Lingüística y Literatura
ISSN: 0120-5587
ISSN: 2422-3174
Universidad de Antioquia

Pérez Pérez, Carlos Mario; Quiroz Herrera, Gabriel Ángel; Tamayo Herrera, Antonio Jesús

¿TIENE CARÁCTER PREDICTIVO LA ESTRUCTURA PREDICATIVA [VERBO + OBJETO DIRECTO]? HACIA UNA CARACTERIZACIÓN SINTÁCTICO-SEMÁNTICA PARA PROPÓSITOS DE ANÁLISIS DE SENTIMIENTOS1

Lingüística y Literatura, núm. 78, 2020, Julio-Diciembre, pp. 11-34
Universidad de Antioquia

DOI: https://doi.org/10.7440/res64.2018.03

Disponible en: https://www.redalyc.org/articulo.oa?id=476569499001

¿TIENE CARÁCTER PREDICTIVO LA ESTRUCTURA PREDICATIVA [VERBO + OBJETO DIRECTO]? HACIA UNA CARACTERIZACIÓN SINTÁCTICO-SEMÁNTICA PARA PROPÓSITOS DE ANÁLISIS DE SENTIMIENTOS[1]

Carlos Mario Pérez Pérez
Universidad de Antioquia (Colombia)
carlosm.perez@udea.edu.co

Gabriel Ángel Quiroz Herrera
Universidad de Antioquia (Colombia)
gabriel.quiroz@udea.edu.co

Antonio Jesús Tamayo Herrera
Universidad de Antioquia (Colombia)
antonio.tamayo@udea.edu.co

Recibido: 15/11/2019 - **Aprobado:** 03/01/2020
DOI: doi.org/10.17533/udea.lyl.n78a01

Resumen: Tradicionalmente, el análisis de sentimientos se ha centrado en la clasificación de textos cortos con unidades lingüísticas simples como el adjetivo. Sin embargo, otras unidades, como elementos predicativos, podrían ser discriminantes de elementos discursivos más grandes como las noticias de prensa. Así, en este artículo, se caracteriza la estructura predicativa [verbo + objeto directo] en un corpus de noticias sobre pobreza en tres diarios colombianos. Se concluye que este tipo de unidades tienen un carácter discriminante en el dominio de este corpus.
Palabras clave: análisis de sentimientos; verbo; complemento de objeto directo; pobreza en Colombia; lingüística de corpus.

1. Artículo resultado del trabajo de investigación en el marco del proyecto *Variación terminológica vertical y horizontal en español: hacia una descripción lingüística de las unidades terminológicas desde la lingüística de corpus*, inscrito en el CODI (Comité para el Desarrollo de la Investigación), Universidad de Antioquia, Colombia. Acta 2018-22970 del 27 de marzo de 2019. Agradecimientos a los miembros del grupo de investigación en Traducción y Nuevas Tecnologías - TNT, perteneciente a la Escuela de Idiomas de la Universidad de Antioquia, por sus aportes en las distintas etapas de la investigación.

LINGÜÍSTICA Y LITERATURA
ISSN 0120-5587
E-ISSN 2422-3174
N. º 78, 2020 11-34

DOES PREDICATIVE STRUCTURE [VERB + DIRECT OBJECT] HAVE PREDICTIVE CHARACTER? A SYNTACTIC-SEMANTIC CHARACTERIZATION FOR SENTIMENT ANALYSIS PURPOSES

Abstract: Traditionally, sentiment analysis has focused on processing less complex textual units, as sentences, through detailed characterization of their components like adjectives. However, there are other types of units, such as predicative structures, that might be discriminant elements of more complex discursive units like the news. Thus, in this article, word occurrences of the predicative structure [verb + direct object] are extracted and characterized from a corpus on poverty compiled from Colombian newspapers. The results demonstrate that these units are discriminant and might help with polarity classification.

Key words: sentiment analysis; verb; direct object complement; poverty in Colombia; corpus linguistics.

1. Introducción

En las últimas décadas, el interés por el análisis de sentimientos (AS) ha aumentado gracias a su automatización en ámbitos académicos e industriales. En términos generales, el AS es el campo de estudio que analiza las opiniones de las personas, sus sentimientos, evaluaciones, apreciaciones, actitudes y emociones sobre productos, servicios, organizaciones, individuos, asuntos, eventos, temas y atributos (Liu, 2012, p. 1). Asimismo, el AS es un área del procesamiento del lenguaje natural (PLN) con aplicaciones en la minería de datos. Las distintas investigaciones en el área han enfocado los análisis, principalmente, en la identificación de opiniones positivas o negativas en los corpus analizados, omitiendo el neutro como categoría de polaridad, por razones tanto teóricas como funcionales (Tan, Wang, & Cheng, 2008; Ding, Liu, & Yu, 2008; Taboada, Brooke, Tofiloski, Voll, & Stede, 2011).

Los adjetivos han sido la unidad de análisis por excelencia, puesto que se ha logrado establecer que estos son portadores de la opinión al igual que determinar su correlación con frases subjetivas (Hatzivassiloglou & McKeown, 1997; Hatzivassiloglou & Wiebe, 2000; Mullen & Collier, 2004; Whitelaw, Garg, & Argamon, 2005). En menor medida, otros estudios buscan identificar la carga de opinión desde otras categorías gramaticales tales como sustantivos subjetivos (Riloff, Wiebe, & Wilson, 2003), verbos y adverbios (Benamara, Cesarano, Picariello, Reforgiato, & Subrahmanian, 2007) y el adjetivo unido a otras categorías gramaticales como el adverbio (Turney, 2002).

LINGÜÍSTICA Y LITERATURA
ISSN 0120-5587
E-ISSN 2422-3174
N. º 78, 2020 11-34

En la literatura sobre el AS, se puede identificar claramente que el enfoque de análisis se ha dirigido, en mayor medida, a corpus de tipologías textuales como revisiones y comentarios en sitios web que se decantan por un nivel oracional; mientras solo se reportan algunos análisis a nivel de documento y, en menor medida, en ámbitos de mayor complejidad semántica como la de los textos de carácter socio-político, además de tomar como unidad de estudio categorías gramaticales distintas al adjetivo (Tamayo, 2018; Joshi & Penstein-Rosé, 2009; Ng, Dasgupta, & Arifin, 2006). En esta misma línea, no se reportan investigaciones que aborden sistemáticamente el neutro como categoría de polaridad, aunque algunas sí mencionan metodologías para excluirlo de los análisis finales sin aportar los resultados encontrados (Taboada *et al*. 2011; Koppel & Schler, 2006).

En este artículo, se pretende ofrecer un acercamiento descriptivo del análisis sintáctico-semántico de la estructura predicativa [verbo + objeto directo] extraída de artículos de prensa escrita en tres periódicos colombianos: *El Tiempo, El Espectador, El Colombiano,* publicados entre los años 2010 y 2014. Se asignó a cada ocurrencia recuperada del corpus el sentido y el tipo de verbo, el área temática del primer complemento del predicado (objeto directo) y la carga de polaridad *positiva, negativa y neutra*. Lo anterior, constituiría un avance en el campo del AS no solo en el tipo de texto, la unidad de análisis y la polaridad, al incluir el *neutro* como variable, sino también en la caracterización de esta estructura como predictiva del fenómeno analizado (la pobreza).

