el Mercado, es decir, en un contexto situacional específico.

3. Lexicón a medida

Observemos ahora el siguiente texto extraído de distintas fuentes periodísticas nacionales, el cual despertó reacciones negativas en el público consumidor:

“YPF aumentó el precio de la nafta en un 4,5%”

Al analizar las opiniones o reacciones —posts en Twitter— desde diferentes contextos situacionales, se pudo detectar que imprimían un impacto diferente en los resultados. Por un lado, consideradas desde un contexto político y socio-cultural, se observó un alto índice de sentimiento negativo, el cual surgió del estudio de la composición semántica referida al aumento de precio al público consumidor, aun cuando el valor de las acciones de YPF ese mismo día subió considerablemente, representando un efecto positivo para los inversores. Ahora bien, para estas mismas opiniones analizadas desde otro contexto más específico, esto es, desde el mundo de las finanzas y mercados, y de acuerdo con los especialistas en esta materia, otra posible interpretación admitiría que si bien el sentimiento “popular” sobre esa noticia fue negativo, la repercusión en el Mercado de Valores fue altamente positiva ya que si aumenta el precio de la nafta, se espera que YPF aumente sus ingresos, por lo cual las acciones que cotizan en el mercado aumentan su valor.

Otro ejemplo interesante, se manifiesta en el titular a continuación:

“Espectacular éxito del Gobierno en el allanamiento de las cuevas financieras para bajar el dólar Blue”

Claramente, esta frase tiene una estructura semántica positiva fácilmente detectable por cualquier algoritmo, aunque la repercusión en el mercado fue sensiblemente negativa. Aquí operaría el mismo principio de análisis mencionado con anterioridad, por lo cual podemos decir que la tendencia “positiva” o “negativa” de un texto además de depender del estudio semántico es contextualmente dependiente.

Dada entonces las características en los textos, los contextos en los que la noticia y comentarios se

generan, la repercusión sea esta positiva o negativa, entre otros factores, y debido a que el mundo de las Finanzas posee un lenguaje técnico muy específico y, tanto la interpretación como la clasificación de las noticias en positivas o negativas debe ser analizada por expertos en la materia, el *Trading Algorítmico* basado en Análisis de Sentimiento en nuestro modelo sienta sus bases en un lexicón compuesto por un diccionario de palabras financieras específicas para el mercado Argentino diseñado por especialistas en Lingüística y Análisis del Discurso y un corpus de noticias clasificadas por analistas financieros, economistas, expertos en política económica, entre otros.

3.1. Características del Lexicón

Como primer paso en la construcción de nuestro lexicón, se extraen del corpus las palabras más relevantes para el estudio de Análisis de Sentimiento. Este corpus está compuesto, a su vez, por una selección representativa de tweets. Luego, a cada una de esas palabras se les asigna un peso que refleja su polaridad positiva o negativa. Dicho peso se encuentra en un rango discreto entre [-5,5], siendo -5 la carga más negativa y 5 la carga más positiva.

El lexicón está conformado por palabras “llenas” o “léxicas” [10, 11], aquellas que poseen un contenido descriptivo inherente es decir que expresan una idea o conceptos y pueden describirse mediante rasgos semánticos, donde hemos excluido palabras “vacías” o “funcionales” [10, 11] las cuales no expresan una idea y su significado es fundamentalmente gramatical que se precisa en relación con otras palabras que transmiten información categorial o entre las que establecen relaciones sintácticas. Por otra parte, consideramos aquellas palabras o ítems léxicos que se “activan” en una situación comunicativa especializada, en este caso la interacción discursiva en los mercados financieros. Así, la información semántica que compone el significado especializado se distingue del no especializado por su configuración. En este sentido, observaremos que una de las propiedades centrales del término especializado refiere a la mayor especificidad referencial que se desprende de la cantidad y tipo de información activada. Asimismo, consideraremos que la configuración semántica de estos términos es más densa y más compleja ya que presentan varios niveles de relación de imbricación entre las predicaciones activadas. [4]

