



Universidad de Valladolid

Grado en Español: Lengua y Literatura

TRABAJO DE FIN DE GRADO
CONVOCATORIA EXTRAORDINARIA 2019

**ANÁLISIS DE EMOCIONES Y SENTIMIENTOS
EN *EL CONDE LUCANOR***

Marina Andrea HERNÁNDEZ SÁNCHEZ

TUTOR ACADÉMICO: José Manuel FRADEJAS RUEDA

ÍNDICE

1. Introducción.....	2
2. Análisis de Sentimientos (AS).....	4
2.1. Procesamiento Natural del Lenguaje (PNL).....	4
2.2. ¿Qué es el Análisis de Sentimientos?.....	5
2.3. Tratamiento de emociones en el Análisis de Sentimientos (AS).....	8
2.4. Aplicaciones del Análisis de Sentimientos (herramientas y otros trabajos).....	10
3. Análisis de Sentimientos (AS) aplicado a la literatura.....	15
4. Análisis de emociones y sentimientos en <i>El Conde Lucanor</i>	21
4.1 ¿Cuáles son los principales problemas que presenta este análisis?	24
4.1.1 Negación.....	24
4.1.2 Semántica.....	25
4.1.3 Verbos.....	27
4.1.4 Otros problemas.....	28
5. Lexicón.....	31
6. Conclusión.....	40
7. Bibliografía.....	42

1. Introducción

El presente Trabajo de Fin de Grado presenta una finalidad clara, la creación de un diccionario de sentimientos de la lengua medieval, teniendo como base del mismo, una obra tan conocida como es *El Conde Lucanor*. En primer lugar, es necesario aclarar que esta variedad de diccionarios se crea a partir de la adjudicación de una determinada polaridad (bien positiva, bien negativa o bien neutra) a los adjetivos, verbos y sustantivos (principalmente) que componen un texto según su significado.

Tal y como veremos en estas páginas, el número de trabajos realizados a partir de estos parámetros crece a medida que pasa el tiempo, pues el nacimiento de esta forma de analizar se remonta al siglo pasado. La globalización en aquél momento era reducida. Por este motivo, la mayoría de información y estudios publicados se encuentra en inglés y sobre el inglés. Aunque bien es cierto que hoy en día podemos encontrar también en español.

Pero, como bien hemos mencionado en el párrafo anterior, todos los trabajos realizados contaban con una base ya existente. Esta, corresponde con los diccionarios de sentimientos o *sentiment lexicon* en inglés, los cuales se forman gracias a un análisis previo, por parte del ser humano, de emociones y sentimientos. La disciplina en la que podemos situar todo lo explicado anteriormente es la del *Análisis de Sentimientos (AS)*.

Hecha esta aclaración, podemos afirmar que la principal diferencia con los trabajos que mencionaremos y el nuestro, es que aquí presentamos la creación de un diccionario de una lengua medieval que será útil para todo aquel que quiera analizar cualquier texto castellano medieval bajo estos presupuestos

Pero antes de llegar a este punto, haremos un repaso de lo que es el *Análisis de Sentimientos* desde lo más general a lo más concreto que, en este caso, corresponde con el diccionario en cuestión. Para ello, en el primer apartado de este trabajo, haremos una breve introducción y explicaremos la relación existente entre el *Lenguaje Natural (LN)* y esta disciplina. A continuación, se hará una breve alusión a su nacimiento para así poder sumergirnos en los problemas, tanto a nivel de nomenclatura como a nivel de definición, que ha sufrido.

Después, se verá todo lo referido al tratamiento de las emociones desde el punto de vista del *Análisis de Sentimientos (AS)* y mencionaremos algunas de las aplicaciones que consideramos más relevantes tanto en inglés como en español. Cerraremos este apartado aludiendo a algunos trabajos que, desde nuestro punto de vista, son relevantes.

En el segundo apartado, como no podía ser de otra manera, veremos cómo este tipo de análisis se puede aplicar a los siguientes aspectos de una obra literaria: 1) el argumento, 2) el título y 3) el inicio. Mencionaremos algunas de las obras consideradas de gran importancia que se han visto analizadas bajo estos parámetros. Concluiremos este capítulo relacionando el fenómeno del *bestseller* con el *Análisis de Sentimientos (AS)*.

En el tercer apartado, nos centraremos en el objeto de estudio en sí, el diccionario de sentimientos en lengua medieval. Veremos cómo lo hemos creado y todos los problemas que han surgido a medida que se realizaba, siendo los principales la negación y la semántica.

Cerraremos con las conclusiones que hemos extraído tras la realización del trabajo y propondremos una cuestión que consideramos importante y que se ha de resolver en futuros estudios. Además, incluiremos, y como es de suponer, el diccionario que es el objeto de estudio de este trabajo.

2. Análisis de sentimientos

2.1. Procesamiento Natural del Lenguaje (PNL)

Antes de comenzar a explicar de una forma más concreta qué es el *Análisis de Sentimientos (AS)*, primero haremos una breve introducción explicando qué es el *Procesamiento Natural del Lenguaje (PNL)*.

El lenguaje, tal y como muestran Vásquez, Quispe y Huayna (2009), se puede definir desde dos puntos de vista diferentes: 1) el funcional y 2) el formal. El funcional se refiere a la mera función de expresar sentimientos y a la posibilidad de comunicarse con el resto de la sociedad y, por el contrario, el formal, trata el lenguaje desde un punto de vista semántico, sintáctico y gramatical.

En otras palabras, el lenguaje natural es el que posibilita a las personas el poder transmitir día a día sus pensamientos además de tener una evolución, como su propio nombre indica, natural (como es el caso del español, italiano, inglés...). El formal, por otro lado, es el que ha desarrollado el ser humano para fines específicos. Un ejemplo es el hecho de hablar sobre un área en concreto como puede ser la ciencia creando así lenguajes lógicos, matemáticos etc.

Pero también existen lenguajes artísticos a los que se les denomina *artlangs*. En el trabajo de Mata Garrido (2017: 12) podemos ver algunos de ellos: «Son lenguas artísticas las lenguas élficas de Tolkien, el klingon de *Star Trek*, el valyrio o el dothraki de *Juego de Tronos* [...]»

Centrándonos en lo que se refiere al lenguaje natural, es necesario recordar que corresponde con la principal característica del ser humano. ¿Por qué decimos esto? Porque como ya conocemos, el resto de seres vivos no goza de dicha capacidad. Por lo tanto, podemos afirmar que el ser humano hace uso de dicho lenguaje con la finalidad de poder expresar todo aquello que piensa, siente, le preocupa, le gusta...En resumen, el lenguaje¹ tiene una clara función, y es la expresiva (ligada a un gran componente semántico).

Esta función expresiva de la que hablamos no siempre se hace de forma oral (ámbito de la producción), también hay veces en las que podemos expresar nuestro parecer de una

¹ El lenguaje tiene dos formas de funcionar muy diferenciadas. La primera corresponde con la parte de la producción (habla y escritura) y la segunda con la de la comprensión (leer y escuchar).

forma escrita (ámbito de la comprensión). De esta manera, el lenguaje adquiere una nueva función, la de vehículo de transmisión de información.

Hoy en día, gracias a la aparición de todas las nuevas tecnologías, han surgido nuevos enfoques sobre un mismo tema. La creación de las máquinas, y más concretamente de los ordenadores, ha abierto una nueva visión al mundo sobre el funcionar de las cosas, ya que estos nos han hecho la vida mucho más fácil gracias a una *Inteligencia Artificial (IA)* que las caracteriza.

Es por eso que, con este trabajo, nos vamos a sumergir en la nueva relación que ha surgido entre el lenguaje y los ordenadores, lo que conocemos como *Procesamiento del Lenguaje Natural (PLN)*. Este nuevo ámbito, muy ligado a la Lingüística Computacional, lo que quiere conseguir es que, por ejemplo, un ordenador sea capaz de analizar sintácticamente una oración, hacer traducciones automáticas, recuperar información² etc. o también un análisis de sentimientos de un texto como veremos a lo largo de este trabajo.

Pero antes de pasar al siguiente punto, quisiera cerrar este apartado con la siguiente cita de Archer y Jockers (2016: 220) en la que hacen un breve resumen de las funciones a las que tiene que hacer frente el *Procesamiento del Lenguaje Natural (PNL)*, *Natural Language Processing (NLP)* en inglés:

The ways that computers are taught to read are manifold, and the details of how all this machine reading gets done is the subject of an entire field of academic research called Natural Language Processing (NLP)

2.2. ¿Qué es el Análisis de Sentimientos (AS)?

Gracias a toda la tecnología que nos rodea, se ha popularizado el simple hecho de dar tu *opinión* sin importar cuál sea el tema del que se habla o si eres un experto en el mismo o no. ¿Por qué decimos que se ha popularizado? Porque nos encontramos en un momento en el que no solo el autor puede darla, si no que todo aquél que tenga acceso a internet puede dejar por escrito sus pensamientos en diversas plataformas como pueden

² Hoy en día contamos con asistentes electrónicos como pueden ser Alexa de *Amazon*, Siri de *Apple* o Cortana de *Microsoft*. Estos son capaces de saber lo que se les está pidiendo descodificando la voz del humano a través de un lenguaje que ha sido etiquetado previamente.

ser blogs, foros, redes sociales, etc. Y estos a su vez, pueden ser rebatidos por más usuarios.

Esta disciplina en la que nos encontramos trabajando, dentro del PNL, no siempre ha tenido la repercusión que tiene hoy en día ni tampoco ha sido tan estudiada como ahora. De hecho, es en el siglo XXI cuando ha empezado a triunfar, sobre todo entre los comerciantes, fabricantes, grandes empresas, etc. ya que supone una herramienta que juega a su favor tal y como afirman Vilares, Alonso y Gómez-Rodríguez (2012). Esto no significa que anteriormente no se haya indagado o trabajado en esta área, es más, los primeros estudios datan en el siglo pasado, más concretamente en los años ochenta.

En el momento que surgió esta cuestión todos los estudios e investigaciones que se realizaban eran únicamente en inglés, situación que dificultaba la expansión de la misma a otras lenguas, así como la posibilidad de profundizar en ella. Es inevitable recordar que cuando se inició el desarrollo de estas técnicas y métodos, el número de anglófonos a nivel mundial era mucho menor, al igual que tampoco existía la globalización actual.

La situación cambió radicalmente entrado el siglo XXI. Esto se debe a que, como ya se ha mencionado anteriormente, empezó a tener gran popularidad, lo que tuvo como principal consecuencia el despertar del interés en otros ámbitos lingüísticos, como el español.

Este interés es perfectamente demostrable ya que el número de publicaciones (artículos, revistas, tesis...) que han aparecido sobre este tema en los últimos tiempos ha aumentado de manera notable. Cabe añadir que no solo se han recogido publicaciones nuevas, sino que se ha trabajado en obras de la historia de la literatura ya escritas antes de que todo esto surgiera, aspecto que del que se hablará en el siguiente capítulo.