2. Marco conceptual

Aunque son vastas las aplicaciones del AS en áreas que van desde lo gubernamental hasta ámbitos comerciales e individuales, son pocas las investigaciones que, teniendo el español como idioma de base, enfocan sus análisis en textos completos con dificultad semántica mayor (Tamayo, 2018; Nasukawa & Yi, 2003). Asimismo, son pocos los estudios en AS que toman como unidad de análisis categorías gramaticales distintas al adjetivo y, adicionalmente, las investigaciones centradas en las cargas de polaridad *positivo* y *negativo* han sido las más destacadas, dejando a un lado al *neutro* como posible categoría de análisis. Después de varios análisis iniciales con los datos recuperados del corpus de trabajo, se definió la estructura

LINGÜÍSTICA Y LITERATURA
ISSN 0120-5587
E-ISSN 2422-3174
N. º 78, 2020 11-34

predicativa [verbo + objeto directo] en textos de prensa colombianos sobre pobreza como objeto de estudio de este artículo. En la elección de los periódicos, se tuvo presente si estos podían aportar información relevante que permitiese fijar la agenda política del país (Chomsky, 1997) para lo cual se siguieron criterios de conocimiento contextual por parte de los diferentes grupos de trabajo de cada país en el marco del proyecto *Poverty, Language and Media* (PoLaMe)[2] (Quiroz, Tamayo & Zuluaga 2017, p. 26). A continuación, se listan algunas precisiones conceptuales entorno a cada componente de la estructura mencionada.

Para entender el concepto de verbo, desde una perspectiva lingüística, se tomó la definición desde la gramática descriptiva:

> Si por flexión se entiende las variaciones de forma de una palabra para manifestar distintos significados gramaticales o funcionales, el verbo es la palabra flexiva por excelencia: por el número de significados, de tiempo, aspecto y modo (TAM), y de número y persona (NP); y por las variaciones que de tales significados pueden expresar las distintas formas de un verbo. El verbo es una clase de palabras que significan un evento, una acción, proceso o estado (Alcoba, 1999, p. 4917).

Para la categorización de cada ocurrencia, se definió la información relacionada con el verbo: tipología (Lorente, 2002), sentido (Levin, 2003; en el proyecto *WordNet*) y polaridad (Pérez Pérez, 2018).

En primera instancia, las categorías expuestas por Lorente (2002, § 5), permiten identificar el papel desempeñado por cada uno de los verbos en un discurso especializado. Así, los verbos discursivos, del tipo *decir, explicar, indicar, señalar, considerar,* se relacionan con las estrategias propiamente textuales empleadas por el emisor, con las cuales se busca activar la competencia pragmática. Por el contrario, los verbos conectores, como *afectar, cambiar, dar, entregar, pedir,* hacen parte de la expresión del conocimiento especializado, aunque no lo proporcionan teniendo en cuenta que su función, como lo indica su denominación, es la de expresar relaciones de equivalencia, igualdad, similitud, dependencia o atribuir cualidades o valores. Los verbos fraseológicos, *aumentar, bajar, combatir, medir, mejorar,* expresan acciones, procesos y estados en los contextos en los cuales confluyen con sus respectivos complementos directos, es decir, cuando forman parte de unidades sintagmáticas en las que adquieren valor especializado y, por lo tanto, transmiten el conocimiento especializado. Por

2. Para mayor información, véase http://www.uib.no/en/project/polame y http://grupotnt.udea.edu.co/polame-corpus/

LINGÜÍSTICA Y LITERATURA
ISSN 0120-5587
E-ISSN 2422-3174
N. º 78, 2020 11-34

último, los verbos-término están ligados, exclusivamente, al tipo de conocimiento específico de cada área.

En segunda instancia, para el sentido verbal, se utilizó la propuesta de etiquetaje semántico de Levin (1993, pp. 111-313) retomada en el proyecto *WordNet*[3], según la cual los verbos pueden agruparse en categorías semánticas al considerar una serie de diátesis (argumentos verbales). De este modo, el autor estableció las siguientes categorías: *body*, *change*, *cognition*, *communication*, *competition*, *consumption*, *contact*, *creation*, *emotion*, *motion*, *perception*, *possession*, *social*, *stative* y *weather*.

En tercera instancia, para asignar la polaridad, se definió una escala con variables discretas, esto es: *positivo* (1), *negativo* (-1) y *neutro* (0). Además, para puntualizar aquello que se entiende por *positivo* o *negativo*, se utilizó la definición de estos términos en el *Diccionario de la lengua española* (2017) y se contrastaron con las presentadas en el *Diccionario combinatorio práctico del español contemporáneo* (2006) de Ignacio Bosque y el *Diccionario de uso del español* (2007) de María Moliner. De este modo, teniendo la *pobreza* como eje temático, se entiende lo *positivo* como aquello que favorece la eliminación de esta condición o el mejoramiento de los criterios con la cual se mide, como puede verse en *aumentar empleo permanente, bajar pobreza extrema, cerrar brecha desigualdad*. Mientras lo *negativo* se entiende como aquello que permite que el fenómeno aumente o se mantenga, es decir, que haya más pobreza. A modo de ejemplo: *afectar empleabilidad*, *mantener pobreza* y *perder dinero.*

Para la asignación de la polaridad del verbo, se siguió un criterio lexicográfico, es decir, se tomó la información contenida en la primera acepción de cada palabra consultada y que correspondiera a los ejes temáticos del análisis y se asignó la carga de polaridad según los argumentos contenidos en ella puesto que, según el orden de indexación descendente, la primera corresponde a la de mayor uso en términos de frecuencia seguida de las de menor frecuencia.

Con relación al objeto directo, se comenzó por definir el núcleo del objeto directo como «el sustantivo nuclear de un sintagma nominal presente en el mismo» (Pérez Pérez, 2018, p. 54). Además, se definieron cuatro áreas temáticas bajo las cuales se agruparon los distintos núcleos de objeto directo, a saber: *economía, sociedad, política* y *general.* Por último, se asignó la carga

3. Para mayor información, véase https://wordnet.princeton.edu/

LINGÜÍSTICA Y LITERATURA
ISSN 0120-5587
E-ISSN 2422-3174
N. º 78, 2020 11-34

de polaridad siguiendo la escala y criterios arriba mencionados y descritos en detalle en la metodología.

3. Metodología

Para llevar a cabo la investigación, se utilizó el corpus sobre pobreza del proyecto PoLaMe, conformado por más de 25 millones de palabras en español. Para el caso de Colombia, el universo de datos quedó conformado por un total de 7 053 artículos (3 782 en *El Tiempo,* 829 en *El Colombiano* y 2 442 en *El Espectador*) (Chiquito & Quiroz, 2017, p. 87).