Además de esta categorización, decidimos trabajar en un segundo nivel de análisis, que se corresponde con el valor relativo que las palabras pueden tener. Sostenemos que las palabras tienen un sesgo implícito de valoración o modalización y de acuerdo a su ordenamiento y selección pueden brindar perspectivas o maneras de valorar lo que se dice, esto es asignarle un valor al texto. En este nivel tomaremos como punto

de partida los postulados de la Teoría de la Valoración [16, 17], desarrollada dentro del enfoque de la Lingüística Sistémico Funcional [6, 12], que se ocupa particularmente del lenguaje (la expresión lingüística) de la valoración, la actitud y la emoción, y del conjunto de recursos que explícitamente posicionan de manera interpersonal las propuestas y proposiciones textuales. Dentro de la Teoría de la Valoración consideramos los recursos evaluativos divididos en tres dominios semánticos: actitud (apreciación, juicio, afecto), orientación/polaridad (positiva/negativa), gradación (foco/fuerza). Así, en base a lo mencionado, nuestro lexicón se compone de términos especializados en economía y finanzas—en su acepción más completa “lexicón económico-financiero relacionado con terminología del discurso político y jurídico”— con los que se puede determinar la actitud, es decir, significados por los cuales los textos/hablantes atribuyen un valor o una evaluación intersubjetiva a los participantes y los procesos, ya sea relacionados con respuestas emocionales o con sistemas de valores culturalmente determinados (afecto, juicio, apreciación). Cabe mencionar que nuestro lexicón actualmente, y al que de ahora en más denominaremos “FSAL”, cuenta con más de 3023 palabras.

3.2. Sistema de filtrado

A fin de eliminar aquellas palabras que no aportan un significado contextual específico al dominio en el que estamos trabajando, hemos decidido aplicar un sistema de exclusión–inclusión de términos (filtrado), donde en primer lugar, se eliminaron aquellos términos que no se corresponden con la temática principal por sus características de generalidad, y en segundo lugar, aquellos que tienen un significado gramatical, que se precisa solamente en relación con otras palabras de las que transmiten información categorial o entre las que establecen relaciones sintácticas. Consideramos, entonces, excluir sustantivos comunes que no tienen un significado semántico con una valoración intrínseca, esto es, aquellos que consideramos “neutros”, por ejemplo: gente, nombre, uso, etc.; así como también sustantivos propios como nombres de personas, países, lugares geográficos, además de gentilicios, siglas, adverbios de lugar y tiempo, el verbo “haber” en todos sus formas (desinencias verbales) y los que denominamos verbos de estado (ser, estar). Por otra parte, se excluyen determinantes (artículos, demostrativos y posesivos) y cuantificadores, conjunciones coordinantes y subordinantes, preposiciones (plenas, semiplenas, y gramaticalizadas).

Para evitar la repetición, omisión e incorporación de un término que no se corresponda con las categorías gramaticales o no contengan en sí mismo una carga

valorativa, se ha desarrollado una herramienta de asistencia al curado del lexicón, la cual nos permite revisar aquellos términos que se encuentran en los tweets relacionado con noticias o comentarios en finanzas que componen el corpus y agregar al listado los términos que fueron omitidos en una primera selección manual, además de asignarle el peso correspondiente de acuerdo a la escala establecida.

4. Discusión

Como se ha señalado, el tema central de nuestro trabajo radica en mostrar que un lexicón hecho a medida para el Mercado de Valores de Buenos Aires (FSAL) permite obtener mejores resultados de clasificación que si se utiliza un lexicón de propósito general (SDAL). En este sentido hemos llevado a cabo un estudio comparativo, que toma como punto de partida el contraste y resultados obtenidos al utilizar ambos lexicones. Así, el procedimiento consiste en, dado un corpus de tweets determinado, previamente clasificado por expertos y manteniendo como constante nuestros algoritmos de aprendizaje supervisado basados en Machine Learning, variar el lexicón, esto es FSAL y SDAL, lo cual nos permitirá a nivel comparativo determinar la efectividad de lexicones así como el grado de certeza en la clasificación de nuestro lexicón (FSAL) al contrastarlo con el de propósito general (SDAL).