Por otro lado, como bien indica Murillo Lanza (2017: 25), uno de los mayores problemas que presenta esta disciplina es la nomenclatura de la misma. Por esto es normal encontrar numerosos nombres para referirse a esta ciencia, como por ejemplo:

Análisis de Sentimiento (*Sentiment Analysis*), Minería de Opinión (*Opinion Mining*), Análisis de Subjetividad (*Subjectivity Analysis*), Extracción de Opinión (*Opinion Extraction*), Análisis de Afectos (*Affect Analysis*), Análisis de Emoción (*Emotion Analysis*), etc

El mismo Murillo Lanza (2017: 25) nos despeja este caos cuando dice:

Sin embargo, de toda esta lista terminológica, los dos primeros, es decir, Minería de Opinión (*Opinion Mining*) y Análisis de Sentimiento (*Sentiment Analysis*), han sido los nombres en torno a los cuales se ha ido reorganizando la disciplina. Estos, a su vez, se han usado indistintamente.

Aunque se da por hecho que estas dos formas de aludir a una misma disciplina significan lo mismo, hay autores que no están de acuerdo y lo consideran cosas diferentes, como expresa Troyano *et al.* (2015):

El análisis de sentimientos se centra en determinar la actitud del autor de un texto con respecto a un determinado tema. La minería de opiniones, por su parte, analiza los textos a un nivel de granularidad más fino y se plantea identificar qué opina el autor del texto sobre aspectos concretos del tema sobre el que escribe (un producto, una institución, una persona, un partido político...).

En lo que se refiere al contexto español, el problema de la nomenclatura está solventado. Se prefiere, de una manera unánime, *Análisis de Sentimientos (AS)*. Sin embargo, de esta situación no puede presumir la definición de esta, ya que todavía existen matices por pulir hasta poder llegar a un acuerdo.

La explicación más extendida sobre qué es el *Análisis de Sentimientos (AS)* o *Minería de Opinión (MO)* es la proporcionada por Pang y Lee (2008): «La m[inería]o[pinión] se centra en tratar automáticamente información con opinión, lo que permite, entre otras cosas, extraer la polaridad (positiva, negativa, neutra o mixta) de un texto.»

No obstante, considero que es de recibo citar la definición que Carrillo Albornoz Cuadrado (2011: 18) propone en su tesis ya que, a mi parecer, es bastante acertada:

El término *análisis sentimental* en procesamiento del lenguaje natural hace referencia, en su más amplia acepción, a la disciplina que comprende la tarea de identificar/clasificar fragmentos de texto que contengan una carga emotiva, opinión o subjetividad.

También cabe señalar que existe la posibilidad de que se dé un paso más allá y se estudie no solo la polaridad de las palabras, sino también el grado de intensidad que suscitan cada una de ellas. Si esto ocurre, la distinción entonces será la siguiente: fuertemente negativo, negativo, neutro, positivo y fuertemente positivo (también se puede realizar con números siendo 0 neutro, 1 fuertemente positivo, -1 fuertemente negativo y, entre medias, los decimales que corresponden con el nivel de cada sentimiento).

2.3 Tratamiento de emociones en el AS

Genéticamente, el ser humano está destinado a dejarse llevar por las emociones. Tal es así, que existen ciertas expresiones faciales (al igual que corporales) que se relacionan con una emoción en concreto a niveles universales. Podemos afirmar que este tema se ha extendido hasta llegar a cualquier aspecto de la vida actual. Ha sido tal la repercusión, que incluso en los distintos tipos de religiones existentes se pueden ver reflejadas, independientemente de la importancia que esta tenga.

Si esto lo trasladamos al ámbito de la literatura, cuando un libro llega tanto al lector que le hace evadirse de todo y centrarse solamente en la historia, puede ser que el corazón se acelere si se está leyendo algo de miedo o, por el contrario, sonreír si está relacionado con el amor.

Las emociones son, al mismo tiempo, la base del *Análisis de Sentimientos (AS)* pero también uno de los mayores problemas, ya que existe una gran «falta de consenso a la hora de definir qué se entiende por emoción o sentimiento, y cómo se representa.» (Albornoz Cuadrado, 2011).

Siguiendo el modelo que Albornoz Cuadrado (2011) propone en su tesis doctoral, existen tres tareas principales a las que el AS tiene que hacer frente: 1) Detección de la subjetividad, 2) Detección de fragmentos del texto que contienen opiniones y 3) Detección o clasificación de la polaridad, siendo la última la que realmente nos compete³.

Esta labor es la que más ligada está con esta disciplina, ya que su principal cometido consiste en hacer una distinción sentimental en textos, dividiendo así las palabras en **positivas** o **negativas** según el grado emocional que estas muestren.

Existen ciertos diccionarios de emociones llamados *sentiment lexicon* (la base del AS) que son, fundamentalmente, listas de palabras catalogadas según la distinción que se acaba de mencionar. Por ejemplo, la palabra *amor* se considera positiva y por el contrario, la palabra *odio* negativa.

Pero antes de que todo esto tenga lugar, hay dos pasos previos considerados de gran relevancia; el primero es tener en cuenta que todo programa informático empieza cuando un ser humano ha hecho las anotaciones pertinentes. Una vez completado este paso, el

³ En esta misma tesis, Albornoz Cuadrado (2011) añade otras dos tareas a las tres mencionadas: la detección de la intensidad emocional y la clasificación sentimental mediante tópicos o características del texto.

ordenador tiene que saber dónde empiezan y dónde acaban las palabras para así poder analizarlas más tarde. Este procedimiento, Archer y Jockers (2016) lo llaman en su trabajo como segmentación de palabras (*tokenización*).

Esto es mucho más difícil de lo que se puede intuir, ya que como todo, presenta una serie de problemas todavía por solucionar. Por ejemplo, en inglés, las contracciones⁴ como *can't* (can not) o *Robert's* (genitivo sajón) son las que presentan complicaciones. Es más, todavía se discute si el ordenador debería considerarlas como una palabra o como dos. En el caso de que las considerara como una, entonces este debería empezar a pensar que los apóstrofes son signos de puntuación.

Otro ejemplo de contracciones de palabras que también ha dado problemas son las del tipo 'Dr.'. La razón es muy sencilla. Previamente se le ha indicado al ordenador que después de un punto hay una mayúscula y, por consiguiente, comienza una nueva frase. Pero en este tipo de casos esa norma no sirve. Es por eso que, de una manera lógica, el ordenador empezaría a contar una nueva frase a partir de ese punto (hecho que no se corresponde con la realidad).

Volviendo al tema de las emociones y haciendo una revisión cronológica de los estudios que se han realizado sobre estas, es de gran importancia mencionar el trabajo de Anscombe y Geach (1970) en el cual afirman que ya René Descartes expuso su teoría en la que defendía que existe un grupo de emociones puras y primitivas y que es a partir de estas que nacen el resto.

Fue siete años más tarde, en el trabajo de Izard (1977), cuando se propuso una agrupación de emociones divididas de la siguiente manera: *anger, contempt, disgust, distress, fear, guilt, interest, joy, shame y surprise* (ira, desprecio, asco, angustia, miedo, culpa, interés, alegría, vergüenza y sorpresa). A los dos años, Plutchik (1980) presenta una clasificación más compacta que la anterior (*anger, anticipation, disgust, joy, fear, sadness y surprise* (ira, anticipación, asco, alegría, miedo, tristeza y sorpresa)) declarando que las restantes son una derivación de estas o en su defecto, una mezcla.

Una de las distribuciones más extensas es la de Ortony *et al.* (1998), en la cual se expone un modelo OCC⁵ formado por veintidós pares antitéticos: *pride - shame, admiration -*

⁴ Una máquina no puede leer de la misma manera que un ser humano, pero puede prestar atención y sacar conclusiones de aspectos imperceptibles a nuestros ojos. Nos referimos a las comas, al por qué el autor ha decidido tal división de capítulos... así como exclamaciones o interrogaciones.

⁵ Este nombre corresponde con las siglas de los nombres de los creadores: Ortony, Clore y Collins.

reproach, happy - resentment, gloating - pity, hope - fear, joy - distress, satisfaction - fear, relief - disappointment, gratification - remorse, gratitude - anger y, por último, *love - hate*. (orgullo - vergüenza, admiración - reproche, felicidad - resentimiento, regocijo - lástima, esperanza - miedo, alegría - angustia, satisfacción – miedo, alivio - decepción, gratificación - remordimiento, gratitud – enfado y amor – odio).

Por último, hemos de aludir al trabajo de Parrot (2001), quien presenta una separación de emociones, incluso más extensa que la anterior, aunque con una diferencia, y es que este autor las subordina en forma de árbol. Es decir, en la base se encuentran las que él considera básicas (*love, joy, surprise, anger, sadness* y *fear* (amor, alegría, sorpresa, enojo, tristeza y miedo)) y en las hojas las emociones secundarias y terciarias.

Este tipo de divisiones son las que siguen la mayoría de estudiosos. Es decir, usan un lexicón de emociones en el que poder basarse para realizar el análisis. Dicho lexicón, como ya se ha mencionado, se corresponde con una lista de palabras que denotan diferentes emociones, normalmente las ocho ya mencionadas de Plutchik (1980).

2.4 Aplicaciones del AS

A medida que el tiempo ha ido pasando, el número de recursos lingüísticos⁶ que se han creado, con la finalidad de ayudar en esta disciplina, ha crecido. Estos han adquirido una gran importancia y es por lo que, bajo este epígrafe, se procederá a hacer referencia a las que, según nuestro criterio, son de mayor importancia.

Antes de comenzar, es preciso aclarar que existen tres formas generales de estudiar la polaridad en los textos estudiados desde el punto de vista del *Análisis de Sentimientos (AS)*: 1) a nivel de frase, 2) a nivel de oración o 3) a nivel de documento. Al mismo tiempo, clarificar que la mayoría de las herramientas desarrolladas están destinadas a un corpus en inglés, hecho que no impide la existencia de las mismas en el ámbito español.

⁶ A la tecnología le resulta más difícil detectar ciertas emociones en frases pequeñas y, por el contrario, encuentra más fácil analizar un texto más extenso.