Para la elección de los textos que conformaron los datos de trabajo, se realizó un muestreo aleatorio estratificado, teniendo como criterios mínimos de extracción los siguientes: 1) la inclusión de la palabra *pobreza*; 2) artículos publicados entre los años 2010 y 2014; y 3) con una extensión mínima de 300 palabras y máxima de 1 000. La muestra se conformó por un total de 205 artículos (138 973 palabras), cuya distribución final se muestra en la Tabla 1, adaptada de Quiroz, Chiquito y Zuluaga (2017, p. 87).

	# de artículos (universo PoLaMe)	**# de artículo (muestra PoLaMe)**	**# de palabras (muestra PoLaMe)**	**# de artículo (muestra trabajo)**	**# de palabras (muestra trabajo)**
El Tiempo	3 782	346	206 553	56	30 239
El Colombiano	829	312	268 274	65	61 722
El Espectador	2 442	332	17 730	84	47 012
Total	**7 053**	**990**	**492 557**	**205**	**138 973**

Tabla 1. *Del universo de datos a la muestra de trabajo*

Una vez se conformó la muestra de trabajo, se procedió con la extracción de las ocurrencias de la estructura predicativa [verbo + objeto directo], mediante el uso del *parser* de *Freeling*[4] y *Scripts* en *Python*. Se recuperaron un total de 6 283 ocurrencias, de las cuales se tomó el 50 % para el análisis, como se explica en la Fase 4. A continuación, se describen las cuatro fases de la depuración de los datos:

4. Para mayor información, véase http://nlp.lsi.upc.edu/freeling/index.php/node/1

LINGÜÍSTICA Y LITERATURA
ISSN 0120-5587
E-ISSN 2422-3174
N. º 78, 2020 11-34

—**Fase 1.** El 100 % de las ocurrencias se organizaron en un archivo de Excel para proceder con el etiquetaje sintáctico-semántico. En principio, se asignaron los metadatos necesarios para la identificación de cada ocurrencia al interior de la base de datos: ID del artículo al interior de la base de datos, URL del artículo, periódico de publicación, artículo completo, como se muestra en la Figura 1. Por último, el etiquetaje se realizó de forma manual, lo que permitió establecer una propuesta metodológica de depuración de la información para la conformación de una base de datos con fines de AS, teniendo como referencia la unidad de análisis trabajada.

—**Fase 2.** Corresponde a la cantidad de ocurrencias restantes posterior a la primera depuración (un total de 3 563 de 6 283).

—**Fase 3.** En la cual se llevó a cabo una segunda depuración posterior a las indicaciones dadas por los expertos que validaron los datos. Se eliminaron 843 ocurrencias, de manera que quedaron un total de 2 720 que conformaron la base de datos.

—**Fase 4.** Tomando la base de datos final de la fase anterior, se seleccionó el 50 % de estas ocurrencias con las cuales se realizaron los análisis. Esta decisión fue de carácter metodológico teniendo en cuenta la delimitación del tiempo y el etiquetaje manual semántico y de polaridad que se realizó. Para ello, se filtró la base de datos por frecuencia de aparición del verbo y se tomó la muestra hasta completar el 50 % de los datos.

3.1. Metodología para la construcción de una base de datos [verbo + objeto directo]

Esta propuesta metodológica aporta a la solución de los problemas inherentes al AS, al constituirse en un aporte significativo en cuanto podría proporcionar elementos que permitirían una extracción de datos con menor ruido (entendido como los datos que no corresponden a los criterios de extracción). De este modo, la propuesta metodológica consta de tres momentos: 1) extracción de las ocurrencias de la estructura predicativa [verbo + objeto directo]; 2) depuración de los datos; y 3) etiquetado. Asimismo, es importante mencionar que esta propuesta, de implementarse a otro tipo de unidades de análisis distintas al aquí trabajado, debería ser adaptada para cumplir las necesidades específicas.

3.1.1. Extracción de las ocurrencias

El trabajo se limitó al análisis de las ocurrencias de la estructura predicativa de tipo [verbo + objeto directo], extraídas por medio del uso del *parser* de *Freeling* y *Scripts* en *Python,* conforme a los objetivos de investigación, los cuales se plantearon en términos de caracterizar y depurar estas ocurrencias y describir el cómo se aborda el fenómeno de la pobreza en la prensa colombiana. Otros patrones poliléxicos como aquellos que contienen el atributo de los verbos copulativos o pseudocopulativos o adjetivos y complementos adjetivales no se tuvieron en cuenta puesto que por sus características pueden ser analizados en futuras investigaciones en el campo. Se listan algunos ejemplos de la información recuperada: *abordar aquellas trampas, disminuir los niveles de pobreza, generar empleo, implementar programa social, incrementar presupuesto, mejorar calidad de vida, reducir índice de pobreza extrema, superar carencia básica, trabajar inclusión social.*

3.2.2. Depuración de los datos

Una vez obtenidos los datos de análisis se procedió con la depuración de los mismos. En los numerales siguientes, se detallan los elementos que se eliminaron de la muestra de trabajo junto a las razones que sustentan cada decisión.

3.2.2.1. Eliminación de errores en la extracción

Se realizó una primera depuración de los datos eliminando las ocurrencias que no dan cuenta de la unidad de análisis. Es importante tener presente algunos errores de extracción que corresponden a formas en idioma inglés como *given all, however promising* y otras formas con errores tipográficos o de edición como *holandésel programa, niñoshay 19 sentencias condenatorias, gratisramiro Velásquez Gómez*.

3.2.2.2. Eliminación de pronombres

LINGÜÍSTICA Y LITERATURA
ISSN 0120-5587
E-ISSN 2422-3174
N. º 78, 2020 11-34

Se eliminaron las ocurrencias correspondientes a los pronombres de tipo personal, indefinido, relativos, reflexivos e interrogativos, puesto que no aportan información susceptible de ser analizada con relación a la unidad de análisis, al no poder recuperar el elemento al que se refieren. Algunos ejemplos de estas ocurrencias son: *imaginar usted, pensar yo mismo, corregir algunos, capacitar él.*

3.2.2.3. Eliminación de formas deícticas

Al igual que con las formas pronominales, los elementos que dan cuenta de deíxis también se eliminaron, como se evidencia en: *acercar los presentes, priorizar muchas de sus acciones, aclarar Gómez.*

3.2.2.4. Eliminación de formas impersonales o incompletas

Las formas impersonales o incompletas se eliminaron puesto que no aportan ningún tipo de información relevante que permita establecer su función en cuanto a la expresión de la opinión, como se evidencia en: *hacer algo, soñar uno*.

3.2.2.5. Eliminación de ocurrencias repetidas

En algunos casos se encontraron varias repeticiones de la misma ocurrencia como el caso de *abrir la puerta* (3 veces), *acabar el negocio* (2 veces), *acabar una década* (2 veces), *accionar los grupos armados* (2 veces); las cuales se eliminaron dejando una sola puesto que en esta investigación no se planteó un análisis estadístico enfocado en la cantidad de ocurrencias, sino en el análisis de cada tipo.