4.1. El corpus

Con el fin de construir nuestro corpus de análisis hemos recolectado tweets cuyo contenido es significativamente relevante para los mercados financieros argentinos. Estos tweets provienen de una selección de cuentas de Twitter de comunidades de expertos de reconocida trayectoria en el ámbito económico-financiero.

Además, dichos tweets fueron clasificados de manera colaborativa por especialistas financieros. La plataforma diseñada para este propósito, consta de una interfaz web, donde diversos especialistas cuentan con un usuario y clave para ingresar. Con esta herramienta, los expertos tienen la oportunidad de ver cada uno de los enunciados y determinar si poseen una carga positiva o negativa en relación al impacto que tiene dicho enunciado en el Mercado. Por otra parte, hemos desechado aquellos tweets cuyo sentimiento no era posible determinar o eran de carácter neutro. Como resultado, se obtuvo un corpus constituido por 500 tweets (250 con carga positiva y 250 con carga negativa).

4.2. Algoritmos de clasificación

A efectos de evaluar el desempeño de nuestros algoritmos en cuanto al grado de certeza en la

clasificación de tweets utilizando tanto el lexicon genérico como nuestro lexicon de términos especializados, hemos recurrido a diferentes instancias.

Los algoritmos de clasificación que hemos utilizado son: Naive Bayes, Random Forest y Support Vector Machines. Tanto el procesamiento de datos y procesamiento de textos, como la implementación de los algoritmos, fueron realizados con la plataforma de análisis estadístico R².

5. Resultados

Para evaluar la efectividad del análisis se aplica la técnica de validación cruzada (X-Validation, con X = 10) comúnmente utilizada en pruebas en el dominio del aprendizaje supervisado. Se inicia con un corpus de textos (tweets) que han sido previamente clasificados en positivos o negativos. A diferencia de otros enfoques que también utilizan aprendizaje supervisado, en nuestro caso, para el armado del vector de características (*feature vector*) sólo se tienen en cuenta las palabras existentes en el lexicon. El clasificador resultante se emplea para predecir el sentimiento de los textos en la partición de prueba. Finalmente, se reportan los aciertos y errores tanto para los textos negativos como los positivos.

Los resultados que se observan en la Tabla 1 corresponden a la ejecución de validación cruzada (para ello se utiliza la implementación de cross-validation provista por la librería Caret de R³). Cada ejecución utiliza una partición diferente del corpus, siempre en una relación 90% para entrenamiento, 10% para prueba.

	SDAL	FSAL
Naive Bayes (POS)	64,80%	80,00%
Naive Bayes (NEG)	66,07%	71,53%
Random Forest(POS)	67,36%	83,81%
Random Forest(NEG)	76,34%	76,52%
SVM(POS)	61,59%	82,29%
SVM(NEG)	65,66%	71,63%

Tabla 1. Resultados

Se observa que, para todos los algoritmos utilizados, nuestro lexicon FSAL resultó en una mejor clasificación que el SDAL.

Cabe señalar que nuestra estrategia de clasificación supervisada sólo considera las palabras presentes en el lexicon. Por lo tanto, sería posible que la diferencia en resultados se deba a una diferencia en el número de palabras de cada lexicon que son relevantes al corpus. Para descartar esta posibilidad, hemos observado dicha variable, obteniendo un valor de 362 palabras del

SDAL presentes en el corpus, contra 391 del FSAL. Es decir que 12.87% de las palabras del SDAL están presentes en el corpus de estudio, mientras que para el FSAL el porcentaje de palabras presentes es de 15.95%. Si bien existe una diferencia en el número de palabras, entendemos que no es suficiente como para explicar la diferencia en los resultados. Complementariamente, la Tabla 2 ofrece un resumen de la presencia de palabras de cada lexicon en los tweets del corpus. Se observa que 31 tweets del corpus no tienen ninguna palabra (columna 0) del SDAL, mientras que sólo 15 no tienen ninguna palabra del FSAL. La Tabla 3 lista las palabras de cada lexicon que aparecen al menos 10 veces en el corpus. Puede observarse que son menos (y más específicas a finanzas) las palabras del FSAL cuya frecuencia impacta en el resultado. Asimismo, puede observarse que el SDAL (por su carácter de lexicon de propósito general) asigna una carga valorativa a palabras que no necesariamente la tienen en el dominio de finanzas (por ejemplo: una, pero, vía).