Para la realización de la siguiente enumeración, hemos tomado como base la propuesta de Murillo Lanza (2017). En un primer lugar, veremos las herramientas destinadas para el inglés:

- a) *WordNet*⁷ (Miller, 1995): Corresponde con una base muy amplia de datos léxicos del inglés. Cataloga los sustantivos, verbos, adjetivos y adverbios en *synsets* (sinónimos cognitivos), expresando cada uno de ellos una concepción diferente.
- b) *WordNet Affect*⁸ (Strapparava y Valitutti, 2004): Esta herramienta se creó teniendo como base al recurso anterior. Se partió de una agrupación de palabras emocionales divididas en seis categorías diferentes según Ekman (1972) (alegría, tristeza, miedo, sorpresa, ira y disgusto) y, a partir de esto, se aplican las relaciones propuestas en *WordNet*.
- c) *SentiWordNet*⁹ (Esuli y Sebastiani, 2006): Este podría ser uno de los recursos que más nos podría interesar ya que se encarga de analizar la polaridad (positivo, negativo y neutro) de los *synsets* de *WordNet*.

Continuando así con el trabajo de Murillo Lanza (2017), pasamos a la enumeración de las herramientas destinadas al español. Anticipar que estas se pueden dividir en dos grandes grupos: corpus de opiniones y corpus de tratamiento del español.

Dentro del primer grupo se encuentran:

- a) *Spanish Movie Reviews*¹⁰: Como bien indica su propio nombre, es un corpus que se formó teniendo el cine como principal meta. Hablando en términos numéricos, es un corpus que cuenta con 3878 opiniones sobre este arte, analizadas así en una escala de intensidad que va del 1 al 5.
- b) *TASS*¹¹ y *COST*¹²: Ambos corpus se han centrado en los tweets (el primero cuenta con cuatro corpus) analizando la polaridad que presenta cada palabra que los compone.

⁷ <https://wordnet.princeton.edu/>

⁸ <http://wndomains.fbk.eu/wnaffect.html>

⁹ <http://sentiwordnet.isti.cnr.it/>

¹⁰ <http://www.lsi.us.es/~fermin/corpusCine.zip>

¹¹ <http://www.daedalus.es/TASS2015/private/general-tweets-train-tagged.xml>

¹² <http://sinai.ujaen.es/cost-2/>

- c) *COAH*¹³: Corpus formado por 1816 opiniones sobre determinados hoteles en Andalucía. Estas, al igual que el corpus mencionado en a), están analizadas bajo una escala que oscila entre el 1 y 5.

En el segundo grupo podemos mencionar las siguientes:

- a) *ML-SentiCon*¹⁴: Los creadores de este corpus lanzan una nueva propuesta sobre el algoritmo en el que se basa *SentiWordNet* pudiendo así crear una lista de *synstes* ya polarizados.
- b) *Spanish Emotion Lexicon*¹⁵: Inventario formado por 2036 palabras divididas en 6 estados de ánimo diferentes (alegría, enfado, miedo, tristeza, sorpresa y disgusto) dependiendo la emoción que susciten.
- c) *iSOL*¹⁶ e *iSOLDomainGlobal*¹⁷: Siendo la segunda una versión de la primera. El corpus ISOL corresponde con una lista de palabras en español que se nació gracias a la investigación realizada por Martínez Cámara en 2015. Este corpus está destinado al ámbito de la traducción.

En el segundo grupo se va más allá y se aplica a sectores como de los viajes, hoteles, coches, música o libros.

Como bien afirman Vilares, *et al.* (2013) en su trabajo, los españoles contamos con el sistema *Spanish SO Calculator* (Brooke, Tofiloski y Taboada, 2009), el cual surgió en la universidad *Simon Fraser* de Canadá.

Este recurso no solo almacena información sobre verbos, sustantivos, verbos y adverbios, también se encarga del estudio de los modificadores que pueden alterar la polaridad de una palabra tales como: *muy, poco, bastante, nada* etc. Además, «detecta y descarta el sentimiento reflejado en el contenido no fáctico del texto, representado, por ejemplo, mediante expresiones condicionales o subjuntivas.» (Vilares *et al.*, 2009)

Para concluir con el ámbito español, queremos hacer mención a los grupos ITALICA (Universidad de Sevilla) y SINAÍ (Universidad de Jaén), ya que estos han hecho una gran labor, en lo que al AS se refiere, en el árabe. Como era de esperar, para

¹³ <http://sinai.ujaen.es/coah/>

¹⁴ <http://www.lsi.us.es/~fermin/ML-SentiCon.zip>

¹⁵ <http://www.cic.ipn.mx/~sidorov/SEL.zip>

¹⁶ <http://sinai.ujaen.es/isol/>

¹⁷ <http://sinai.ujaen.es/esoldomainglobal/>

esta lengua también existe un recurso denominado OCA¹⁸, en el que se puede ver un corpus creado a partir de opiniones escritas en una página web especializada en cine. Contiene 500 críticas, de las cuales 250 son positivas y 250 negativas.

Aunque la principal finalidad de estas herramientas es ayudar, también existen una serie de desventajas de las que hay que ser consciente a la hora de trabajar con este tipo de elementos. No hay que olvidar que son máquinas y que muchos aspectos del lenguaje humano no los entienden, como puede ser el sentido figurado, la ironía, las metáforas (en definitiva, todo lo referido a la semántica), etc. Es decir, recursos literarios que se ven presentes en la mayoría de textos debido a que están incluidos en el lenguaje.

También, como veremos en el cuarto apartado, hay ciertos aspectos de la lengua que la tecnología no es capaz de interpretar aún. Con esta afirmación nos referimos a la negación, matiz con el que la gran mayoría de los estudiosos están de acuerdo, ya que existen varias maneras de negar; no es lo mismo decir “estoy triste” a “no estoy triste”.

Para afrontar la siguiente idea es necesario tener en mente la frase tan conocida ‘nada es perfecto’. Con esto queremos afirmar que incluso estos soportes pueden equivocarse tal y como veremos a través del ejemplo que Archer y Jockers (2016) exponen en su trabajo. Estos mencionan el caso de *Pride and Prejudice and Zombies* por una razón, y es que la máquina erró¹⁹ al considerar esta obra como un *no-bestseller* cuando en realidad sí lo era.

Aprovechando la alusión al hecho de que esta disciplina cada vez crece un poco más, me gustaría destacar un estudio bastante reciente, *Measuring historical emotions and their evolution* de Eichler *et al.* (2019) un trabajo que, desde mi punto de vista, es innovador. Estos autores se han puesto en la piel de los que se encuentran en la otra cara de la moneda en lo que se refiere a la conquista de América. Se plantean si en los textos que tratan sobre este tema se ven reflejados los sentimientos de ambas partes, conquistador y conquistado en este caso.

Para ello han tenido que tener en cuenta que el vocabulario que usaban en esa época no se corresponde con el actual (ahora es mucho más amplio y rico). Es decir, se pueden

¹⁸ <http://sinai.ujaen.es/oca-corpus/>

¹⁹ Para poder afirmar que la máquina se ha equivocado, primero hay que aclarar que esta sigue unos algoritmos a partir de los cuales decide si una obra es un *bestseller* o no. Los creadores de dichos algoritmos son Archer y Jockers (2016).

encontrar palabras que hoy en día están en desuso o incluso que el significado de alguna de ellas ha cambiado.

Pero no solo se ha hecho un análisis de polaridad a textos narrativos o que hayan marcado la historia del mundo. También, como ya se ha visto, se ha aplicado a comentarios que escriben los usuarios en las redes sociales (Troyano *et al.* 2015) y, más concretamente, en la plataforma de *Twitter* como vemos en el trabajo de Escortell, Giménez y Rosso (2016), o incluso a las críticas que ciertos consumidores han dejado en una página de críticas gastronómicas como vemos en Dubiau y Ale (2013).

Siguiendo en la línea de *internet*, también ha habido quien ha aplicado este tipo de análisis a los correos electrónicos, ayudando así a diferenciar un correo que deseado de los no deseados (lo que se conoce como *spam*). El ordenador distingue unos correos de otros dependiendo de las palabras que aparezcan en los mismos.

Por último, no querría cerrar este punto sin hacer referencia a la estilometría, ya que dentro de esta también se encuentra la *Minería de Textos (MT)*. Lo que la mayoría de la gente entiende por estilometría es «el análisis estadístico del estilo literario». Se encarga de estudiar, gracias a determinados programas estadísticos, la autoría de determinados textos (aunque en este caso, las máquinas te dirán quién **no** es el autor).

Como conclusión de este apartado decir que gracias a la facilidad de acceder a la tecnología de la que gozamos hoy en día, podemos observar una serie de datos interesantes en grandes textos como puede ser el guion de una serie, sabiendo así cuál es el personaje que más se pronuncia o el ratio²⁰ de una palabra dentro de un mismo capítulo.

²⁰ Número de veces que aparece una palabra.

3. Análisis de sentimientos aplicado a la literatura

Desde el momento en el que cada individuo se ve dotado del don de la palabra, este comienza a manifestar los primeros deseos de opinar sobre aquello que le rodea, exteriorizar sus pensamientos o, simplemente, sus sentimientos. Estos deseos de los que se habla se pueden expresar bien de una manera oral o bien dejarlos plasmados en papel de manera escrita.

El principal ámbito en el que se ve implicada la escritura es en el de la literatura. En esta arte, como en cualquiera de las seis restantes, siempre se ha podido ver una clara subjetividad por parte del autor. Es decir, el escritor crea en función de su manera de concebir el mundo, sus pensamientos, sus emociones... y con una idea clara de qué es lo que quiere dar a entender.

La *Real Academia Española* (RAE)²¹ define literatura como «arte de la expresión verbal» englobando así textos escritos, hablados o incluso cantados. Es decir, ha sido, es y será el principal canal de transmisión del ser humano. Pero si nos centramos en los textos escritos, podemos contemplar cómo desde el principio de los tiempos, estos han servido entre otros aspectos, como método de desahogo del autor.

Estos escritos se recogen en un mismo ejemplar, se editan y se venden en un determinado formato, lo que conocemos como libro. El lector compra un libro, y cuando empieza a leerlo es el momento en el que entra en juego la interacción entre el escritor y el lector.

Todo lector alguna vez ha creado algún tipo vínculo con una obra de teatro, una novela, un poema etc., y, si todo ha ido bien, le ha tenido que producir tanto llanto como risa, lágrimas, sorpresas... simplificando, algún tipo de emoción o sentimiento.

Cuando se escoge un libro se hace de manera voluntaria, pero es tarea del autor engancharte a través de sus palabras. Las emociones son el principal vehículo con el que poder crear un vínculo entre autor y lector, si esto falla, falla todo. Este vínculo del que se habla se crea a través de los personajes y, por supuesto, del narrador.

²¹ Para no entrar en las complejidades que supone una definición técnica (sea del término que sea), se ha preferido hacer uso de la definición propuesta por la RAE, tanto en este caso como en los siguientes (tal y como veremos más adelante).