3.2.2.6. Eliminación de cifras

Con relación a las cifras, estas no se tuvieron en cuenta para el análisis debido a que representan desafíos tanto teóricos como metodológicos que ameritan ser objeto de estudio en un

LINGÜÍSTICA Y LITERATURA
ISSN 0120-5587
E-ISSN 2422-3174
N. º 78, 2020 11-34

nuevo trabajo de investigación. Ocurrencias del tipo: *beneficiar 1 151 000 hectáreas, recibir 27 371 créditos, superar 16 mil millones de dólares* hacen parte de esta categoría.

3.2.2.7. Eliminación de nombres propios

Se omitieron las ocurrencias que contenían nombres propios dado que, en términos de polaridad, estas no la aportan, toda vez que su función es la de brindar otro tipo de información complementaria acerca del sujeto de la oración, como se evidencia en: *explicar Sanabria, expresar Emilio García Méndez, agregar Zavala, decir el presidente Santos*; acerca de lugares como en: *mantener América Latina, vivir Medellín*; u organizaciones o programas estatales: *fundar Cambio Radical, quejar Ecopetrol, recomendar el Banco Mundial*.

3.2.2.8. Eliminación de los verbos modales

Según Gili Gaya (1980), los verbos modales expresan el *modus* explícito de la oración, mientras el verbo en infinitivo que los acompaña expresa el *dictum* (p. 119). Debido al carácter auxiliar de los verbos *deber* y *poder,* encontrados en el corpus, se excluyeron en el análisis, toda vez que las características inherentes a estos contribuyen a especificar el significado del verbo que auxilian. Por lo tanto, no cumplen con la estructura de análisis establecida [verbo + objeto directo].

3.3.3. Etiquetado

Con el etiquetaje semántico, se pretendió caracterizar no solo el verbo en cuanto al tipo y el sentido, sino también identificar los temas en los cuales se enmarcan los artículos analizados. En consecuencia, se asignó al núcleo del objeto directo su área temática. Además, para definir las cargas de polaridad, se tomaron los dos componentes de la estructura por separado, el verbo y el objeto directo y se etiquetó su valor correspondiente. Finalmente, se computó la suma de las cargas y, de este modo, se obtuvo la polaridad final de la estructura analizada teniendo en cuenta la premodificación negativa de las ocurrencias, cuando se presentaba, a la hora de asignar la

LINGÜÍSTICA Y LITERATURA
ISSN 0120-5587
E-ISSN 2422-3174
N. º 78, 2020 11-34

carga global, como en el caso de: *no contar trabajo, no crear impuesto, no dar limosna, no encontrar solución.*

4. Resultados y análisis de los datos

Los datos obtenidos en este estudio corresponden a un acercamiento no solo teórico sino metodológico desde el AS que permitió, desde lo lingüístico, identificar características y generar una descripción de la unidad de estudio. Sin embargo, los análisis y conclusiones que se presentan en este apartado tienen un carácter descriptivo, por lo que es necesario extrapolar el análisis a un corpus de mayor alcance, puesto que permitiría establecer las posibles tendencias de comportamiento del fenómeno estudiado.

En primer lugar, se relaciona una descripción de los verbos analizados teniendo en cuenta su tipología, carga de polaridad y sentido. Como segundo elemento, se describe la carga y el área temática del núcleo de objeto directo. En tercer lugar, se abordan los datos hasta ahora descritos al relacionar los distintos componentes de la estructura predicativa analizada. Finalmente, se enumeran las consideraciones generales entorno al fenómeno de la pobreza.

4.1. Ocurrencias de la estructura predicativa [verbo + objeto directo]

La unidad de análisis se entiende como la estructura predicativa [verbo + objeto directo] de la cual se recuperaron automáticamente las ocurrencias aquí analizadas. A modo de ejemplo: *aumentar los niveles de pobreza, disminuir las brechas, lograr la jubilación, perder el empleo, superar la pobreza extrema.*

4.2. Ocurrencias según los periódicos analizados

Los periódicos seleccionados dan cuenta de tres posiciones políticas que permiten analizar el fenómeno de la pobreza en su amplitud política: *El Tiempo* (centro derecha), *El Espectador* (centro), *El Colombiano* (centro derecha) (Quiroz *et al*. 2017, p. 86). La Figura 1 ilustra la distribución de las ocurrencias analizadas según cada periódico:

LINGÜÍSTICA Y LITERATURA
ISSN 0120-5587
E-ISSN 2422-3174
N. º 78, 2020 11-34

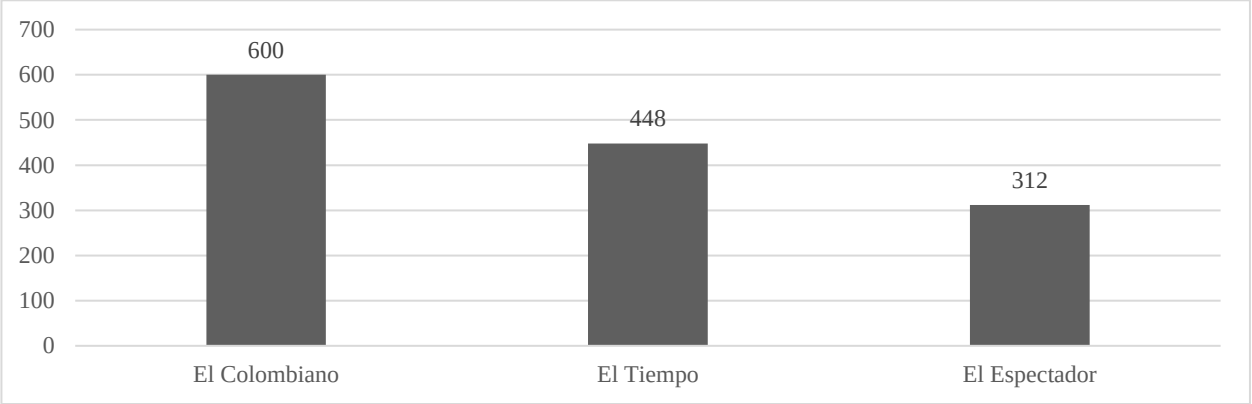

Figura 1. *Distribución de las ocurrencias por periódico*

4.3. Verbos

Concluida la Fase 4 de depuración, se consolidó la muestra de trabajo con un total de 89 verbos que conformaron las 1 360 ocurrencias analizadas (50 % de las consolidadas en la Fase 3), donde el verbo de mayor frecuencia de aparición es *hacer* (78 ocurrencias) mientras *avanzar* es el de menor frecuencia (4 ocurrencias).

