

# LAS METÁFORAS Y SU IDENTIFICACIÓN AUTOMÁTICA: EVALUACIÓN DE DIFERENTES MÉTODOS DE DESAMBIGUACIÓN LINGÜÍSTICA SUPERVISADOS Y SEMI-SUPERVISADOS

Aantal woorden: 12.061

Catalina Capatina

Studentennummer: 01700397

Promotor(en): Prof. dr. Patrick Goethals

Masterproef voorgelegd voor het behalen van de graad master in de meertalige communicatie:  
talencombinatie Nederlands, Frans, Spaans

Academiejaar: 2020 – 2021



## **DECLARACIÓN SOBRE DERECHOS DE AUTOR**

De auteur en de promotor(en) geven de toelating deze studie als geheel voor consultatie beschikbaar te stellen voor persoonlijk gebruik. Elk ander gebruik valt onder de beperkingen van het auteursrecht, in het bijzonder met betrekking tot de verplichting de bron uitdrukkelijk te vermelden bij het aanhalen van gegevens uit deze studie.

## **AGRADECIMIENTOS**

En primer lugar, quisiera ofrecer mis más sinceros agradecimientos al profesor doctor Patrick Goethals. El tiempo que ha dedicado a guiarme en la escritura de esta tesina, sus comentarios y sus consejos han sido de gran ayuda. También me gustaría darle las gracias por haber despertado mi interés por este tema y por haberme hecho comprender su importancia en el campo de la lingüística computacional.

Luego, quisiera agradecer a Jasper Degraeuwe haberme guiado en la parte técnica de la presente tesina. Le agradezco especialmente su paciencia y sus múltiples explicaciones sobre la investigación. También me gustaría agradecerle el tiempo que ha dedicado codificando, procesando los datos y generando los resultados.

Por último, me gustaría agradecer a mi familia y amigos su apoyo emocional al largo de mi carrera académica.

## RESUMEN

El objetivo de la presente tesina es analizar si un sistema de desambiguación léxica automática (WSD) basado en el modelo de procesamiento de lenguaje BERT (Devlin et al., 2018) es capaz de clasificar correctamente los significados literales y metafóricos de palabras ambiguas. Más concretamente, se supone que el sistema debe ser capaz de superar el valor de la línea de base del sentido más frecuente (*most frequent sense*).

Para la recolección manual de datos se utilizan los corpus de SCAP (Goethals et al., 2018) y para su anotación se emplea la versión en línea del diccionario de uso del español actual, Clave (s. f.).

Mediante anotaciones manuales de 29 palabras ambiguas de diferentes clases morfológicas (12 sustantivos, 8 adjetivos y 9 verbos) y frases de semilla, el sistema clasifica los significados de las palabras de manera supervisada. Además, el sistema se evalúa sobre un enfoque de clasificación semi-supervisado, en el que el sistema añade frases de ejemplo para aprender nuevos aspectos de la tarea de clasificación. Sin embargo, los resultados mejoran tras añadir anotaciones de las frases que el sistema no logra clasificar. Por lo tanto, el sistema ha conseguido superar el valor del MFS para la mayoría de las palabras ambiguas.