A propósito de lo dicho, Bermúdez Antúnez (2010: 2) hace referencia a lo explicado en el párrafo anterior de la siguiente manera:

[...] Los mundos ficcionales son creaciones intencionales de los autores en los cuales se proponen estados de cosas ficticias. La única forma de entrar en ellos es a través del proceso de lectura. Para que este acercamiento se haga lo suficientemente duradero (lo bastante como para conocerlos) es necesario que dichos mundos estimulen al lector a permanecer inmerso en su interior o contemplación. El principal estímulo que propicia la permanencia del lector en un mundo ficcional es que éste le suscite emociones.

También hay ocasiones en las que sucede todo lo contrario. No conectamos con lo que estamos leyendo, nos aburrimos... Incluso puede llegar un punto en el que se deja de leer aunque hayamos avanzado en la obra. Esto tiene una explicación, y es que las emociones que el autor nos debería transmitir no han despertado ningún tipo de interés en el lector, es decir, no se ha sentido atraído.

Por este motivo, creo que el propio título (sea del género que sea) nos tiene que suscitar algún tipo de emoción, ya sea positiva o negativa, para así tener un mayor interés a la hora de leerlo.

Lo vamos a ver con los siguientes tres ejemplos: el primero es *Amor constante más allá de la muerte*, poema de Francisco de Quevedo en el que podemos ver la palabra *amor* y *muerte* a la vez. El segundo es *Cien años de soledad* de Gabriel García Márquez en el que vemos la palabra *soledad* o, siguiendo en la línea de este autor, *Crónica de una muerte anunciada* título en el que vemos de nuevo la palabra *muerte*.

Cabe añadir que no solo el título de una obra tiene que llamarnos la atención, también el inicio de esta es muy importante. Se han dado casos, y de hecho no es tan extraño que suceda, en los que el título no llama para nada la atención pero en cambio las primeras líneas sí, aspecto que vamos a ver ejemplificado a continuación.

El primer ejemplo es *El extranjero* y su comienzo es el siguiente: «Hoy ha muerto mamá. O quizá ayer. No lo sé. Recibí un telegrama del asilo: "Falleció su madre. Entierro mañana. Sentidas condolencias". Pero no quiere decir nada. Quizá haya sido ayer.» (Camus, 1942). Vemos cómo el autor crea la intriga perfecta en el lector.

El segundo es el comienzo de *El camino*, «Las cosas podían haber sucedido de cualquier otra manera y, sin embargo, sucedieron así» (Delibes, 1950), en el que podemos percibir cómo el autor se refiere a un hecho pasado del cual el lector no tiene constancia. Así se crea un misterio en él que le empuja a seguir leyendo.

Proponemos como último ejemplo el comienzo de *Asfixia* de Palahniuk (2001):

Si vas a leer esto, no te preocupes. Al cabo de un par de páginas ya no querrás estar aquí. Así que olvídalos. Aléjate. Lárgate mientras sigas entero. Sálvate. Seguro que hay algo mejor en la televisión. O, ya que tienes tanto tiempo libre, a lo mejor puedes hacer un cursillo nocturno. Hazte médico. Puedes hacer algo útil con tu vida. Llévate a ti mismo a cenar. Tíñete el pelo. No te vas a volver más joven. Al principio lo que se cuenta aquí te va a cabrear. Luego se volverá cada vez peor.

Este inicio ha sido seleccionado porque verdaderamente se sale de lo establecido, ya que es el propio autor el que te incita a que dejes de leer su obra. Es justamente esa acción la que crea una cierta expectación en el lector.

Todo esto, a su vez, se puede conectar con el principal deseo que tiene el ser humano al tener un libro entre sus manos, y es que busca el máximo acercamiento a la realidad que se le está contando, siendo consciente en todo momento que lo que está leyendo es ficción ligada directamente a una verosimilitud.

Para poder realizar este análisis de una manera informática, el texto debe estar en lo que se conoce como 'texto plano', es decir, libre de todos los aspectos visuales como pueden ser las negritas, cursivas, sangrados...

Las obras que se han analizado hasta el momento desde el punto de vista del *Análisis de Sentimientos (AS)*, aunque no son muchas, corresponden en su mayoría con aquellas que han dejado huella en la historia de la literatura. Entre ellas se encuentra el autor tan conocido como es William Shakespeare con *Otelo: el moro de Venecia*, *La tragedia de Hamlet, príncipe de Dinamarca*, *Romeo y Julieta* o *La fierecilla domada* entre otras (Plata Rosas, 2018).

Pero en el análisis de estas obras anteriormente mencionadas es necesario hacer una puntualización, y es que se ha hecho a partir de sentimientos en parejas. Es decir, después de haber observado las gráficas de la obra en cuestión, se han analizado cómo los sentimientos de cada personaje cambiaban en función de a quién se estuvieran dirigiendo y viceversa.

Por lo tanto, las parejas analizadas respecto al orden establecido hace dos párrafos han sido: Otelo y Desdémona, Hamlet y la reina Gertrudis (su madre), Romeo y Julieta y, por último, Petrucho y Catalina.

No solo las obras de este autor se han visto sometidas al AS, también las de Gabriel García Márquez de las cuales destacan *Cien años de soledad* o *El amor en los tiempos de cólera* así como *Orgullo y Prejuicio* de Jane Austen.

Otro género literario que también se ha estudiado bajo estas normas ha sido el de los cuentos/fábulas, como podemos ver en el trabajo realizado por Saif Mohammad (2013), el cual se realizó con el objetivo de mostrar que los cuentos tienen una distribución mucho más amplia, en cuanto a la densidad de emociones en las palabras se refiere, que las novelas.

Para la realización de dicho trabajo, Mohammad se sirvió en las ocho emociones prototípicas de Plutchik (1980), todas ellas mencionadas en el capítulo anterior. Después se dio cuenta de que los cuentos, normalmente, empiezan con una frase neutra y acaban con una feliz.

Por último, estableció una serie de similitudes entre las novelas y los cuentos. Así pudo observar que los cuentos son bastante más populares y, que es a través de las emociones, que llegan más a los lectores. Pero no solo eso, también contempló que los cuentos, notoriamente, tienen un número mucho más bajo de palabras negativas y un número mucho más alto de palabras positivas respecto a las novelas.

Otro género estudiado es la narrativa de ficción. Para autores como Patrick Colm Hogan (2003) está directamente relacionada con las emociones. Es más, llega a afirmar que ciertos géneros literarios, como por ejemplo el mencionado al principio de este párrafo, están abocados a producir determinadas emociones en los lectores.

Antes de cerrar este capítulo, es necesario mencionar un fenómeno que se ha dado en la literatura gracias a la globalización en la que nos encontramos en estos momentos, el del *best seller*, término que, según Buonocore (1976: 61), recibe la siguiente definición: «Libro de actualidad que logra obtener durante un determinado espacio de tiempo el más alto nivel de ejemplares vendidos al público».

Es decir, nos encontramos ante una clase de libros que llegan a los lectores masivamente, y que, con el tiempo, sufren una mayor demanda sobre todo entre el público femenino.

John Sutherland (2007), de una manera u otra apoya la definición anteriormente dada afirmando lo siguiente: «As a rule of thumb what defines the bestseller is bestselling.

Nothing else». Por otro lado, y de acuerdo con Archer y Jockers (2016: 14), podemos destacar la siguiente definición:

Bestsellers are a class of books that are more often dismissed as objects for amusement than studied as works of literary art, or, at the very least, as works of considerable craftsmanship [...] Beyond value in terms of millions of dollars, the value of those writers on the bestseller lists is that these books make us read. They make us imagine, feel, discuss, think and empathize [...] fantasize, spy, escape.

Observamos cómo estos dos investigadores mencionan el hecho de que un libro tiene que hacer imaginar, sentir, pensar, empatizar (el lector se entromete dentro del mundo ficcional a través de la empatía) con lo que estás leyendo...siendo esto justamente lo que llevamos defendiendo en todo este apartado.

Archer y Jockers (2016) empiezan su trabajo poniendo un ejemplo de *bestseller* con el libro titulado *The girl who kicked the Hornets' Nest* (*La reina en el palacio de las corrientes de aire* en español) del autor sueco Stieg Larsson. Este falleció antes de que el libro se publicara (es la última entrega de la trilogía). Lo que no sabía Larsson era lo que su obra iba a suceder al poco de ser publicada.

Supuso un auténtico *boom*, se expandió por absolutamente todo el mundo. Nadie entendía el porqué, nadie sabía qué tenía ese libro que había llegado a tanta gente. Incluso *Amazon*, al mes de que este libro se publicara, proclamó a Larsson como el primer escritor que había vendido más de un millón de ejemplares en *Kindle*, cifra que llegaría a los 75 millones en los siguientes dos años.

La crítica no entendía qué estaba pasando. Los ejemplares del libro se vendían tan pronto como salían de las prensas. Daba igual en qué parte del mundo estuvieras, alguien estaba leyendo este libro, hecho que no pasa muy a menudo en el mundo de la lectura. Es más, solo volvería a suceder con *Cincuenta sombras de Grey* (E. L. James, 2011).

En un cierto momento se pensó que los *bestseller* son tan aclamados porque simplemente te entretienen, son fáciles de leer y luego se olvidan. Pero Archer y Jockers están convencidos de que tiene que haber algo más.

Cincuenta sombras de Grey fue una obra atacada por tener malas habilidades a la hora de escribir, un mal estilo, inmadurez...Algunos lectores se sintieron ofendidos y otros, en cambio, les daban la razón mientras se estaban acabando la serie de libros. Hubo

a quienes que les enganchó de tal manera que incluso se podría llegar a comparar con el efecto de las drogas.

En este caso, ese 'algo más' que defienden Archer y Jockers (2016), es el tema del sexo; es eso lo que mantiene a los lectores enganchados. Podemos afirmar que este libro ha conseguido lo que todos anhelan... Según testimonios de algunos de los lectores, les ha llegado a producir incluso la sensación de tener calor en el cuerpo, es decir, este libro ha superado con creces el hecho de crear sentimientos o emociones en la otra persona.

4. Análisis de emociones y sentimientos en *El Conde Lucanor*

Gracias a la idea que tuvo Saif Mohammad (2013) de analizar los cuentos de los Hermanos Grimm tal y como propuso en su trabajo, hemos decidido seguir ese planteamiento pero con un matiz diferente, la época. El texto que será analizado pertenece a la Edad Media castellana.

Así, tal y como se anuncia en el título del presente trabajo, el texto²² al que hemos aludido en el párrafo anterior trata del Conde Lucanor de Don Juan Manuel. Esta obra y la de los hermanos Grimm, presentan un punto en común que es el que nos ha empujado a realizar este estudio. En ambos, la división del texto es igual, por fragmentos (en los Hermanos Grimm son cuentos y en el Conde Lucanor los *exemplos*).

Pero en el caso del Conde Lucanor la división es la siguiente: se trata de una obra dividida en cinco partes escritas en lengua vulgar de las cuales destaca una sucesión de cincuenta y un *exemplos* con un tono moralizante, pues es al final de cada uno donde se encuentra una moraleja de lo narrado.

