

Enriquecimiento automático de un léxico afectivo basado en relaciones semánticas obtenidas de un diccionario explicativo en español

Noé Alejandro Castro-Sánchez y Bernardo López-Santiago

Centro Nacional de Investigación y Desarrollo Tecnológico (CENIDET), Cuernavaca, Morelos, México

{ncastro, bernals}@cenidet.edu.mx

Resumen. Las relaciones semánticas de sinonimia e hiperonimia son estudiadas y utilizadas en varios problemas de PLN. En este trabajo se estudian y procesan para incrementar las palabras en un léxico afectivo, con base en la determinación de patrones que utiliza el diccionario de la Real Academia de la Lengua Española en las definiciones de sustantivos y verbos. La definición de los diccionarios tiene patrones que permiten, a partir de heurísticas, extraer relaciones de sinonimia e inclusión. En este trabajo se desarrollaron dichas heurísticas y se comprobó que obtienen los sinónimos e hiperónimos de las definiciones a un bajo costo computacional, logrando agregar poco más de 1300 palabras al léxico.

Palabras clave: Relaciones semánticas, sinonimia, hiperonimia, definiciones lexicográficas, léxico afectivo.

1. Introducción

Las relaciones semánticas se han estudiado y utilizado ampliamente en diversas tareas de Procesamiento de Lenguaje Natural (PLN). Podemos dividir el estudio de estas relaciones según la manera en que se presenta el texto en los documentos procesados, esto es, como estructurado (por ejemplo diccionarios) y no estructurado (por ejemplo páginas web).

Las ventajas de utilizar texto estructurado es que se requiere una cantidad menor de documentos en comparación con el texto no estructurado, lo que repercute que su procesamiento tenga un bajo costo computacional. El hecho de que esté estructurado, significa que presenta una regularidad estructural y sintáctica, que permite a los métodos basados en reglas obtener buenos resultados.

El descubrimiento de hiperónimos con texto no estructurado generalmente necesita de un gran corpus (hablamos de millones de documentos). En el idioma inglés se han

realizado trabajos para descubrir hiperónimos utilizando los llamados “Patrones de Hearts” [1]. Por otro lado, el descubrimiento de sinónimos en este tipo de texto no es una tarea trivial, ya que las relaciones de sinonimia se establecen con mayor frecuencia por la semántica y no por la sintaxis. Se opta por un enfoque distributivo para el descubrimiento de sinónimos en texto no estructurado [2].

En este artículo nos centraremos en las definiciones utilizadas en el Diccionario de la Real Academia Española (DRAE) basado en el análisis de la estructura de palabras dentro de la categoría de verbos y sustantivos, esto con el fin de encontrar patrones que nos faciliten la identificación de relaciones semánticas de sinonimia e inclusión y así poder realizar el enriquecimiento de un léxico afectivo.

2. Léxico afectivo utilizado

Los indicadores más importantes de sentimientos son *sentiment words*, también llamadas *opinion words*. Estas palabras se utilizan comúnmente para expresar sentimientos positivos o negativos. Por ejemplo: bueno (*good*), maravilloso (*wonderful*) y asombroso (*amazing*) son palabras que expresan sentimientos positivos, y malo (*bad*), escaso (*poor*) y terrible (*terrible*) son palabras que expresan sentimientos negativos. Una lista de esas palabras es llamada léxico afectivo [3].

Un léxico afectivo puede estar dividido en ciertas clases de sentimientos, por ejemplo: positivo/negativo o en una única lista de palabras asociadas a un valor numérico que representa su polaridad [4].

En este trabajo se procesa el léxico afectivo en español (llamado LAfE) [5] el cual se realizó a partir de una traducción manual de recursos como léxicos afectivos ya existentes en el idioma inglés (General Inquirer [6], Opinion Finder [7], WordNetAffect [8]) y de teorías psicológicas (Geneva Emotion Wheel [9], Geneva Affect Label Coder [10], A Circumplex Model of Affect [11], Structure of Emotions [12]).

En su fase de creación, el léxico afectivo se enriqueció de manera manual a través de relaciones de sinonimia, inclusión y familia léxica de las palabras que se localizaron en los léxicos en inglés.

El léxico afectivo se compone de un total de 3,325 palabras, de las cuales 1,178 son sustantivos, 1,275 verbos, 820 adjetivos y 50 adverbios.