Número de palabras: 200

Palabras clave: desambiguación léxica automática; WSD; BERT; MFS

## TABLA DE CONTENIDOS

|                                                            |    |
|------------------------------------------------------------|----|
| ABREVIATURAS .....                                         | 7  |
| LISTA DE TABLAS .....                                      | 8  |
| LISTA DE FIGURAS .....                                     | 8  |
| 1. INTRODUCCIÓN .....                                      | 10 |
| 2. LA METÁFORA: UN CASO CONCRETO DE AMBIGÜEDAD LÉXICA..... | 11 |
| 2.1 LA METÁFORA CONCEPTUAL .....                           | 12 |
| 2.2 SIGNIFICADOS METAFÓRICOS.....                          | 16 |
| 2.3 LA AMBIGÜEDAD LÉXICA .....                             | 17 |
| 2.4 HOMONIMIA.....                                         | 19 |
| 2.5 POLISEMIA .....                                        | 19 |
| 3. LA DESAMBIGUACIÓN LÉXICA AUTOMÁTICA .....               | 22 |
| 3.1 DESAFÍOS EN LA DESAMBIGUACIÓN LÉXICA AUTOMÁTICA .....  | 25 |
| 3.2 EL MODELO DE PROCESAMIENTO DE LENGUA BERT.....         | 27 |
| 4. METODOLOGÍA .....                                       | 28 |
| 4.1 ELECCIÓN DE PALABRAS AMBIGUAS EN SCAP.....             | 29 |
| 4.2 SIGNIFICADOS DE CLAVE.....                             | 30 |
| 4.3 ANOTACIÓN MANUAL DE LOS SIGNIFICADOS .....             | 31 |
| 4.4 ANOTACIÓN MANUAL DE LOS TEST SETS .....                | 35 |
| 4.5 ANOTACIONES ADICIONALES.....                           | 35 |
| 5. RESULTADOS Y ANÁLISIS DE ERRORES.....                   | 36 |
| 5.1 SUSTANTIVOS .....                                      | 37 |
| 5.2 ADJETIVOS .....                                        | 44 |
| 5.3 VERBOS.....                                            | 51 |
| 6. CONCLUSIÓN Y DISCUSIÓN.....                             | 56 |
| 7. BIBLIOGRAFÍA .....                                      | 60 |
| 8. APÉNDICE.....                                           | 62 |

## **ABREVIATURAS**

|      |                                                                  |
|------|------------------------------------------------------------------|
| WSD  | Word Sense Disambiguation (Desambiguación Lingüística)           |
| MFS  | Most Frequent Sense                                              |
| NLP  | Natural Language Processing (Procesamiento del Lenguaje Natural) |
| BERT | Bidirectional Encoder Representations from Transformers          |
| SCAP | Spanish Corpus Annotation Project                                |

## LISTA DE TABLAS

Tabla (1): Lista de lemas elegidos de SCAP después de haber consultado el diccionario Clave

Tabla (2): Lista definitiva de lemas después los pasos de filtración

Tabla (3): Datos obtenidos de la tarea de clasificación de los sustantivos

Tabla (4): Evolución del valor F del lexema 'salud'

Tabla (5): Análisis de errores del lexema 'salud'

Tabla (6): Frases dudosas del lexema 'salud'

Tabla (7): Evolución del valor F del lexema 'cielo'

Tabla (8): Análisis de errores del lexema 'cielo'

Tabla (9): Datos obtenidos de la tarea de clasificación de los adjetivos

Tabla (10): Evolución del valor F del lexema 'negro'

Tabla (11): Análisis de errores del lexema 'negro'

Tabla (12): Análisis de errores del lexema 'negro'

Tabla (13): Evolución del valor F del lexema 'bajo'

Tabla (14): Análisis de errores del lexema 'bajo'

Tabla (15): Datos obtenidos de la tarea de clasificación de los verbos

Tabla (16): Evolución del valor F del lexema 'desembocar'

Tabla (17): Análisis de errores del lexema 'desembocar'

Tabla (18): Análisis de errores del lexema 'desembocar'

Tabla (19): Evolución del valor F del lexema 'amasar'

Tabla (20): Análisis de errores del lexema 'amasar'

Tabla (21): Ejemplos de anotaciones erróneas

## LISTA DE FIGURAS

Figura (1): Significado literal y metafórico de la palabra 'calor'

Figura (2): Anotación manual del verbo 'amasar'

Figura (3): Frases en SCAP para cada significado del adjetivo 'ancho'

Figura (4): Frases en SCAP para cada significado del verbo 'mojar'

Figura (5): Evolución del valor F del lexema 'salud'

Figura (6): Evolución del valor F del lexema 'cielo'



Figura (7): Evolución del valor F del lexema 'negro'

Figura (8): Evolución del valor F del lexema 'bajo'

Figura (9): Evolución del valor F del lexema 'desembocar'

Figura (10): Evolución del valor F del lexema 'amasar'