La referencia a la época de este libro no es vana, es algo que hay que tener realmente en cuenta a la hora de analizar esta obra, ya que las palabras que se utilizaban y el significado de las mismas han variado considerablemente hasta el día de hoy. Por eso, ha sido lo primero que se ha tenido en cuenta para hacer el estudio de esta obra.

Otro aspecto presente debido a la relación tan estrecha que presenta con la época (contexto histórico), es el contexto social. Así, tanto la forma que tenían de ganarse la vida, como la clara división de clases, la religión o las preocupaciones típicas de dicho tiempo han sido cuestiones que hemos tenido en mente y que, además, se veían reflejadas en todo momento en el texto.

Un ejemplo muy claro es el que podemos observar a través de la palabra 'moro' que tantas veces se repite a lo largo de la obra. Para las personas de aquél momento denotaba un sentido claramente **negativo** por el simple hecho de no ser 'cristiano'. En casos como estos podemos ver la relación de contrarios que se establece, es decir, moro = negativo / cristiano = positivo.

²² He analizado la versión de José Manuel Fradejas Rueda. Véase la referencia de la misma en la bibliografía.

Aprovechando lo dicho en el párrafo anterior, es necesario repetir que el análisis de estos *ejemplos* se ha realizado bajo el punto de vista del análisis de sentimientos y, más concretamente, de la polaridad que los distintos adjetivos, verbos y sustantivos de la obra presentan.

Pero bien es cierto que, al ser esta la primera vez que alguien se decide a analizar un texto medieval desde estos parámetros, hemos decidido dejar de lado la polaridad neutra para así centrarnos en la **positiva** y **negativa**.

Por último, es de gran relevancia cerrar esta idea haciendo alusión al contexto argumental. Con esto nos referimos tanto a la historia que se nos está contando y cómo se nos está contando, como a las palabras que utiliza el autor para ello. Estas presentan una gran importancia, pues han sido seleccionadas desde el punto de vista del escritor (subjetividad).

Estas afirmaciones se relacionan también, muy estrechamente, con la polisemia. Es nuestra labor saber diferenciar a cuál de todos los significados se refiere el autor. En ocasiones ha sido tarea fácil gracias al contexto argumental pero bien es cierto que, en otras, esto se ha complicado tal y como veremos más adelante.

En nuestro caso, como ya hemos mencionado, nos hemos centrado en clasificar los adjetivos, verbos y sustantivos que presenta El Conde Lucanor en positivos y negativos según los significados que estos tenían en la Edad Media.

Cuando una novela, un cuento, o una obra literaria (independientemente del género al que pertenezca) se ha visto analizada bajo las bases de estos tipos de diccionarios, se ha convertido en texto plano y, por consiguiente, se ha pasado al ordenador trazando así gráficas que muestran la trayectoria... es cuando entran en juego determinados programas informáticos capaces de crear ciertas curvas argumentales que representan la trama de dicha obra. Estas sirven, entre otros, para mostrarnos el destino de los personajes o para indicarnos en qué parte de la línea argumental nos encontramos.

Cada palabra analizada supone un tipo, una forma gráfica. Como ya se ha dicho, hay veces (y de hecho es lo más normal) en las que una palabra se repite dentro de una misma obra. Es por eso, que aunque esto suceda, dicha palabra va a quedar fijada como un tipo, lo que significa que en nuestro estudio va a aparecer solo una vez.

Para entender de una manera más sencilla la diferencia entre los *tokens* y los tipos, veamos el siguiente ejemplo: «E quel mio pariente tiene quel es muy grave cosa de sufrir aquellas terrerías quel fazen, e querría aventurarlo todo ante que sufrir tanto pesar de cada día. E porque yo querría que él acertasse en lo mejor, ruégovos que me digades en qué manera lo conseje porque passe lo mejor que pudiere en aquella tierra» (*Exemplo XXIX*).

Podemos ver cómo esta cita está compuesta por 56 *tokens*. Es decir, 56 palabras libres de cualquier aspecto visual. Por el contrario, el número de tipos no corresponde con el de los tokens. Contamos con 44 tipos debido a que las palabras *e*, *quel*, *sofrir*, *que*, *de*, *lo* y *mejor* se repiten y, como ya se ha explicado, estas quedan registradas la primera vez que aparecen.

En lo referido a los verbos sucede algo parecido. Las veces que estos pueden aparecer conjugados de una determinada manera, es realmente elevada. En estos casos, *lematizaremos* la raíz de los mismos para así clasificar, de una manera mucho más rápida, todas las formas verbales que nos vayamos encontrando en el texto. Al mismo tiempo, esta cuestión en ocasiones presenta problemas, tema que abordaremos en el siguiente apartado.

Aprovechando todas las menciones que se han hecho a los *token* a lo largo de este trabajo, quisiera hacer una aclaración que realmente nos va a ser útil para entender mejor el siguiente punto. Para ello, nos vamos a servir de la explicación que Brezina (2018) ofrece en su trabajo.

Un *token* es una sola ocurrencia de una forma de palabra en el texto. Cada vez que nos encontramos con una cadena de letras o números separados por el espacio en blanco (o puntuación) contamos esto como uno solo. Cuando se utilizan herramientas de corpus debemos tener muy claro cómo se implementa exactamente la definición de *token* y que cuando se obtienen los datos estadísticos (especialmente cuando se basan en herramientas diferentes) son comparables entre ellos debido a que estos tienen un efecto importante en todas las medidas basadas en la frecuencia de una palabra o frase.

Por otro lado, en el caso de que nos interesáramos por la base de la forma de las palabras, nos encontraríamos ante lo que conocemos como lema y lexema. Brezina (2018: 39) defiende que los lemas están basados en un análisis gramatical, mientras que los lexemas están basados en un análisis gramatical y semántico.

Además, hace otra división muy interesante de estos dos conceptos. Afirma que los *lemas* son «un grupo de todas las formas de inflexión relacionadas con una raíz que pertenecen a la misma clase de palabra» y que los *lexemas* (noción por la que nos interesaremos en este trabajo) son «un lema con un significado particular, que es necesario para distinguir las palabras polisémicas».

Aunque Brezina (2018) haya hecho esta distinción, a su vez, también afirma que *lexema* y *lema* son términos que se pueden intercambiar. Para ello, se ha servido de la definición de *lexema* propuesta por Biber *et al.* (1999: 54): «a group of related forms which share the same meaning and belong to the same word class (part of speech)»

4.1 ¿Cuáles son los principales problemas que presenta este análisis?

Mientras realizábamos este estudio nos hemos ido encontrando con una serie de problemas (ya mencionados en otros trabajos) a los que hemos tenido que hacer frente, así como a los nuevos que iban surgiendo.

4.1.1 Negación

Se trata de no afirmar un cierto enunciado, es decir, negarlo. Esta, se incluye dentro de la modalidad oracional marcada debido a que presenta una serie de marcadores (léxicos, sintácticos, etc.) que a reconocer si se trata de un enunciado negativo o no.

La negación presenta dos tipos principales: 1) la negación absoluta y 2) la negación relativa. El primer grupo se emplea para indicar un rechazo total, tanto por el adverbio *no*²³ como por expresiones complejas del tipo: *de ninguna manera, ni hablar...* Y, por el contrario, el segundo grupo indica la negación de un predicado que generalmente ha sido afirmado previamente.

Ambos presentan problemas a la hora de ser *tokenizadas*. En el caso de la negación absoluta, no podemos adjudicar la polaridad negativa a una palabra por el simple hecho de ir precedida del adverbio *non*.

²³ En esta obra aparece como *non*. Recuérdese que nos encontramos en el siglo XIV.

Veámoslo con el siguiente ejemplo. Si apareciese *matar*, inmediatamente se le adjudicaría la polaridad negativa (así como a todas las formas verbales que aparezcan en la obra). Pero si por el contrario apareciese *no matar*, es cuando nos encontraríamos ante el problema, ya que con la unión de dos palabras negativas hemos formado un sintagma que denota un sentido positivo.

Entonces, el ordenador haría el siguiente análisis: En un primer lugar, no sería consciente de que se encuentra ante un sintagma, por lo que lo analizaría palabra por palabra como ya hemos mencionado. Es decir, es este el momento que realmente diferencia a las máquinas de los seres humanos en lo que a la lectura se refiere. Se ha perdido el sentido positivo.

Y en segundo lugar, omitiría la presencia de *no* porque es un adverbio (recuérdese que analizamos adjetivos, verbos y sustantivos) y se centraría en la palabra *matar*, ya analizada previamente como negativa.

Nos encontramos ante un análisis erróneo por parte del ordenador, pues ha adjudicado el sentido opuesto. Esto tendrá una serie de consecuencias que se podrán ver reflejadas en las curvas argumentales entre otros.

Suponiendo ahora que esta negación absoluta se ve representada por expresiones complejas, estas, tampoco podrían ser *tokenizadas* por el mismo motivo, y es que el ordenador se encarga de analizar palabra por palabra, no sintagmas.

Por otro lado, en el caso de la negación relativa ocurre algo parecido. Esta, al referirse a un enunciado previamente afirmado, depende directamente del contexto argumental de la obra. Este aspecto el ordenador no lo detecta tal y como veremos en el siguiente apartado.

4.1.2 Semántica

La semántica es la «disciplina que estudia el significado de las unidades lingüísticas y de sus combinaciones» (RAE, 2019). Estas, pueden ser diferentes, y es por eso que dependiendo de la forma en la que se combinen, pueden significar una cosa o todo lo contrario.

Como bien afirman Archer y Jockers (2016: 20), esta es la principal diferencia entre las lecturas que hacemos los humanos y la que hacen las máquinas de un texto (aspecto que, desde mi punto de vista, es de gran relevancia): «The difference between the human reader and the machine reader is that the human understands that the content being read has meaning».

Con esto quiero decir que, al igual que hemos visto con el caso de la negación, el sentido y el significado que tiene un sintagma es lo primero que se pierde debido a la incapacidad de detectarlos que tienen los ordenadores, aumentando así las posibilidades de error.

Esta afirmación que acabamos de realizar en el párrafo anterior está muy ligada al ámbito, ya mencionado, de la polisemia y esta, a su vez, al contexto argumental. La pluralidad de significados en una misma palabra es algo que, lógicamente, el ordenador tampoco detecta y mucho menos, la variación de estos dependiendo del contexto argumental. El porqué está claro y es que, entre otros motivos, cuando un ordenador *tokeniza* una palabra, este ya la cataloga como tal para el resto de la obra.

Veámoslo con el siguiente ejemplo. Nos encontramos ante el caso del verbo *encubrir*. Este, aparece por primera vez en el *ejemplo XXVII* en la forma condicional, *encubriría*. Es a partir de este momento que se le adjudica la polaridad negativa a este verbo y por consiguiente, al resto de sus variantes en toda la obra.