Las palabras contenidas en el léxico se encuentran etiquetadas en cuatro dimensiones de polaridad: positiva, muy positiva, negativa y muy negativa. En algunas teorías psicológicas y en WordNetAffect, algunas palabras se encuentran asociadas a emociones, en esos casos se agregó la palabra que designa la emoción asociada. Se registró el recurso de dónde se extrajo cada palabra y las palabras en inglés de las que provienen.

En la Tabla 1 se muestra un extracto del contenido de dicho léxico. Los léxicos afectivos se indican como: GI que referencia a *General Inquirer*, WNA a *WordNetAffect* y OF a *Opinion Finder*. Por otro lado, las teorías psicológicas se indican con el nombre de sus autores: Morgan-Heise, Rusell y Scherer. La simbología para representar la polaridad es la siguiente: ++ hace referencia a una polaridad muy

positiva, + indica polaridad positiva, - se asocia a una polaridad negativa y finalmente – hace referencia a muy negativa.

Table 1. Estructura del léxico afectivo en español.

Palabra	Polaridad	Emoción	Morgan- Heise	Rusell	Scherer	GI	WNA	OF	inglés
Regocijo	+	Alegría				+			rejoice
Angustiado	--	Ansiedad	--						anguish
Tranquilo	+	Serenidad		+					calm
Deplorable	-	Tristeza					-		deplorable
Despojado	-	Tristeza					-		bereft
Abatido	--	Tristeza			--				dejected
Furioso	--	Ira	--						furious
Furiosamente	--	Ira							
Colérico	--	Ira							
Solidaridad	+	Ninguna						+	solidarity

3. Definiciones en el diccionario de la RAE

3.1. Estructura de los artículos del diccionario

Las secciones textuales dispuestas ordenadamente en un diccionario se denominan artículos, y están conformadas por una entrada también denominada unidad léxica, y la información que la define o describe. Además de estos dos elementos, se ha llegado también a considerar la categoría gramatical de la entrada como parte del artículo.

Las entradas pueden ser simples (una sola palabra) o complejas (más de una palabra), y aparecen ordenadas alfabéticamente en el diccionario en su forma lematizada.

Delante de la unidad léxica se disponen información relativa de ella, de la cual puede distinguirse una serie de elementos que señalan sus restricciones y condiciones de uso, y la información semántica, o definición, que constituye el contenido básico del artículo lexicográfico.

3.2. Tipos de definiciones en el diccionario

El principio para que una definición se considere correcta es que ésta debe abarcar todo lo definido pero nada más que lo definido, y una forma de comprobar si se cumple esta condición es aplicando la prueba de la *sustituibilidad* [13].

Se puede considerar de manera muy general una distinción entre dos tipos de definiciones: de contenido y de signo. En las definiciones de contenido o conceptuales, se considera el definido como una unidad que hace referencia a la realidad y se pretende traducir en otras palabras de la misma lengua su contenido significativo. En las definiciones de signo o funcionales, se considera al definido

como elemento o signo del sistema de la lengua, por lo tanto se informa de sus valores y funciones dentro del sistema.

La definición conceptual se divide en: definición perifrástica y definición sinonímica.

En la definición sinonímica se define una palabra en términos de otra que tiene el mismo (o cercano) significado que la que se desea definir [14,15].

La definición conceptual perifrástica puede manifestarse bajo diversas formas teniendo en cuenta la naturaleza de la unidad a definir. En [13] se menciona el siguiente esquema con los diferentes tipos de definición:

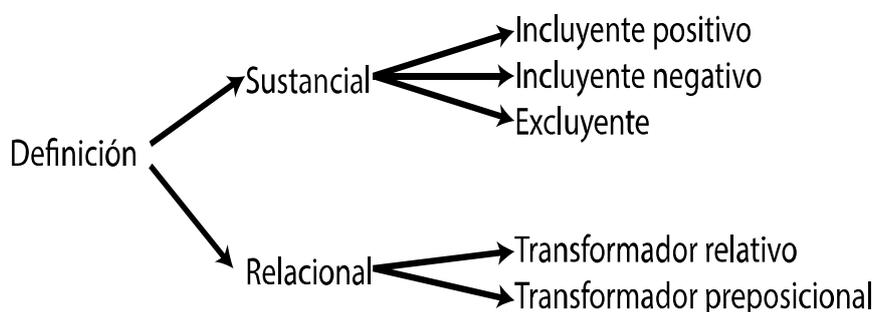


Fig. 1. Tipos de definición conceptual perifrástica.