Palabras	0	1	2	3	4	5	6	7+
SDAL	31	45	55	39	29	25	8	4
FSAL	15	62	82	40	21	11	3	2

Tabla 3: Presencia de los lexicones en los tweets

FSAL		SDAL	
Palabra	Frecuencia	Palabra	Frecuencia
no	87	dólar	43
mas	33	mas	33
inflación	31	bolsa	30
baja	29	una	30
default	14	gobierno	25
deuda	14	peso	18
problema	14	mercado	16
sube	13	deuda	14
caída	11	país	14
		problema	14
		economía	12
		vía	12
		caída	11
		pago	11
		año	10
		emisión	10
		hasta	10
		ley	10

Tabla 2: Palabras con frecuencia de 10 o más

6. Conclusiones y trabajo futuro

En el presente estudio se han mostrado las relaciones que existen entre Análisis de Sentimiento y Trading Algorítmico y cómo un lexicon diseñado a medida de terminología específica utilizado en los mercados Argentinos mejora los resultados si lo comparamos con uno de propósito general para la clasificación de noticias en positivas y negativas.

Hasta el momento hemos logrado construir el lexicon a medida (FSAL), y desarrollado los

²Disponible en: <http://r-project.org>

³ Librería Caret R (Classification and Regression Training). <https://cran.r-project.org/web/packages/caret/index.html>

algoritmos que clasifican con un nivel de certeza aceptable los enunciados en positivos y negativos. También hemos mostrado la necesidad de crear una *interfaz colaborativa* para la clasificación manual de noticias debido a que la interpretación de si una noticia impacta de manera positiva o negativa no siempre es trivial y se necesita un profundo conocimiento del sistema financiero para dicha tarea. Es imprescindible enfatizar que aún nos encontramos en la fase de desarrollo y entrenamiento de algoritmos a nivel semántico, es decir de la unidad léxica que se “activa” en una situación de comunicación determinada, esto es del “valor de la palabra” al “valor del término” [4].

Nuestro próximo paso es correlacionar el índice de sentimiento calculado con los movimientos de precios del mercado y llevar a cabo una simulación para demostrar que un algoritmo de Análisis de Sentimiento puede servir como estrategia en la toma de decisiones de compra/venta de instrumentos financieros. De modo que, una vez finalizadas todas las tareas descritas crearemos una interfaz web, donde diversos inversores puedan probar esta tecnología.