Pero en el *ejemplo XLIX* nos encontramos con el problema y, más concretamente, en el siguiente enunciado: «E hizo la morada en un lugar tan **encubierto**, que nunca gelo pudieron entender los de aquella tierra quel dieron aquel señorío». Entendemos que denota un significado positivo por una razón principal: antes de esta cita se habla de todas las cosas malas que se le habían hecho a un hombre en concreto. Pero hubo otro bastante más astuto, que al observar lo que le estaban haciendo al prójimo, decidió llevar todas sus pertenencias a una casa más escondida, así estarían a salvo.

Al haber clasificado el verbo *encubrir* previamente como negativo, el ordenador hará el mismo procedimiento en este caso volviendo a errar. La solución más lógica a este problema viene de la mano de la estadística en lo que a este trabajo se refiere. Es decir, teniendo en cuenta el ratio de la palabra que estamos interesados, tendremos que ver el número de veces que esta aparece como positiva y como negativa. Dependiendo de la polaridad que presente el número más elevado, así será catalogada.

También se han dado casos, al igual que en la negación, en los que un único significado viene dado por la unión de, al menos, dos palabras. Un ejemplo de esto puede ser el caso del sintagma: *gatos religiosos* (*Exemplo XLII*). En conjunto, denota un sentido negativo ya que significa 'hipócrita'. Pero al ser analizado por el ordenador, y siguiendo las pautas ya explicadas, sabemos que, en un primer lugar, va a analizar estas dos palabras por separado (perdiendo así el sentido negativo). Y, en un segundo lugar, obviará *gatos* (por ser un sustantivo con polaridad neutra) y calificará *religiosos* como positivo. Es decir, obtendremos una vez más, un resultado opuesto al real.

4.1.3 Verbos

Por otro lado, tal y como afirma la RAE (2019), los verbos son una «clase de palabras cuyos elementos pueden tener variación de persona, número, tiempo, modo y aspecto». La combinación de estos elementos tiene como resultado las diferentes formas verbales que cada verbo presenta. Combinaciones que, aunque sean finitas, son lo suficientemente elevadas como para suponer un problema a la hora de ser analizadas.

Frente a esto y con la mirada puesta en toda la obra, proponemos dos opciones de análisis: 1) *tokenización* y 2) *lematización*. La primera es la que, desde mi punto de vista, menos complicaciones presenta ya que analizar un texto palabra por palabra garantiza de una manera u otra, que el sentido de estas se va a perder en menos ocasiones que si se hiciera lematizando.

Para lematizar verbos, hay que tener en cuenta la forma del infinitivo. Primero se hace una división de la palabra en raíz y desinencia verbal para a continuación atribuir una polaridad determinada a la raíz, catalogando de tal manera todas las variaciones que el verbo presente.

Veámoslo con el siguiente ejemplo. Si del verbo *matar* separáramos la desinencia verbal, nos quedaríamos con la raíz *mat-*. A esta le adjudicamos la polaridad negativa por lo que el infinitivo significa. Es entonces cuando aparezcan palabras como *mataron*, *matando*, *mata*, *matassen*, *matarse* o *matáronlo*, al presentar esta raíz, quedarían catalogadas directamente como negativas.

Este es un hecho que facilita, de una manera notoria, este tipo de análisis al ser humano además de suponer un ahorro del tiempo que se invierte. Pero hay ocasiones, como en los verbos irregulares, en las que esta lematización no es fructífera.

En el caso de *morir* podemos verlo de una manera muy clara. Si siguiéramos el proceso anteriormente mencionado, obtendríamos como raíz *mor-* y adjudicaríamos a esta una polaridad negativa. Es decir, a palabras como *moría*, *moriesse*, *morredes* o *moriéredes* se las asignaría automáticamente una polaridad negativa.

Pero a *muerto*, *murió* o *murades*, palabras que corresponden al paradigma verbal de *morir*, no se les atribuiría esta polaridad ya que su raíz no corresponde con la que se ha lematizado en un primer momento²⁴.

No quisiera cerrar este apartado sin hacer referencia a otro ejemplo que, desde mi punto de vista, es bastante interesante. En el caso del verbo *vencer* hay una peculiaridad. La polaridad que denota este verbo es claramente positiva, tanto en el infinitivo como en casi todas sus formas verbales. Pero es en la forma del participio pasado donde esta cambia, ya que *vencido* es una palabra con una polaridad negativa, incluso entendiéndolo como un adjetivo.

4.1.4 Otros problemas

Siguiendo con la idea de la lematización, añadir que no solo en el ámbito de los verbos presenta problemas, también en el de los sustantivos. Esta vez, la parte que se separa de la raíz corresponde con la de la flexión, es decir, el **género** y **número**. La principal consecuencia que tiene este hecho es, una vez más, la pérdida del sentido que denota cada palabra.

Continuando con esta idea, hay ocasiones en las que el cambio de género en una palabra, supone un cambio en el significado de la misma. Este tipo de ejemplos no aparecen en nuestro texto, la principal razón es por ser una cuestión bastante reciente. Nos referimos a los casos del tipo: *barco-barca*, *política-político*, *punto-punta*, *cesta-cesto* etc.

²⁴ Ocurre lo mismo con el verbo *encubrir*, cuyo participio pasado es *encubierto*.

Observamos que, por el hecho de que una palabra presente una raíz determinada, puede tener significados diferentes dependiendo de la desinencia que le acompañe. En el caso de los pares *barco-barca* y *cesto-cesta* la diferencia es mínima, ya que solo difiere en el tamaño de dicha entidad. Pero en *política-político* y *punto-punta* sabemos por su significado, que son entidades totalmente diferentes.

Por otro lado, en el caso de lematizar la palabra *niño*, la raíz sería *niñ-*. Así, palabras como *niña*, *niños*, *niñas*, *niñitos* etc. quedarían ligadas de manera automática a dicha raíz, obviando que entre *niño* y *niña* existe una clara diferencia en el sexo, al igual que entre *niño* y *niñito* hay un matiz diferenciador en el tamaño o entre *niñas* y *niña*, donde la diferencia es la cantidad de personas que denota dicha palabra.

Es decir, en cierto modo, cuando se lematiza en el caso de los sustantivos se hace inconscientemente, una división en lo que conocemos como familias léxicas o familias de palabras. Esto se debe al hecho de que el conjunto de palabras anteriormente mencionado presenta una misma raíz.

Siguiendo con esta idea, nos preguntamos qué pasaría con los prefijos en concreto²⁵, ya que estos son afijos que se colocan al principio de cada palabra, antes de la raíz. En el momento en el que lematizamos, estamos estableciendo que todas las variantes que va a presentar dicha raíz se van a colocar después de esta.

A partir de esta afirmación llegamos a pensar que, entonces, la palabra que esté formada por la presencia de un afijo, quedará excluida de dicha familia léxica. Podemos verlo con el ejemplo de *poder* (entendiéndolo como un sustantivo, no como un verbo), ya que dos de las palabras que forman la familia léxica son *poderoso* y *poderío*.

Estas dos palabras corresponden con la clara minoría de palabras que forma esta familia de palabras. Por el contrario, la mayoría corresponde a palabras que presentan un prefijo. Estas son: *apoderado*, *apoderados*, *desapoderado* y todas aquellas formas correspondientes del verbo *apoderar*.

También podemos verlo con la palabra *vergüenza* (o *vergüença*). Esta, presenta un gran número de palabras que forman parte de su familia léxica y están formadas con un prefijo. Son: *envergoñarse*, *envergonça*, *envergonçado*, *envergonçadamente* etc.

²⁵ Antes de lematizar habría que crear una lista de posibles desinencias que puede presentar un sustantivo. De otra manera, el ordenador no podría reconocerlos.

Aprovechando la palabra entre paréntesis del ejemplo anterior y basándonos en el propio texto, nos hemos dado cuenta de otra cuestión que se escapa del radio de la lematización. Nos referimos al hecho de que una misma palabra puede aparecer escrita de muchas maneras diferentes por razones históricas y gráficas. Estas variaciones, usualmente, se deben al cambio de una letra que corresponde en su mayoría con la última, como es el caso de *verdad* – *verdat*, *bondad* – *bondat* etc.

Pero también se dan ocasiones en las que cambia más de una letra y no precisamente la última. En la palabra *peligro* y sus variantes podemos verlo reflejado: *peligroso*, *peliglo*, *periglo*, *periglos*, *periglosos* etc.

Sirviéndonos del ejemplo anterior, quisiera añadir el caso de la palabra *milagro*, la cual también podemos encontrar con una apariencia diferente, y es la de *miraglo*. Como bien se ha mencionado, en casi el total de las ocasiones en las que sucede esto, la explicación se debe a razones histórico-gramaticales. Este ejemplo no iba a ser dispar, pues nos encontramos ante el fenómeno de la metátesis²⁶, correspondiente al ámbito de la fonética.

²⁶ «Cambio de lugar de algún sonido en un vocablo, como en *perlado* por *prelado*.» (RAE, 2019)

5. Lexicón

En este apartado presentamos el objeto de estudio de este trabajo, el diccionario de sentimientos en lengua medieval. Aclarar que las palabras, primero las positivas y luego las negativas (en ambos casos ordenadas alfabéticamente), son las correspondientes a la edición analizada.