La definición sustancial intenta responder a la pregunta « ¿Qué es el definido? ». La respuesta suele darse de las siguientes formas:

- el definido es «tal cosa» (incluyente positiva),
- el definido es «no tal cosa» (incluyente negativa),
- el definido «es contrario» o «carece de tal cosa» (excluyente).

La definición incluyente positiva viene a ser el prototipo de definición lógica aristotélica que analiza al definido mediante el género próximo (palabra cuya carga semántica abarca al definido, también se le conoce como hiperónimo) y la diferencia específica (encargada de concretar el significado del definido) [13][15].

La definición relacional remite a la relación entre el definido cualificante y otra palabra cualificada. Consta de dos partes: un elemento transformador, que puede ser un relativo (definición relativa) o una preposición (definición preposicional), el cual confiere a la palabra u oración que sirve de definidor (elemento traspuesto) el carácter de adjetivo o adverbio [13].

Dejando los modelos formales de definición cabe destacar la llamada definición morfo-semántica [13]. La definición morfo-semántica es utilizada en la definición de palabras compuestas y derivadas, la definición viene a resultar en la descomposición de los elementos componentes. Relacionadas con este grupo se encuentran los sustantivos derivados formados a partir de un verbo con la intervención de distintos sufijos nominales o de derivados adjetivos a partir de sustantivos.

4. Metodología de solución

4.1. Etiquetas de vigencia de uso y voces técnicas en el diccionario

Como primer filtro para encontrar las relaciones de sinonimia se discriminaron las entradas con las etiquetas de vigencia de uso (ant., desus., p. us.) y voces técnicas [16]. La marcación sobre la vigencia de la palabra aparece cuando se pretende informar sobre su bajo empleo [13][17].

Analizando las muestras notamos que las palabras con etiquetas de vigencia ya no se utilizan de la misma forma en el español actual y que comúnmente tienen un significado diferente, por ejemplo en la palabra “**enojo**” en su acepción tercera nos remite a la definición de “**agravio**” y más particularmente a la acepción de ofensa, que actualmente en ningún contexto se utilizan como sinónimos, lo mismo sucede con definiciones con la marca “desus.” y “p.us.”, como se muestra a continuación:

Enojo: 3. m. ant. agravio (//ofensa).

Calma: 6. f. desus. Angustia, pena.

Aprobación: 2. f. desus. prueba.

Agrado: 3. m. Ec. p. us. obsequio (// regalo).

También se observó que las palabras marcadas con voces técnicas no son relevantes para el estudio ya que no mostraban rasgos de afectividad. Por ejemplo:

Idiocia: 1. f. Med. Trastorno caracterizado por una deficiencia muy profunda de las facultades mentales, congénita o adquirida en las primeras edades de la vida.

Fuego: 11. m. Med. cauterio.

Deducción: 3. f. Fil. Método por el cual se procede lógicamente de lo universal a lo particular.

4. f. Mús. Serie de notas que ascienden o descienden diatónicamente o de tono en tono sucesivos.

4.2. Definiciones morfo-demánticas

Después de este filtrado se buscaron patrones en las definiciones llegando a los siguientes resultados: En el trabajo se identificaron sustantivos que para definirse hacen uso de la definición morfo-semántica utilizando la paráfrasis “*Acción y efecto de + verbo*” y “*Acción de + verbo*”. Las palabras obtenidas en estas definiciones pertenecen a la familia léxica del sustantivo. Por ejemplo:

Desmoralización: 1. acción y efecto de **desmoralizar**.

Quebranto: 1. acción y efecto de **quebrantar** o **quebrantarse**.

Arresto: 1. acción de **arrestar**.

Perdición: 1. acción de **perder** o **perderse**.

Dentro de la misma categoría de definiciones morfo-semánticas, en la categoría de sustantivos se localizaron definiciones del tipo “*calidad de*” y la palabra que la complementa puede ser un adjetivo, sustantivo o verbo. Estas palabras también forman parte de la familia léxica del sustantivo definido. Por ejemplo:

Fealdad: 1. calidad de feo.
Desigualdad: 1. calidad de desigual.
Ternura: 1. calidad de tierno.
Caballerosidad: 1. calidad de caballero.