4.3.1. Tipología verbal

Siguiendo la propuesta de Lorente (2002) se identificó la tipología de los 89 verbos de la muestra de trabajo lo que permitió identificar, según los criterios establecidos por la autora, que el discurso analizado es temático: 54 verbos fraseológicos, 22 verbos discursivos y 13 verbos conectores. De igual manera, se estableció la frecuencia de las ocurrencias por cada tipo de verbo como se ilustra en la Tabla 2, seguida de algunos ejemplos de verbos según su tipología:

Tipo de verbo	Frecuencia	%
Fraseológico	827	60,81
Discursivo	324	23,82
Conector	209	15,37
Total	**1 360**	**100 %**

Tabla 2. *Frecuencia de las ocurrencias con relación al tipo de verbo*

LINGÜÍSTICA Y LITERATURA
ISSN 0120-5587
E-ISSN 2422-3174
N. º 78, 2020 11-34

—Fraseológicos: *aumentar, construir, enfrentar, fortalecer, implementar, salir, trabajar*

—Conectores: *cambiar, cumplir, encontrar, llevar, dar, incluir, quedar, pedir, tomar*

—Discursivos: *explicar, hablar, indicar, llamar, mostrar, presentar, resaltar, señalar, ver*

4.3.2. Sentido verbal

Para establecer el sentido del verbo, se utilizó la clasificación ofrecida por el proyecto *WordNet* (basada en la propuesta de Levin, 1993). En los datos analizados se da cuenta de 14 de los 15 sentidos verbales, exceptuando solo aquellos con el sentido *weather*. En consecuencia, con esta información, podría identificarse la forma como se tematiza el fenómeno de la pobreza en la prensa colombiana. La Figura 2 da cuenta de los sentidos verbales hallados (se conservaron los nombres de las categorías en inglés por razones metodológicas) y la frecuencia de ocurrencias por cada uno. Se listan algunos ejemplos de verbos según los tres sentidos de mayor frecuencia con los cuales da cuenta del 50 % de las ocurrencias analizadas: brindar, comprar, ganar (*possession*); asegurar, decir, destacar (*communication*); fortalecer, incrementar, reducir (*change*).

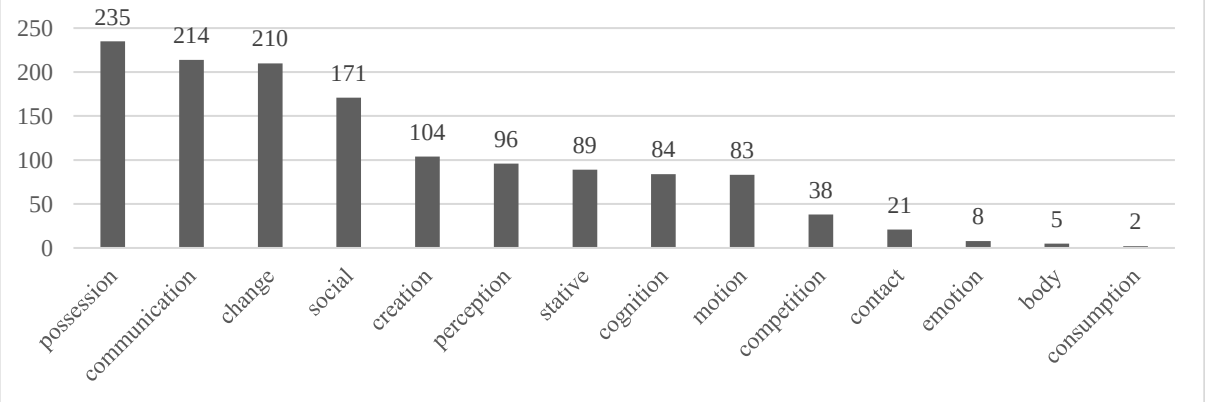

Figura 2. *Frecuencia de aparición del sentido de los verbos*

4.3.3. Carga de polaridad del verbo

LINGÜÍSTICA Y LITERATURA
ISSN 0120-5587
E-ISSN 2422-3174
N. º 78, 2020 11-34

A diferencia de acercamientos teóricos en el área del AS (véase Tan, Wang, & Cheng, 2008; Taboada *et al*. 2011; entre otros), donde la asignación de las cargas de polaridad se da solo si dan cuenta de las estrategias discursivas utilizadas en el corpus entre las categorías *positivo* o *negativo*, en la presente investigación se trabajó con las tres polaridades: *positiva*, *negativa* y *neutra*. La Tabla 3 relaciona la información obtenida según el tipo de verbo y la carga asignada.

Tipo de verbo	**Fraseológico**		**Discursivo**		**Conector**	
	Frecuencia	**%**	**Frecuencia**	**%**	**Frecuencia**	**%**
Positivo	411	49,7	0	0	89	42,58
Negativo	140	16,93	0	0	24	11,48
Neutro	276	33,37	324	100	96	45,93
Total	**827**		**324**		**209**	

Tabla 3. *Frecuencia cargas de polaridad de las ocurrencias según el tipo del verbo*

Con relación al *neutro* y teniendo en cuenta las características analizadas en la estructura predicativa [verbo + objeto directo], este se entiende *como los elementos textuales no poseedores de carácter temático ni especializado, cuya función es discursiva, toda vez que no aportan a la polaridad del texto, sino que ayudan en la construcción del discurso desde las estrategias de coherencia y cohesión* (Pérez Pérez, 2018, p. 86). Algunos ejemplos de verbos según su carga de polaridad:

—Positivos: *apoyar, aumentar, construir, generar, implementar, mejorar, producir, superar*
—Negativos: *afectar, bajar, cerrar, disminuir, perder, preocupar, reducir, sufrir*
—Neutros: *considerar, contar, decir, destacar, explicar, hablar, indicar, mostrar, presentar*

4.2. Núcleo del objeto directo

Retomando la estructura analizada y teniendo en cuenta el núcleo del objeto directo como *el sustantivo nuclear de un sintagma nominal presente en el mismo*, la asignación de las áreas temáticas correspondió a las secciones de las cuales se recuperaron los artículos de prensa, esto

LINGÜÍSTICA Y LITERATURA
ISSN 0120-5587
E-ISSN 2422-3174
N. º 78, 2020 11-34

es: *economía*, *sociedad* y *política*. En este sentido, el carácter temático del fenómeno recae, principalmente, en los núcleos de objeto directo agrupados bajo las etiquetas de *economía* (397 ocurrencias, 29,19 %), *sociedad* (279 ocurrencias, 20,51 %) y *política* (259 ocurrencias, 19,04 %). Adicionalmente, una cuarta área denominada *general* (425 ocurrencias, 31,25 %), se agruparon los núcleos sin marca temática definida, que dan cuenta de las estrategias discursivas utilizadas en el corpus.

Tradicionalmente, la pobreza ha sido vista desde un enfoque, predominantemente, económico. Sin embargo, desde la perspectiva semiótica de Pardo y Ruiz, se afirma que la representación de este fenómeno en América Latina supera el abordaje de su dimensión económica al reconocer que «la pobreza no solamente implica la carencia material y simbólica de recursos sin los cuales ningún ser humano puede desarrollar sus potencialidades, sino la forma como cada sociedad interpreta dichas carencias y las relaciones de poder que sustentan estas interpretaciones» (Pardo & Ruiz, 2017, p. 265). Además, desde lo sociológico, en Álvarez y Naharro (2017), la pobreza se aborda desde distintas visiones (estadístico-técnica, desarrollista productivista, individualista, negativa clasista y racista, médico social, piadosa, crítica) que en última instancia «naturalizan y ponen en sus víctimas las causas de la pobreza» (p. 261). Ahora bien, con la distribución de los núcleos de objeto directo según el área temática, sería factible identificar cómo el tratamiento de la pobreza, en la prensa colombiana, se da desde elementos de tipo económico y estadístico, principalmente.