POSITIVAS

abogados	aman	assossiego
abraçatvos	amar	avenidos
acabar	amar	avenidos
acertaba	amasse	aveniencia
acertar	amigas	avenir
açertar	amigo	ayuda
acertaron	amigos	ayudándole
acompañada	amo	ayudándose
acordaron	amola	ayudar
acorrer	amor	ayudarién
acorriese	amparad	ayudasen
acuçiosa	andança	ayudol
acuçioso	ángel	beber
acuerdo	ángeles	beguina
adormido	aparejamiento	bendicho
adoró	apercebimiento	bendiçión
afanes	aperçibido	bendito
agorero	apoderábanse	besáronle
aguisadas	apoderado	besó
aguisadas	apoderados	bien
aguisado	apoderándose	bien
alababa	apoderarse	bienaventurada
alabamiento	apostura	bienaventurado
alabando	aprender	bienes
alabar	aprendían	bodas
alcaldes	aprovecha	bondad
alegre	aprovechar	bondades
alegres	aprovechosa	bondat
alegría	aprovechosas	buen
alimpiábanle	aprovechoso	buena
alimpiar	apuesta	buenamente
alimpiassen	apuestos	buenas
allegar	arçobispado	buenaventura
alma	arçobispo	bueno
almas	ardides	buenos
alquilé	armados	buenos
altar	arripintiéronse	caballero
amaba	arte	cae
	assessegados	cardenal
	assossegado	cardenales

casada	criado	eguidas
casado	cristiano	emendaría
casamiento	cristianos	emendedes
casar	cristo	emienda
casaría	cruz	emiendas
casó	cuerda	empeçe
cassasen	cuerdo	emperador
castigar	cuerdos	emperadriz
catedral	cuida	ençalçar
católica	cuidado	endereçar
christiano	cuidan	enderesçar
çierta	cuidando	enrequescer
ciertamente	cuidar	ensalça
çiertamente	cuidaría	ensalçó
çierto	cuidase	enseñoreaban
clérigo	cumple	entençión
cobrar	cumplió	entendimiento
cobraría	cumplir	entendió
cobro	cumpliría	enxalçamiento
coidado	deán	ermar
comer	defender	esforçado
comía	defendimiento	esforçarán
comían	defienden	esforçose
comiese	defiéndete	esfuerço
compañero	delectosas	esperança
complían	delibrar	espulgaría
complida	derecha	esquimo
complidamente	derechamente	estados
complido	derechas	evangelio
complimientos	derecho	exentos
compliría	derechurero	facienda
conde	descansar	falagar
confessase	desengaña	falagos
conorta	desengañar	falaguera
conorte	desengañasse	falagueras
conosçiente	desengaños	falagues
conquistas	deseo	fama
consagraba	desfazer	fazaña
consagró	dignidades	faziendas
consejar	dinero	fe
consejedes	dineros	fechos
consejero	dios	fermosa
consejo	discípulos	fermosas
convidó	don	fermoso
convierta	donaire	fiaba
coraçón	dones	fiança
corte	donzella	fiando
creen	dormir	fiar
creuçendo	durmía	fidalgúa
creuçido	eglesia	fiesta

físicos	homenajes	loar
fiuza	homildat	loaredes
folgar	homildoso	logrero
folgasse	honra	loor
fortaleza	honrada	luzía
fortalezas	honrado	madre
fraire	honras	maduro
fraires	hostia	maestría
franco	hueste	maestrías
fremosa	humildat	maestro
fremosas	humillósele	maestros
fremosura	infante	mañeço
fuzas	jesu	manjares
fuzas	joglar	mantenencia
galardón	joglars	mantenimiento
galardone	jubglars	maravilla
ganaban	juez	maravillaba
ganado	juglars	maravillado
ganancias	juglar	maravillas
ganar	juizio	maravillase
ganaron	jura	maravilledes
ganó	justicia	maravilló
gloria	justiçia	maravillosas
governarse	justiçias	maravilloso
gracia	labredes	maravillosos
gracias	lavándole	mártir
gradeció	leal	mayoral
gradesçería	lealdad	mejor
gradesçió	lealdat	mejora
gradesçiólo	leales	mejores
gradesco	lealtad	mejoría
grasdeçia	león	mensura
gualardonaba	lesonía	merçed
guardado	librado	merçedes
guardar	librar	merçet
guardaría	libraría	meresçiera
guardedes	librasse	meresçieron
guardó	libró	meresçimiento
guaresçer	librol	meresçimientos
guarido	librólo	meresçió
guerrero	lidiar	mesurado
guerrero	ligeramente	milagro
guisa	ligereza	miraglo
guisado	limpias	misa
haber	loaban	misacantano
heredades	loades	moçedat
heredado	loado	moço
heredados	loados	naçer
heredan	loados	naçian
heredat	loamiento	naçido

nasçiera	poridades	sabio
nasçieron	postura	sabios
nasçiesse	preçiábala	salud
nasçió	preçiado	salvación
naturalmente	preçiadós	salvamiento
noble	preçiar	salvar
nobleza	preçiosas	salvaría
obediente	preciosas	salvasse
obispado	predigación	salvo
obras	predigasse	sancta
ofiçio	presçiaaba	sancto
oraçión	presçiar	sanctos
oraciones	príncipe	sano
orden	privado	sanos
orgullosamente	privados	santa cruz
oro	privança	sciencia
pagaba	pro	scriptura
pagado	proes	seglares
palabra	puñad	segurados
palacio	puñó	segurança
palaçio	quito	seguranças
papa	recabdo	sembraba
paraíso	reçibiolo	sembrar
parar	reçio	sentença
pararse	recodir	sentir
pareçieron	redemir	señor
partir	redimió	señores
partirá	regaço	señorío
paz	regalada	serviçio
pechados	reina	seso
perdón	reir	sesos
perdonado	reír	solaz
perdonasse	reis	soldán
perdonó	religiosos	soltar
perdonol	renta	sotil
philósopho	reverencia	sotiles
piadat	rey	sotileza
piadoso	rezio	sotileza
piedat	rica	talante
plata	rico	tesoro
plaze	riendo	tesoros
plazer	riesse	tierno
plazeres	ripintimiento	trabajar
plazería	riqueza	trabar
plazerme	riquezas	vagar
plogól	romería	valía
pobló	romerías	valiente
poder	sabidor	vençe
poderío	sabidora	vençedores
poderoso	sabiduría	vençer

vençieron
vençimiento
vençió
vender
vendiolas
ventura
verdad
verdadera
verdaderamente
verdadero
verdat
vevir
vianda
vida
virgen
virtud
virtudes
viva
vivir
vivo
voluntad
voluntades
voluntat
yantado

NEGATIVAS

aborreçe
aborrecen
achaque
acondidamente
acusadores
adebdado
adolesçió
afincado
afincamiento
afincó
afogó
alboroço
algarabía
amargos
amenanças
amenazas
amorteçer
amortiguamiento
aojen
apartado
apartados
apartamento
apremiaba

apremiar
aquexaba
aquexamiento
armáronse
armas
armassen
arrancar
arrastrábanla
arreatadamente
arreatades
arreató
arremetiose
arrincassen
arterias
artes
asañaba
asco
ascondidos
asperezas
assacador
aventura
balde
batalla
beguenería
bestias
bolliçio
brava
bravamente
bravo
burladores
çafondaba
çafondar
caídas
cansado
cansados
cara
cárcel
carga
casa
castiga
castigándol
castigar
castigaredes
cativo
cativo
cayó
ciego
cobarde
cobdiça
cobdiciars

cochiello
coita
coitado
colpe
colpes
conjuraste
conquerido
conquiera
contienda
contiedades
contradezir
contrallo
contrarios
corromper
corrompimiento
cortól
crebando
cruel
cruels
cruelmente
cuervo
cuidados
cuita
cuitado
culpa
culpada
culpado
daños
daño
daños
dañosa
dañoso
dapño
dar
debdos
degollaba
degollar
degollasse
degollóla
demandar
denuesto
denuestro
derranchado
derrangedes
desaguisadas
desaguisado
desamaba
desamor
desampararían
desapoderado

desavenencia	duelo	erraran
desavenençia	duelo	errárien
desavenir	echaron	erraron
desaviniésedes	embargaba	errasen
descabeçara	embargado	escapa
descalabrase	embargo	escapan
descendiendo	embargo	escariçiéndolo
descendiesse	empeçer	escarmiento
descoyuntamiento	empeorado	escarnecer
descoyuntara	enbraveçer	escarnecido
descoyuntolo	encantador	escarnezcán
desechar	encantamientos	escarnida
deserviçio	encerrado	escarnidos
desfecha	encerrar	escarnio
desfecho	enemigo	escarnios
desgradece	enfermedat	escasso
desheredados	enflaqueciera	escatimas
deshonra	enforçar	escondidas
deshonradamente	enforcassen	escondido
deshonrado	engafezió	escusa
desmayedes	engaña	escusar
desmayo	engañades	esforçar
despagaba	engañado	espada
despagar	engañar	espantado
despedeçaría	engañasse	espantarían
despendades	engañe	espante
despensa	engaño	espantedes
despreçiada	engaños	espanto
despreçiado	engañosa	espantos
despreçiar	engañosamente	estorbaba
desterrado	enloqueçido	estrañámelo
desterramiento	enojados	estrañeza
desterrar	enojo	estrañezas
destruyeron	enojos	estroído
desuso	enojosa	estroír
detruya	ensangrentado	estroir
devaneo	ensangrentó	falleçedera
diablo	ensañar	falleçedero
dolença	entristeçió	falleçeras
dolía	envergoçadamente	fallesçen
doliente	envergonça	fallesçido
dolientevo	envergonçado	falló
doliere	envergoñarse	falsa
dolor	envidia	falsedades
dubda	errábades	falsedat
dubdaban	errábamos	falso
dubdosas	erraban	falsos
dubdosas	errar	fambre
duda	errara	fea
duele	errara	fecho

feo	lazdrados	mentira
feos	lazería	mentiras
ferida	lazerías	mesquina
feridas	lid	mesquino
ferido	lides	miedo
feridos	llaga	miedos
feriéndola	llagas	mente
ferir	llorando	mintroso
ferrados	llorar	mora
fieramente	loco	moría
fieras	locura	moriéredes
finado	locuras	moriesse
fincar	maça	morir
finó	mal	moro
finquen	mala	moros
flaca	malas	morrá
flaco	malas	morredes
floxamente	maldad	mortal
foír	maldades	movimiento
folgura	malenconía	muchedumbre
forca	malfechor	muera
forçada	malo	muere
forçadamente	malos	muerta
fuessa	malquerencia	muerte
fuir	maltraer	muertes
fulano	maltraídos	muerto
gafedat	maltrayendol	muertos
gafo	maltrecho	muessos
golfin	mandado	murades
grave	manzellamientos	muriese
graves	mata	murió
guerra	matando	navaja
guerras	matar	negar
hereje	matara	nesçio
hipocrisía	mataría	nigromançia
homildades	mataron	ocasión
homillado	matáronlo	ocasionado
huyen	matarse	olvidar
infierno	matasse	olvidasse
infinta	matassen	ordir
ira	medrosa	orgullo
jubgado	menaçar	palancadas
judgado	mengua	panadizos
judgar	menguaba	partiese
judgaron	menguada	passiones
judguedes	menguaría	pecado
lágrimas	menguasse	pecador
lanças	menospreciada	pecadores
lazdrada	menosprecio	pecados
lazdrado	mentir	pelea

peleador	priessa	soberbios
peliglo	prior	soberbioso
peligroso	prisión	sofre
pena	probar	sofría
penas	puñada	sofrid
penitencia	quebrada	sofrió
penitência	quebrantado	sofrir
peor	quebrantar	sofrirá
peora	quebrantassen	sospecha
peores	quebról	suaba
perder	quebrólo	sufrades
perderán	quebrose	sufriéredes
perdería	quemar	sufriesse
perderían	quemarle	sufrió
perderla	quemasse	temo
perdían	quexa	temptaçión
pérdida	quexábase	temptar
perdido	quexarse	tentar
perdiera	quexasse	tenuda
perdieron	quexasse	terrorás
perdiesse	quexó	tienta
perdimiento	quexose	tirósvoslo
perdió	quexosos	tomar
periglo	rabia	tormentos
periglos	rabiar	torniose
periglosos	rebato	torpe
pesa	reçelaba	trabajos
pésame	reçelades	trabar
pesar	reçelando	trabarán
pesares	reçelo	trabarian
pesó	reçeló	trabaron
pésoles	renegado	traición
pierda	retrayéndol	traidor
pleito	revesadas	traidores
pobre	revesada	trebejo
pobreza	revoltoso	tremer
pobrezuelo	revoltosos	triste
podre	revuelta	tristeza
podrida	robados	tuerto
porfia	roído	tuerto
porfiaba	roído	tuerto
porfiado	rompiendo	ufana
porfiar	rompió	vana
porfiaron	rotos	vanagloria
porfias	sangrienta	vanas
poridat	saña	vanidat
premia	sañudo	vasallo
premia	sañudos	vençido
prender	sarna	vençidos
preso	soberbia	vençimiento

vengança
vengar
vengaría
vengasse
vergonçosamente
vergonçoso
vergüença
vergüenza
vicio
viçio
viçios
viçioso
vicioso
viles
yemas
yerra
yerran
yerro
yerro

6. Conclusión

Desde que el hombre se ha visto dotado del don de la palabra, se ha comunicado con el resto de los individuos y, en cierto modo, a dar su opinión de una manera tanto objetiva como subjetiva. Este hecho se ha visto muy ligado a la escritura, siendo esta el principal vehículo de información y de expresión desde el momento en el que se inventó.