# 1. INTRODUCCIÓN

El lenguaje diario y natural es lleno de expresiones metafóricas, más bien conocidas como simplemente metáforas o metáforas conceptuales. No solo sirven como recursos retóricos en las obras literarias para establecer algún tipo de comparación entre dos cosas, sino que también reflejan la realidad que nos rodea. La realidad es que las expresiones metafóricas sirven para expresarnos mejor, ya que las metáforas nacen de la realidad y la cultura. De hecho, estas expresiones son tan comunes que a menudo ni siquiera nos damos cuenta de que las estamos utilizando. Este estudio demuestra, por tanto, que las expresiones metafóricas se esconden en las palabras más banales, pero no se puede subestimar la forma en que se crean y el papel que desempeñan en el lenguaje. Además, varios estudios demuestran con qué frecuencia aparecen las metáforas en el lenguaje diario. Por lo tanto, estos estudios se centran en diferentes tipos de temas, como las "metáforas de la crisis en la prensa" (Arrese, 2015), "metáforas del cuerpo" (Pérez, 2010) o "metáforas del léxico futbolístico" (Medina Montero, 2007). Esta variedad de estudios demuestra, por supuesto, que las metáforas en sí mismas pueden ser muy variadas y que se producen con más frecuencia de lo que podríamos pensar.

La presente tesina presenta una investigación sobre un sistema de desambiguación lingüística, o desambiguación léxica automática, (WSD) basado en el modelo de procesamiento de lenguaje Bidirectional Encoder Representations from Transformers (BERT). La investigación no se centra en las palabras polisémicas en general, sino que se delimita por la identificación y clasificación automática de las palabras con al menos un significado literal y un significado metafórico. Estas palabras son un ejemplo de polisemia metafórica que surge del vocabulario de las metáforas conceptuales, tal y como lo definen Lakoff y Johnson (1980).

Desde la aparición de la desambiguación lingüística, las palabras ambiguas o con múltiples significados han planteado varios problemas. Las metáforas, y otros recursos estilísticos, son una forma especial de ambigüedad. En el contexto más amplio del procesamiento del lenguaje natural (NLP), la identificación y la clasificación correcta de las metáforas puede ser un paso importante para la traducción automática, pero también para el análisis de sentimientos o la extracción de opiniones de textos.

La metodología utilizada consiste en anotaciones manuales y evaluaciones de los resultados. En primer lugar, se han seleccionado palabras con un potencial significado metafórico de los corpus de Spanish Corpus Annotation Project (SCAP). A continuación, se ha utilizado el diccionario de uso del español actual Clave (s. f.) para determinar los diferentes significados de las palabras seleccionadas. Luego, se han anotado manualmente fichas para la parte supervisada del estudio. La parte semi-supervisada de la investigación consistió en que el sistema añadiera frases de ejemplo automáticamente y aprendiera de ellas. Las frases que el sistema no pudo clasificar porque los significados de las palabras resultaban demasiado vagos, se han anotado adicionalmente. Por último, se han evaluado los resultados obtenidos tras las diferentes etapas de la tarea de desambiguación.

El objetivo de la tesina consiste en evaluar si un sistema automático será capaz de superar el umbral del sentido más frecuente (MFS) de una palabra. El sistema aprende de las frases de ejemplo anotadas y luego añade nuevos ejemplos por sí mismo. Se evalúa tanto la parte en la que solo se utilizan las anotaciones manuales (supervisada) como la parte en la que el propio sistema ha añadido ejemplos automáticamente (semi-supervisada). Más concretamente, se evalúa qué enfoque para la clasificación logra mejores resultados. Además, el sistema también se evalúa después de haber anotado manualmente las frases sobre las que el sistema tenía dudas después de la parte semi-supervisada.

En los siguientes capítulos se explica qué son las metáforas y cómo se relacionan con la polisemia. A continuación, se aborda el problema de la traducción automática de las palabras polisémicas en el contexto de WSD. También se presenta el modelo de procesamiento de lenguaje BERT, seguido del apartado de la metodología. Los resultados se presentan por categoría morfológica, es decir, sustantivos, adjetivos y verbos.