4.2.1. Conformación de las áreas temáticas

Inicialmente, los núcleos de objeto directo que se agruparon bajo las cuatro categorías ya mencionadas, se etiquetaron dentro de una serie de subcategorías más amplias como se detalla a continuación:

—*Economía*: se agruparon las ocurrencias relativas al empleo, desarrollo de los sectores económicos, inversiones por parte de empresas privadas o públicas, índices con relación a los datos de pobreza, explotación o producción de bienes, oferta de servicios, explotación de recursos y demás ítems relacionados con dinero como el tema de regalías, pensiones y presupuestos.

LINGÜÍSTICA Y LITERATURA
ISSN 0120-5587
E-ISSN 2422-3174
N. º 78, 2020 11-34

—*Sociedad*: se agruparon las ocurrencias correspondientes a educación, tecnología, construcción, medicina, medio ambiente, cultura, deporte, medios de transporte, comunicación y religión.

—*Política*: se agruparon las ocurrencias correspondientes a información estatal desde los ámbitos judicial, legal y administrativo del tipo tratados, leyes, proyectos, ámbito electoral, militar y política internacional.

—*General*: se agruparon las ocurrencias cuya marca temática era nula. Gran parte de estas ocurrencias son de uso discursivo en cuanto pueden introducir o contextualizar las ocurrencias agrupadas en las tres categorías anteriores. Algunos ejemplos de núcleos de objeto directo por área temática son:

—Economía: *empleabilidad, empleo, sistema productivo, sistema tributario, desempleo*
—Sociedad: *calidad de vida, embarazo adolescente, EPS, gestión educativa, transporte*
—Política: *derecho de réplica, Estatuto Anticorrupción, ley, política regional, regla electoral*
—General: *atención, ayuda, avance, cobertura, conferencia, giro nacional, privación, temor*

4.2.2. Carga de núcleos de objeto directo con relación al área temática

La relación que se estableció entre la carga de polaridad y las áreas temáticas permitiría dar cuenta de la forma como se presenta el fenómeno de la pobreza, es decir, tomando únicamente los núcleos marcados como *positivos* y *negativos* —los cuales establecen el carácter de especialidad del fenómeno—, se afirmaría que la pobreza se aborda desde argumentos *positivos* de carácter económico, principalmente.

4.3. Estructura predicativa [verbo + objeto directo]

La polaridad total de la unidad de análisis equivale a la aplicación de 26 reglas de combinación que permiten asignar una carga final a cada una de las ocurrencias analizadas. Las etiquetas empleadas POS, NEG, NEU, corresponden a *positivo*, *negativo* y *neutro*, respectivamente. Un total de 43 ocurrencias cuentan con premodificación negativa del tipo *no* o

LINGÜÍSTICA Y LITERATURA
ISSN 0120-5587
E-ISSN 2422-3174
N. º 78, 2020 11-34

no sé, cuya función es la de conmutar la carga final de la unidad de análisis. En consecuencia, el 86 % de estas ocurrencias se marcaron con una carga global negativa mientras el 14 % restante se consideraron como positivas. Por otra parte, no se encontraron ocurrencias con esta característica que al conmutar su carga se definieron como neutras. En la Tabla 4 se relacionan cada una de estas reglas de combinación junto a su correspondiente ejemplo recuperado del corpus.

Reglas de combinación	**Ejemplos**
POS+POS=POS	garantizar educación de alta calidad
POS+POS=NEG	(no) dar subsidio
POS+NEG=POS	(no) dar limosna
POS+NEG=NEG	apoyar minería informal
POS+NEG=NEU	dar pesar
POS+NEU=POS	cumplir meta
POS+NEU=NEG	incrementar costo
POS+NEU=NEU	ofrecer información
NEG+POS=POS	(no) perder subsidio
NEG+POS=NEG	afectar empleabilidad
NEG+POS=NEU	cerrar venta de activo
NEG+NEG=POS	bajar línea de pobreza
NEG+NEG=NEG	sufrir escasez de alimento
NEG+NEG=NEU	preocupar desplazamiento
NEG+NEU=POS	bajar precio medicamento
NEG+NEU=NEG	perder participación
NEG+NEU=NEU	bajar tendencia
NEU+POS=POS	hacer inversión
NEU+POS=NEG	(no) contar servicio de gas
NEU+POS=NEU	señalar política de vivienda
NEU+NEG=POS	sacar miseria
NEU+NEG=NEG	tomar terrible decisión
NEU+NEG=NEU	medir pobreza
NEU+NEU=POS	realizar investigación
NEU+NEU=NEG	(no) pasar prueba
NEU+NEU=NEU	hacer acta

Tabla 4. *Reglas de combinación*

La Figura 3 presenta la distribución según la polaridad total de las ocurrencias analizadas, teniendo en cuenta la premodificación arriba descrita.

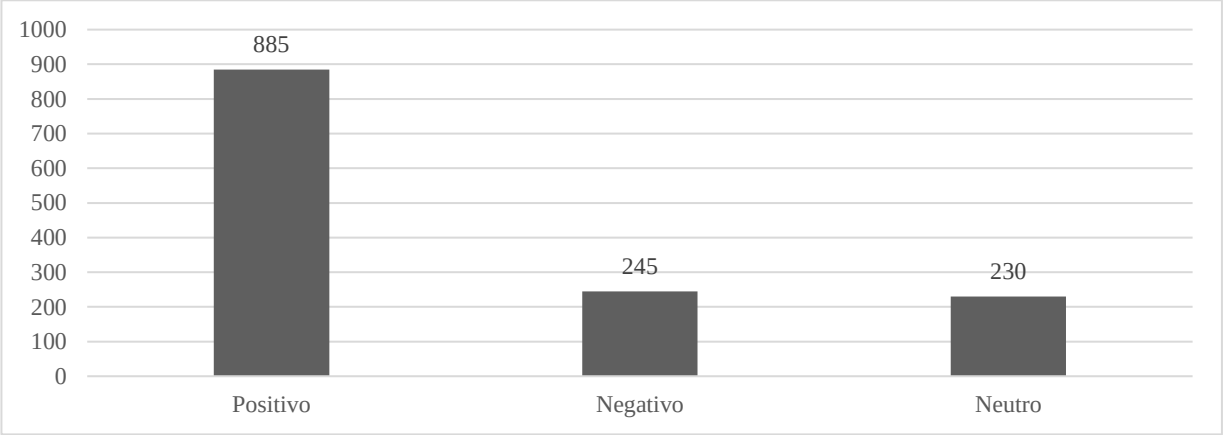

Figura 3. *Distribución polaridad total de la estructura [verbo + objeto directo]*

4.4.1. Polaridad de la estructura predicativa [verbo + objeto directo] con relación a la tipología verbal

Los verbos fraseológicos y conectores son los encargados de expresar el grado de especialidad del discurso empleado en los artículos analizados. Por consiguiente, al tomar la carga de estas ocurrencias son estas las discriminantes en polaridad y, en consecuencia, se confirma la idea que el discurso de pobreza en Colombia sería abordado positivamente en los medios analizados. De igual forma, al analizar el *neutro* se corroboró su función discursiva. En la Tabla 5, se muestra la distribución de las 890 ocurrencias con carga global positiva y su relación con la tipología verbal.