Todos los trabajos que se han realizado a lo largo de la historia teniendo la escritura como base, sin importar la lengua en que esté escrita, son incalculables. Pero en este trabajo hemos añadido algo innovador, ya que se ha estudiado el texto de *El Conde Lucanor* desde el punto de vista del análisis de emociones y sentimientos.

Esta nueva forma de tratar un texto es relativamente reciente, ya que surgió a finales del siglo pasado. Pero ha sido al realizar este trabajo que hemos podido comprobar la gran popularidad que ha ido ganando a lo largo de estos últimos años, más concretamente en los siguientes dos aspectos: 1) se ha aplicado a un mayor número de lenguas y 2) el número de trabajos, estudios, etc. que se han realizado según estas bases, ha aumentado de manera notoria.

Otro aspecto que hemos podido comprobar es que este tipo de análisis es totalmente aplicable a textos literarios (ya que estos están cargados de emociones y de sentimientos) como pueden ser los cuentos de los hermanos Grimm o varias obras como las de Gabriel García Márquez. Pero no solo eso, les hay que han dado un paso más allá y han utilizado este método como una herramienta de trabajo, pudiendo así analizar la opinión de los clientes en positivas y negativas basándose en las reseñas publicadas.

Tras haber realizado este estudio, hemos podido comprobar que este tipo de análisis presenta una serie de problemas que todavía no se han solventado. Esto se debe a que primero se crea un diccionario de sentimientos por parte del ser humano y después se pasa al ordenador, perdiendo en ese proceso muchos sentidos que son esenciales para entender cualquier texto.

Es decir, las máquinas no son capaces de detectar si un texto presenta cualquier recurso estilístico (destacando la ironía) o no, debido a que estos corresponden con las principales características del habla. En otras palabras, todo lo referido a la semántica no será detectado y por consiguiente, se perderá gran parte del sentido del texto.

Como ya hemos visto, se han planteado dos formas principales de analizar este texto. La primera consiste en tokenizarlo y la segunda en lematizarlo. En ambos casos corremos el riesgo de perder algunos sentidos de determinadas palabras.

En el caso de tokenizar, afirmar que el principal problema es el de la negación. En este texto, los únicos aspectos negativos que van a aparecer vienen dados por la polaridad de determinadas palabras negativas, no por la negación. Esta se pierde en el momento en el que se analiza palabra por palabra porque en la mayoría de los casos la negación viene dada por la unión de dos palabras.

Por otro lado, en el caso de la lematización, la cantidad de sentidos que se pierden es mucho mayor debido a que esta no entiende ni de género ni de número, ni tampoco de verbos irregulares o de si una misma palabra aparece de diferentes formas por razones de la época.

Para concluir, quisiera dar mi opinión diciendo que, aunque el tiempo que se invierte en lematizar parece menor al que se invierte tokenizando, es mucho más productivo *tokenizar* ya que, aunque se sigan perdiendo muchos de los sentidos, el número de estos disminuye.

El porqué de esta afirmación se debe al número de excepciones que un mismo texto presenta y a las que, obligatoriamente, hay que hacer frente. Con lo cual, considero que es mejor tokenizar desde un principio a establecer una raíz, polarizarla y a continuación seguir trabajando para intentar solventar todos los problemas que las excepciones presentan.

En lo que a trabajos futuros se refiere, propongo intentar solucionar todo lo que tiene que ver con la negación, ya que considero que es un aspecto de gran relevancia dentro de un texto, pues puede cambiar por completo el argumento de cualquier obra.

7. Bibliografía

- Albornoz Cuadrado, J. C. de (2011). *Un modelo lingüístico-semántico basado en emociones para la clasificación de textos según su polaridad e intensidad* (Doctoral dissertation, Universidad Complutense de Madrid Madrid).
- Antúnez, S. B. (2009). «Las emociones como centro del impacto narrativo en la interacción con mundos ficcionales. Su necesidad para la teoría literaria», *Espéculo: Revista de Estudios Literarios*, (41), 71.
- Anscombe, E. and P. Geach. (1970). *Descartes Philosophical Writings*. Nelson: The Open University.
- Archer, J., & Jockers, M. L. (2016). *The bestseller code: Anatomy of the blockbuster novel*. Nueva York.
- Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Longman.
- Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide*. Cambridge University Press.
- Buonocore, D. (1976). *Diccionario de Bibliotecología: términos relativos a la bibliotecología, bibliografía, bibliofilia, biblioteconomía, archivología, documentología, tipografía y materias afines* Buenos Aires, Marymar, p. 61.
- Camus, A. (1991). *El extranjero*. Madrid.
- Delibes, M. (1950). *El camino*. Barcelona.
- Dubiau, L., & Ale, J. M. (2013). «Análisis de Sentimientos sobre un Corpus en Español: Experimentación con un Caso de Estudio», en *XIV Argentine Symposium on Artificial Intelligence (ASAI)-JAIIO*, 42, 36-47.
- Escortell Pérez, M. A. (2017). *El impacto de las emociones en el análisis de la polaridad en textos con lenguaje figurado en Twitter*. TFM. Valencia: Universidad de Valencia. (<http://hdl.handle.net/10251/86127>)
- Ehrlicher, H., Klinger, R., Lehmann, J., & Pado, S. (2019). «*Measuring Historical Emotions and Their Evolution: An Interdisciplinary Endeavour to Investigate The 'Emotions of Encounte*», *Liinc em Revista*, 15(1), en línea: <http://revista.ibict.br/liinc/article/view/4557>.
- Fernández Marcial, V. (2008). «*Marketing del libro: por qué funcionan los best sellers*», en *El libro en perspectiva. Una aproximación interdisciplinaria*. A Coruña: Universidad de A Coruña, 219-237
- Hogan, P. C. (2003). *The mind and its stories: Narrative universals and human emotion*. Cambridge University Press.
- Izard, C. E. (2013). *Human emotions*. Springer Science & Business Media.

- Jiménez, J. A. T., López, L. A. U., López, M. J. M., Mata, F. C., & de Salamanca Ros, F. E. (2015). "AORESCU: análisis de opinión en redes sociales y contenidos generados por usuarios", *Procesamiento del Lenguaje Natural*, (55), 153-156.
- Juan Manuel (2015), *El Conde Lucanor*, ed. Fradejas Rueda, JM. Barcelona: Penguin
- Mata Garrido, I. (2017). *Las técnicas del conlanging: un capítulo sobre la lingüística aplicada a la creación de lenguas*. TFG. Valladolid: Universidad de Valladolid. (<http://uvadoc.uva.es/handle/10324/25524>)
- Mohammad, S. (2011). *From once upon a time to happily ever after: Tracking emotions in novels and fairy tales*, En *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities* (pp. 105-114). Association for Computational Linguistics.
- Molina-González, M. D., Martínez-Cámara, E., Martín-Valdivia, M. T., Ureña-López, L. A. (2014). «Cross-domain sentiment analysis using Spanish opinionated words». *Natural Language Processing and Information Systems*, Lecture Notes in Computer Science, vol. 8455, pp. 214-219. Springer International Publishing. DOI: 10.1007/978-3-319-07983-7_28
- Molina-González, M. D., Martínez-Cámara, E., Martín-Valdivia, M. T., & Perea-Ortega, J. M. (2013). «Semantic orientation for polarity classification in Spanish reviews». *Expert Systems with Applications*, 40(18), pp 7250-7257.
- Murillo Lanza, D. F. (2017). *Análisis de sentimientos de Cien años de soledad y El amor en los tiempos del cólera de Gabriel García Márquez*. TFM. Valladolid: Universidad de Valladolid. (<http://uvadoc.uva.es/handle/10324/25437>)
- Ortony, A., Clore, G. L., & Collins, A. (1988). *The cognitive structure of emotions*. Cambridge.
- Palahniuk, C. (2001). *Asfixia*. Londres.
- Pang, B. y L. Lee. (2008). «Opinion Mining and Sentiment Analysis». *Foundations and trends in information retrieval*, 2(1-2):1– 135.
- Pang, B. & L. Lee. (2008). *Opinion Mining and Sentiment Analysis*. Now Publishers, Hanover, MA, USA.
- Pang, B., L. Lee, y S. Vaithyanathan. (2002) «Thumbs up?: Sentiment classification using machine learning techniques.» En *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10* (pp. 79–86). Association for Computational Linguistics.
- Parrott, W. G. (Ed.). (2001). *Emotions in social psychology: Essential readings*. Filadelfia: Psychology Press.
- Plata Rosas, L. J (2008). *Análisis sentimental y literatura*. Disponible en: <https://www.nexos.com.mx/?p=35866> [Fecha de consulta: 17 octubre 2019]
- Plutchik, R., & Kellerman, H. (1980). *Emotion, theory, research, and experience*. San Diego: Academic Press.

- Real Academia Española (2019). *Diccionario de la Lengua Española*. Disponible en: <https://dle.rae.es/>
- Sutherland, J. (2007). *Bestsellers: a very short introduction*. Oxford University Press.
- Vásquez, A. C., Quispe, J. P., & Huayna, A. M. (2009). «Procesamiento de lenguaje natural». *Revista de investigación de Sistemas e Informática*, 6(2), pp 45-54.
- Vilares, D., Alonso, M. A., & Gómez-Rodríguez, C. (2013). «Clasificación de polaridad en textos con opiniones en español mediante análisis sintáctico de dependencias.», *Procesamiento del lenguaje natural*, 50, pp 13-20.