## **2. LA METÁFORA: UN CASO CONCRETO DE AMBIGÜEDAD LÉXICA**

Las metáforas conceptuales no solo aparecen como figuras retóricas en las frases, sino que también pueden aparecer como lexemas. Una palabra con múltiples significados se conoce principalmente como una palabra polisémica. Sin embargo, si un significado de esa misma

palabra tiene es metafórico, se trata concretamente de polisemia metafórica. Los significados de una palabra metafórica están conectados de la misma manera que las diferentes palabras están conectadas en una metáfora conceptual.

En las siguientes partes de este apartado se explica paso a paso cómo se relacionan las metáforas y la polisemia, y como nacen las palabras metafóricas.

## 2.1 LA METÁFORA CONCEPTUAL

Una de las ideas fundamentales de Lakoff y Johnson (1980) es que las metáforas lingüísticas reflejan nuestras experiencias, o nuestra forma de entender y conceptualizar las experiencias. En concreto, una metáfora es la experiencia de una “cosa” que pertenece a un dominio conceptual abstracto, lo que representa el dominio conceptual A, en términos de otra “cosa”, que pertenece a un dominio conceptual físico o más concreto, también conocido como el dominio conceptual B. Así, dos conceptos que a primera vista no parecen estar relacionados, se conectan en el sistema conceptual humano. La razón por la que las metáforas funcionan de esa manera es porque existe una proyección entre los conceptos vinculados que tiene su base en la correlación de los dos dominios. Por consiguiente, el concepto concreto puede servir de ejemplo o base para hablar del concepto abstracto. Un ejemplo para ilustrar la idea anterior es la metáfora conceptual siguiente:

(1) <El tiempo es dinero.>

(Lakoff & Johnson, 1980, p.7)

En este ejemplo las palabras *tiempo* y *dinero* corresponden con un dominio conceptual meta (dominio conceptual A) y un dominio conceptual fuente (dominio conceptual B), respectivamente. Así que, el tiempo, que es un término más abstracto, se vincula con el dinero, que es un concepto más fácil de entender porque se puede tocar y ver. La idea de que el tiempo puede entenderse utilizando los mismos términos con los que hablamos del dinero se expresa de la siguiente manera:

(2) <No tengo tiempo para ir a la ciudad.>

(3) <No quiero gastar mi tiempo en esto.>

(4) <Nos quedaremos en un hotel por la noche para ahorrar tiempo mañana.>

A partir de los ejemplos (2), (3) y (4) resulta claro que, al igual que el dinero, el tiempo tiene mucho valor, se puede perder o se puede ahorrar. Al establecer esta conexión entre dos conceptos, podemos entender mejor el concepto más abstracto, que en este caso es el tiempo. Según Lakoff y Johnson (1980, pp.3 – 6) esta expresión muestra claramente que las metáforas conceptuales tienen su base en la correlación de los dominios involucrados. Esta conexión entre los dos dominios se expresa a veces mediante una forma de ambigüedad léxica, que se explicará con mayor detalle en el apartado 2.2.

Las expresiones metafóricas conceptuales se pueden categorizar en tres tipos en función del concepto metafórico expresado (Lakoff & Johnson, 1980):

a) metáforas estructurales

Las metáforas estructurales son el tipo más común de metáfora conceptual, ya que son expresiones en las que un concepto se estructura en términos de otro. El dominio fuente, es decir, el dominio B, determina la forma en que pensamos y hablamos de las entidades y actividades a las que se refiere el dominio meta (dominio conceptual A), e incluso la forma en que nos comportamos o actuamos como seres humanos.

La expresión <el tiempo es dinero> ilustra perfectamente este tipo de metáfora. Asimismo, existen muchas otras expresiones que pertenecen a esta categoría:

(5) <Una discusión es una guerra.>

(6) <El amor es un viaje.>

(7) <Las ideas son modas.>

(Lakoff & Johnson, 1980, p.14)

b) metáforas orientacionales

Estas expresiones metafóricas se distinguen de las estructurales por su forma de estructurar los conceptos. Más concretamente, las metáforas orientacionales no vinculan solamente los conceptos, sino que “organizan un sistema global de conceptos con relación a otro”. En este contexto se trata de conceptos que tienen que ver con la orientación espacial como *arriba-abajo*, *profundo- superficial*, *dentro-fuera*, etc.