4.3. Definiciones sinonímicas

Los patrones encontrados en la definición de tipo sinonímica (tanto para sustantivos como verbos) son los siguientes:

- Sustantivos separados por el signo coma (,). Por ejemplo:
Quebranto: Lástima, conmiseración, piedad.
Vacilación: Perplejidad, irresolución.
Equilibrio: contrapeso, contrarresto, armonía entre cosas diversas.
- Sustantivos separados por una disyunción, por ejemplo:
Flojera: Debilidad o cansancio.
Remedio: Enmienda o corrección.
- Sustantivos separados por comas y una disyunción, por ejemplo:
Declinación: Caída, descenso o declive.
Hiel: Amargura, aspereza o desabrimiento.

4.4. Definiciones de la lógica aristotélica (relaciones de tipo hipónimo-hiperónimo)

En [13] el autor hace distinción de los tipos de definiciones que se utilizan en el diccionario e identifica las más utilizadas para definir las palabras que caen en la categoría gramatical de sustantivos, así argumenta que los sustantivos utilizan comúnmente la definición conceptual perifrástica de tipo sustancial en cualquier modalidad. Es muy común que en este tipo de definiciones el hiperónimo de la palabra definida sea la primera palabra de la definición. Este patrón se observó en muchas de las definiciones de los sustantivos, por ejemplo en la definición de “llanto”, podemos notar que su hiperónimo es “efusión”, el llanto es un tipo de efusión. Algunos ejemplos son:

Efusión: Derramamiento de un líquido, y más comúnmente de la sangre.
Llanto: Efusión de lágrimas acompañada frecuentemente de lamentos y sollozos.

Aversión: Rechazo o repugnancia frente a alguien o algo
Horror: Aversión profunda hacia alguien o algo.

5. Resultados

Extrayendo las palabras de las definiciones morfo-semánticas se pudieron agregar 125 palabras nuevas al léxico afectivo, dichas palabras pertenecen a la familia léxica de los sustantivos extraídos del mismo léxico procesado.

Para agregar las nuevas palabras encontradas por sinonimia se verificó que al menos una palabra de la acepción de la definición procesada estuviera en el léxico afectivo, esto con el fin de garantizar que la palabra encontrada sea una palabra afectiva. Después de esta verificación se agregaron los sinónimos que aún no se encontraban en el léxico afectivo.

El método para la extracción de hiperónimos tuvo una precisión del 76%, la validación se realizó extrayendo 200 palabras de manera aleatoria de la lista de posibles hiperónimos obtenidos y buscando su definición en el diccionario para comprobar que el hiperónimo es correcto. Para determinar si es correcto el hiperónimo la definición del mismo debe abarcar al definido. Los hiperónimos nuevos encontrados para sustantivos y que se agregarían al léxico afectivo suman la cantidad de 749 palabras.

Las palabras agregadas al léxico conservaron la polaridad de la palabra que se utilizó en el léxico para extraer sus relaciones semánticas, es decir si utilizamos una palabra positiva, sus sinónimos, hiperónimos y familia léxica conservan esa polaridad positiva.

En la siguiente tabla se resumen los resultados obtenidos:

Table 2. Palabras nuevas agregadas al léxico afectivo.

<i>Relación semántica</i>	<i>Palabras agregadas al léxico de manera automática</i>
Sinonimia de sustantivos	167
Sinonimia de verbos	357
Hiperonimia	749
Familia léxica	125

5. Conclusiones

Podemos determinar parte de la familia léxica de los sustantivos y verbos de una forma relativamente fácil. La forma estructurada de las definiciones del diccionario hace idóneo el uso de patrones. Se obtienen buenos resultados en la identificación de sinonimia con un procesamiento relativamente sencillo y rápido, en comparación con los métodos basados en corpus donde además es necesaria una gran cantidad de documentos para obtener una precisión aceptable.