Tipo de verbo	**Frecuencia**	**%**
Fraseológico	589	66,18
Discursivo	173	19,44
Conector	128	14,38
Total	**890**	**100**

Tabla 5. *Tipología verbal según carga positiva de la estructura [verbo + objeto directo]*

4.5.1. Polaridad de la estructura predicativa [verbo + objeto directo] con relación al sentido verbal

LINGÜÍSTICA Y LITERATURA
ISSN 0120-5587
E-ISSN 2422-3174
N. º 78, 2020 11-34

El sentido del verbo da cuenta de la elección del vocabulario con relación a los criterios del Índice de Pobreza Multidimensional (IMP) y la forma como estos se desarrollan en la representación de la pobreza en los periódicos elegidos. Por consiguiente, la predominancia de ocurrencias positivas unidas a verbos de tipo *possession, change, social* (50,45 % de 890 ocurrencias) darían cuenta de un tratamiento *positivo* de la pobreza en cuanto se evidencia una mejoría según los parámetros del IMP. Entre algunos ejemplos se encuentran: *aumentar empleo permanente, cerrar brecha rural urbana, disminuir desigualdad social, disminuir brecha, incrementar recaudo, mejorar condición alimentaria, reducir pobreza extrema.*

4.6.1. Polaridad estructura predicativa [verbo + objeto directo] con relación al área temática

El área temática del objeto directo determina los ejes desde los cuales se trabaja la pobreza en los periódicos los cuales, podría afirmarse, están en directa relación con los parámetros del IMP y las Naciones Unidas tomando como fuente el documento sobre los Objetivos del Milenio (Naciones Unidas, 2015). Por consiguiente, omitiendo lo *neutro*, las áreas temáticas se listan en un mismo orden: *economía, sociedad* y *política*, lo que establecería una jerarquía temática a la hora de hablar sobre pobreza. Igualmente, teniendo en cuenta la carga de polaridad según las áreas temáticas mencionadas, este orden se conserva en los argumentos *positivos* y *negativos*.

5. A modo de cierre

A las 1 360 ocurrencias analizadas, que corresponden al 50 % de los datos extraídos conforme a la estructura predicativa [verbo + objeto directo], se les asignó la información correspondiente al tipo y sentido para los verbos y el área temática para los objetos directos. Con relación a la polaridad, se definieron tres variables discretas: *positivo* (1), *negativo* (-1) y *neutro* (0). Se etiquetó cada elemento según la carga de polaridad correspondiente y al final, se definieron las posibles combinaciones según la estructura [carga del verbo + carga del objeto directo = carga final de la estructura].

LINGÜÍSTICA Y LITERATURA
ISSN 0120-5587
E-ISSN 2422-3174
N. º 78, 2020 11-34

Frente a la tipología verbal, los verbos más frecuentes son los fraseológicos (60,67 %); seguidos de los discursivos (24,72 %); y, por último, los verbos conectores (14,61 %). Según Lorente (2002), los verbos conectores ayudan en la construcción del discurso especializado. En consecuencia, los artículos analizados en este trabajo son de carácter especializado (discurso sobre la pobreza), dado que los verbos utilizados son, en su mayoría, temáticos (75,28 % del total). Del mismo modo, sobre el sentido verbal, estos se refieren a la forma cómo se aborda la temática, puesto que permiten establecer un esquema alrededor de la descripción y evolución del tema en cuestión. El 61,03 % de los datos analizados dan cuenta de cuatro sentidos verbales: *possession* (17,28 %); *communication* (15,74 %); *change* (14,44 %) y *social* (12,57 %).

Con relación al área temática de los objetos directos, se identificó que la pobreza en Colombia sería presentada desde los aspectos económicos, sociales y políticos. Este orden se puede establecer gracias a la frecuencia del área temática: *economía* (29,19 %); *sociedad* (20,51 %) y *política* (19,04 %). Al respecto, las ocurrencias agrupadas en el área general (49,49 %) corresponden a estrategias discursivas.

Con referencia a la carga de los verbos, el fenómeno de la pobreza es descrito mediante el uso predominante de verbos *positivos*, los cuales dan cuenta del 51,69 % de los datos; mientras que los *negativos* corresponden al 34,83 % y los *neutros*, al 13,48 %. Asimismo, en cuanto a la carga de los objetos directos, estos son en su mayoría *neutros* (49,49 %). Ahora bien, lo *neutro* está más cerca de ser *positivo* que de ser *negativo* en cuanto que «words with SO = 0 were sometimes interpreted as positive, but almost never interpreted as negative» (Taboada *et al*. 2011, p. 290). En consecuencia, los complementos directos utilizados en la representación de la pobreza son *positivos* en un 80,45 %, teniendo en cuenta las 1 360 ocurrencias analizadas.

Por último, con los datos sobre la carga final de la estructura [verbo + objeto directo], se establecería que el fenómeno de la pobreza se representaría positivamente en la prensa colombiana. El 65,44 % de las ocurrencias (890) son *positivas*, mientras el 17,94 % (244) son *negativas* y el 16,62 % (226) *neutras*. Teniendo en cuenta el comportamiento del *neutro*, la representación de la pobreza en la prensa colombiana se evidencia con el 82,06 % de las ocurrencias analizadas (1 116 de 1 360).

6. Conclusiones

Los resultados descritos en el presente artículo son relevantes en cuanto constituyen un aporte a la solución de la clasificación automática de los textos a partir del AS. Desde la descripción lingüística del verbo y su objeto directo, se ha logrado presentar una guía metodológica con esta estructura predicativa como unidad de análisis, lo que ha permitido establecer que el verbo, como unidad lingüística, unido al objeto directo, sería discriminante en términos de polaridad, al igual que otras categorías lingüísticas como el adjetivo; lo que amplía el panorama de trabajo a otras unidades de análisis poco estudiadas. Además, los textos de carácter socio-político que hemos analizado, han sido poco trabajados con fines de AS, lo que también constituye un aporte en la descripción de estos para tales propósitos.

De igual forma, la propuesta —con los ajustes necesarios—, puede ser aplicada o extrapolada a otros fenómenos que se deseen investigar. Por último, a partir de los datos analizados, se presenta un acercamiento teórico al *neutro* como variable de polaridad, al ofrecer una definición, lo que constituye un aporte al área del AS, en cuanto a la construcción de una tipología del *neutro*, variable que en la literatura no ha sido abordada ampliamente.