Referencias

- [1] Bollen, J., Mao, H., and Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1-8.
- [2] Brown, E. (2012). Will twitter make you a better investor? A look at sentiment, user reputation and their effect on the stock market. En actas de SAIS.
- [3] Chen, R., and Lazer, M. (2011). Sentiment analysis of twitter feeds for the prediction of stock market movement.
- [4] Ciapuscio, G. et al. (2009). *De la palabra al texto. Estudios lingüísticos del español*. Buenos Aires: Eudeba.
- [5] Devitt A. and Ahmad, K. (2007). Sentiment Polarity Identification in Financial News: A Cohesion-based Approach. Prague, Czech Republic: Association for Computational Linguistics.
- [6] Eggins, S. (2002). *Introduction to Systemic Functional Linguistics*. London and New York: Pinter Publishers.
- [7] Esuli A. and Sebastiani F. (2006). SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining. En actas de the 5th Conf. Language Resource Evaluation, pp. 417–422.
- [8] Feldman, R.; Rosenfeld, B.; Bar-Haim, R., and Fresko, M. (2011). The stock sonar—sentiment analysis of stocks based on a hybrid approach. En Twenty-Third IAAI Conference.
- [9] Finter, P.; Alexandra Niessen-Ruenzi & Stefan Ruenzi (2010). *The Impact of Investor Sentiment on the German Stock Market*.
- [10] Giammatteo, M. y Albano, H. (2006). *¿Cómo se clasifican las palabras?* Buenos Aires: Littera Ediciones.
- [11] Giammatteo, M y Albano, H. (2009). *Lengua, Léxico, Gramática y Texto. Un enfoque para su enseñanza basado en estrategias múltiples*. Buenos Aires: Biblos.
- [12] Halliday, M.A.K. (1994). *An Introduction to Functional Grammar*. London: Edward Arnold Publishers.
- [13] Hendershott, T., Jones, C.M. and Menkveld A. (2011). Does Algorithmic Trading Improve Liquidity? *The Journal of Finance*, 66 (6): 1-34.
- [14] Liu, B. (2015). *Sentiment Analysis. Mining Opinions, Sentiments and Emotions*. New York: Cambridge University Press.
- [15] Mao, H; Gao, P.; Wang, Y. and Bollen, J. (2014). Automatic Construction of Financial Semantic Orientation Lexicon from Large-Scale Chinese News Corpus. *7th Financial Risks International Forum*, 20 (20): 1-18
- [16] Martin, J.R. and White, P.R.R. (2005). *The Language of Evaluation, Appraisal in English*. London and New York: Palgrave Macmillan.
- [17] Martin, J. R. and D. Rose (2003). Appraisal. En *Working with Discourse. Meaning beyond the clause*. London: Continuum.
- [18] Pang, B. and Lee L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135
- [19] Rao, T., & Srivastava, S. (2012). Analyzing stock market movements using twitter sentiment analysis. En actas de the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012) (pp. 119-123). IEEE Computer Society.
- [20] Ríos A., Gravano M., and Gravano A. (2013). Spanish DAL: A Spanish Dictionary of Affect in Language. En actas de the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. *Association for Computational Linguistics*.
- [21] Sul, K. A.R. Dennis and L. Yuan (2014). Trading on Twitter: The Financial Information Content of Emotion in Social Media. *System Sciences (HICSS)*, 47th Hawaii International Conference.
- [22] Taboada, M.; Brooke, J.; Tofiloski, M.; Voll, K. and Stede, M. (2011). Lexicon-Based Methods for Sentiment Analysis. *Association for Computational Linguistics*, 37 (2): 267-307.
- [23] Tausczik, Y. R. and Pennebaker, J. W. (2010). “The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods,” *J. Lang. Soc. Psychol.*, 29 (1): 24–54.
- [24] Tetlock, Paul C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3):1139–1168.
- [25] Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *ICWSM*, 10, 178-185.
- [26] Whissell, C., M. Fournier, R. Pelland, D. Weir, and K. Makarec. (1986). *A dictionary of affect in language: IV. Reliability, validity, and applications. Perceptual and Motor Skills*.
- [27] Whissell, C. (1989). The dictionary of affect in language. *Emotion: Theory, research, and experience*, 4:113–131.
- [28] White, P.R.R. (2006). Evaluative Semantics and Ideological Positioning in Journalistic Discourse. En *Image and Ideology in the Mass Media*, Lassen, I. (ed.), Amsterdam/Philadelphia, John Benjamins: 45- 73.
- [29] Željko, A.; Nikola Ljubešić and Marko Tadić (2010). Towards Sentiment Analysis of Financial Texts in Croatian. En actas de the Seventh International Conference on Language Resources and Evaluation. European Language Resources Association.
- [30] Zhang, L. (2013). Sentiment Analysis on Twitter with stock price and significant keyword correlation. Ph D. Dissertation.
- [31] Zhang, W. and Steven, S. (2010). Trading strategies to exploit blog and news sentiment. The 4th International AAAI Conference on Weblogs and Social Media.
- [32] Zhang, L. et al. (2011) Combining Lexicon - based and Learning - based Methods for Twitter Sentiment Analysis. HP Laboratories.