La polaridad en los hiperónimos no siempre se mantiene como habíamos supuesto en un principio: esto se debe a que nos encontramos hiperónimos que no tienen un sentido afectivo, es decir son entes abstractos, por ejemplo en la palabra “compasión” se tiene como hiperónimo a la unidad léxica “sentimiento” al cuál no se le puede

asignar alguna polaridad. Algunos de los hiperónimos que se encontraron con esta falta de sentido afectivo, son “acto”, “estado”, “cosa”, “persona”, “dispositivo”, “sentido”, “capacidad” y “sentimiento”.

Además en la búsqueda de sinónimos nos encontramos con que algunas palabras tienen ambas polaridades y esto se debe a que en una acepción o bajo determinado contexto tienen una polaridad positiva, pero en otra acepción o contexto tienen una polaridad negativa. Por ejemplo la palabra “atreimiento” es sinónimo de las siguientes palabras existentes en el léxico afectivo: insolencia (polaridad negativa), descaro (polaridad negativa), audacia (polaridad muy positiva) y osadía (polaridad muy positiva). Las palabras que presentaron esta característica (16 palabras), no fueron agregadas al léxico.

Bibliografía

1. Ritter, A., Soderland, S., & Etzioni, O.: What Is This, Anyway: Automatic Hypernym Discovery. AAAI Spring Symposium: Learning by Reading and Learning to Read, 88–93 (2009)
2. Wang, T., & Hirst, G.: Exploring patterns in dictionary definitions for synonym extraction. *Natural Language Engineering*, 18(03), 313–342 (2012)
3. Liu, B.: *Sentiment Analysis and Opinion Mining*. (G. Hirst, Ed.) Morgan & Claypool Publishers (2012)
4. Pak, A.: *Automatic, adaptive and applicative Sentiment Analysis*. Paris: Université Paris SUD. A thesis submitted in fulfillment of the requirements for the degree of Philosophy Doctor in Computer Science (2012)
5. Baca Gómez, R. Y.: *Desarrollo de un servicio web para determinar la polaridad de textos de redes sociales en español*. Cuernavaca, Morelos, Mexico: CENIDET (2014)
6. Stone, P., Dunphy, D., Smith, M., & Ogilvie, D.: *The General Inquirer: A Computer Approach to Content Analysis*. MIT Press (1966)
7. Wilson, T., Hoffman, P., Somasundaran, S., Kessler, J., Wiebe, J., Choi, Y., et al.: *OpinionFinder: A system for subjectivity analysis*. In *Proceedings of HLT/EMNLP Interactive Demonstrations* (2005)
8. Strapparava, C., & Valitutti, A.: *WordNet-Affect: an Affective Extension of WordNet*. In *Proceedings of the 4th International Conference on Language Resources and Evaluation*, 1083–1086 (2004)
9. Sacharin, V., Schlegel, K., & Scherer, K. R.: *Geneva Emotion Wheel rating study (Report)*. Geneva, Switzerland: University of Geneva. Swiss Center for Affective Sciences (2012)
10. Scherer, K. R.: What are emotions? And how can they be measured? *Social science information*, 44(4), 695–729 (2005)
11. Russell, J.: A Circumplex Model of Affect. *Journal of Personality and Social Psychology*, 39(6), 1161–1178 (1980)
12. Morgan, R., & Heise, D.: Structure of Emotions. *Social Psychology Quarterly*, 51(1), 19–31 (1988)
13. Pérez Lagos, Manuel Fernando: *Sobre algunos aspectos del quehacer lexicográfico*. ELUA. *Estudios de Lingüística*. Vol. 12, pp. 163–179 (1998)

14. Wilson, T., Hoffman, P., Somasundaran, S., Kessler, J., Wiebe, J., Choi, Y., et al.: OpinionFinder: A system for subjectivity analysis. In Proceedings of HLT/EMNLP Interactive Demonstrations (2005)
15. Real Academia Española, «RAE» [En línea]. Available: <http://www.rae.es/publicaciones/62-definiciones>. [Último acceso: 20 Octubre 2014]
16. Real Academia Española, «RAE» [En línea]. Available: http://www.rae.es/diccionario-de-la-lengua-espanola/que-contiene/item-numero-2#_Toc85519269. [Último acceso: 21 Octubre 2014]
17. Real Academia Española, «RAE» [En línea]. Available: http://www.rae.es/diccionario-de-la-lengua-espanola/que-contiene/item-numero-2#_Toc85519266. [Último acceso: 21 Octubre 2014]