Finalmente, se puede concluir, por un lado, que el fenómeno de la pobreza en Colombia sería representado positivamente en la prensa. Por otro lado, una adecuada clasificación de las ocurrencias de la estructura predicativa [verbo + objeto directo] permitiría no solo mejorar la clasificación de la polaridad de los textos sino también la definición de los parámetros de extracción utilizados en las herramientas de AS. Además, la propuesta de definición del *neutro*, desde su caracterización y función discursiva al interior del discurso especializado, constituye un punto de partida en la discusión en el área sobre esta carga de polaridad.

Referencias bibliográficas

1. Alcoba, S. (1999). La flexión verbal. En Bosque, I., & Demonte, V. (Eds.). *Gramática Descriptiva de la Lengua Española* (p. 4917).

2. Álvarez, S. & Naharro, N. (2017). Representaciones de la pobreza en la prensa hegemónica Argentina. En Chiquito, A., & Quiroz, G. (Eds.). *Pobreza, Lenguaje y Medios en América Latina* (pp. 235–264). Berna: Peter Lang.

3. Benamara, F., Cesarano, C., Picariello, A., Reforgiato, D., & Subrahmanian, V. (2007). Sentiment Analysis: Adjectives and Adverbs are better than Adjectives Alone. En *Proceedings of International Conference on Weblogs and Social Media*.

4. Bosque, I. (2006). *Diccionario combinatorio práctico del español contemporáneo*. Madrid: Ediciones SM.

5. Chiquito, A. & Quiroz, G. (Eds.) (2017). *Pobreza, lenguaje y medios en América Latina*. Berna: Peter Lang.

6. Chomsky, N. (1997). *What Makes Mainstream Media Mainstream*. Recuperado de http://www.zmag.org/chomsky/articles/z9710-mainstream-media.html

7. Colombia. DANE. (2012). *Pobreza en Colombia*. Bogotá D.C.: Oficina de Prensa DANE.

8. Ding, X., Liu, B., & Yu, P. S. (2008). A Holistic Lexicon-based Approach to Opinion Mining. En *Proceedings of the International Conference on Web Search and Web Data Mining - WSDM'08* (pp. 231-240).

9. Gili Gaya, S. (1980). *Curso superior de sintaxis española*. Barcelona: Vox Biblograf.

10. Hatzivassiloglou, V. & McKeown, K. R. (1997). Predicting the Semantic Orientation of Adjectives. En *Proceedings of the 35th Annual Meeting on Association for Computational Linguistics* (pp. 174–181). Association for Computational Linguistics.

11. Hatzivassiloglou, V. & Wiebe, J. M. (2000). Effects of Adjective Orientation and Gradability on Sentence Subjectivity. En *Proceedings of the 18th Conference on Computational Linguistics*, (1), (pp. 299-305). Association for Computational Linguistics.

12. Joshi, M. & Penstein-Rosé, C. (2009). Generalizing Dependency Features for Opinion Mining. En *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers* (pp. 313-316).

13. Koppel, M. & Schler, J. (2006). The Importance of Neutral Examples for Learning Sentiment. *Computational Intelligence*, *22*(2), pp. 100-109.

14. Levin, B. (1993). *English Verb Classes and Alternations*. Chicago: The University of Chicago Press.

15. Liu, B. (2012). Sentiment Analysis and Opinion Mining. En Hirst, G. (Ed.). *Synthesis Lectures on Human Language Technologies* (pp. 1-167). California: Morgan & Claypool.

16. Lorente, M. (2002). Verbos y discurso especializado. *Estudios de Lingüística del Español* [En línea]. Recuperado de http://elies.rediris.es/elies16/Lorente.html

17. Mullen, T., & Collier, N. (2004). Sentiment Analysis Using Support Vector Machines with Diverse Information Sources. En *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing* (4), (pp. 412-418).

18. Nasukawa, T., & Yi, J. (2003). Sentiment Analysis: Capturing Favorability Using Natural Language Processing. En *Proceedings of the 2nd international conference on Knowledge capture* (pp. 70-77).

19. Naciones Unidas. (2015). Objetivos de Desarrollo del Milenio. Nuevas Ediciones S.A. Recuperado de http://www.un.org/es/millenniumgoals/pdf/PR_Global_MDG09_SP.pdf

20. Ng, V., Dasgupta, S., & Arifin, S. (2006). Examining the Role of Linguistic Knowledge Sources in the Automatic Identification and Classification of Reviews. En *Proceedings of the COLING/ACL on Main Conference Poster Sessions* (pp. 611-618).

21. Pardo, N. & Ruiz, J. (2017). Pobreza y bienestar en Colombia, construcción de referenciales en perspectiva mediática. En Chiquito, A. & Quiroz, G. (Eds.). *Pobreza, lenguaje y medios en América Latina* (pp. 265-302). Berna: Peter Lang.

22. Pérez Pérez, C. M. (2018). *Metodología para la caracterización de patrones [verbo + objeto directo] sobre pobreza en la prensa colombiana para propósitos de análisis de sentimientos* (tesis de maestría). Universidad de Antioquia, Medellín, Colombia.

23. Quiroz, G., Tamayo, A., & Zuluaga, J. (2017). Cuestiones metodológicas y técnicas en la recolección de un corpus de prensa con la palabra «pobreza». En Chiquito, A. & Quiroz, G. (Eds.). *Pobreza, lenguaje y medios en América Latina* (pp. 21-41). Berna: Peter Lang.

24. Quiroz, G., Chiquito, A., & Zuluaga, J. (2017). Cómo se representa lingüísticamente la pobreza en la prensa colombiana. En Chiquito, A. & Quiroz, G. (Eds.). *Pobreza, lenguaje y medios en América Latina* (pp. 83-107). Berna: Peter Lang.

25. Real Academia Española. (2017). *Diccionario de la lengua española*. Madrid: Espasa.

26. Riloff, E., & Wiebe, J. (2003). Learning Extraction Patterns for Subjective Expressions. En *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing* (pp. 105-112). Association for Computational Linguistics.

27. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based Methods for Sentiment Analysis. *Computational Linguistics*, *37*(2), 267-307.

28. Tamayo, A. (2018). *Metodología para el análisis automático de sentimientos en documentos a partir de técnicas de aprendizaje automático y lingüística computacional* (Tesis de maestría). Universidad de Antioquia, Medellín, Colombia.

29. Tan, S., Wang, Y. & Cheng, X. (2008). Combining Learn-Based and Lexicon-Based Techniques for Sentiment Detection without Using Labeled Examples. SIGIR 2008.

30. Turney, P. D. (2002). Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. En *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics* (pp. 417-424). Association for Computational Linguistics.

31. Whitelaw, C., Garg, N., & Argamon, S. (2005). Using Appraisal Groups for Sentiment Analysis. En *Proceedings of the 14th ACM International Conference on Information and Knowledge Management* (pp. 625-631). ACM.