(8) <Alto es positivo.>

(9) <Bajo es negativo.>

En consecuencia, se deduce que estos conceptos se utilizan en el lenguaje para describir no solo el espacio, sino también sentimientos, relaciones entre las personas o cantidades. Para ilustrar esta idea, se puede utilizar la expresión *<estar bajo de ánimo>*. La palabra *bajo* relaciona el concepto de emociones con una orientación espacial. La emoción de tristeza o decepción se vincula con la orientación espacial de abajo. Por otra parte, la emoción de alegría siempre estaría relacionada con la orientación espacial de arriba. Este vínculo entre el espacio y otros dominios conceptuales se manifiesta en muchas más expresiones, como *<tener una posición elevada>*, *<alcanzar un pico>* o *<ser de alta calidad>*.

(Lakoff & Johnson; 1980, pp.14 – 17)

#### c) metáforas ontológicas o metáforas de sustancia y entidad

Los conceptos que se expresan mediante este tipo de metáfora se presentan como sustancias o entidades. Las metáforas ontológicas nos permiten conceptualizar cosas, experiencias y procesos muy vagos o abstractos, como si tuvieran propiedades físicas definidas.

(10)       *<La inflación está bajando nuestro nivel de vida.>*

En el ejemplo, la inflación se presenta como una entidad que luego se puede cuantificar, al que se pueden dedicar características y que tiene diferentes dimensiones. Por otro lado, mediante estas metáforas, también se pueden identificar los aspectos particulares de un concepto (*<el lado malo de la inflación>*), las causas de un concepto (*<la demanda de bienes fue la causa de la inflación>*), etc.

(Lakoff & Johnson, 1980, pp.25 – 32)

Sin embargo, los límites de esas tres categorías no son claras en algunos casos. Las categorías pueden solaparse, ya que las metáforas estructurales y orientacionales pueden tener también algunas características de las metáforas ontológicas. Esas últimas, a su vez, se caracterizan por el dominio conceptual fuente que tiene que ser estructurado. (Knowles & Moon, 2004, p.31)

Las metáforas conceptuales son mecanismos cognitivos, que organizan nuestra manera de pensar, y que luego se pueden traducir en expresiones lingüísticas metafóricas. Por así decirlo, las expresiones lingüísticas son los síntomas de nuestra manera de pensar. Las metáforas

lingüísticas son palabras u otras expresiones que nacen del lenguaje o la terminología del dominio conceptual más concreto, es decir, del dominio B (Kövecses, 2010, p.4). Así, la metáfora conceptual <el tiempo es dinero> se refleja en el lenguaje cotidiano en frases comunes como las siguientes (11) y (12):

(11) <Me estás haciendo **perder** el tiempo.>

(12) <He **invertido** mucho tiempo en ella.>

(Lakoff & Johnson, 1980, pp.7 – 8)

En Lakoff y Johnson (1980, p.7), también se explica el papel de las expresiones lingüísticas metafóricas. Concretamente, el vocabulario que puede vincularse al dominio conceptual B es una forma de hablar también del dominio conceptual A. En los ejemplos (11) y (12), los verbos *perder* e *invertir*, hacen que el concepto del tiempo se entienda sistemáticamente en términos de dinero.

También cabe distinguir las metáforas creativas o nuevas de las metáforas convencionales o muertas. Las metáforas creativas se describen generalmente como las que se construyen para expresar una idea concreta en un contexto determinado. Además, el lector tiene que saber que se trata de una metáfora, o un lexema con un significado metafórico, para entender lo que el autor quiere decir. Este no es el caso de las metáforas convencionales, ya que están completamente integradas en la lengua. El lector, o el autor, a veces ni siquiera se da cuenta de que está utilizando expresiones metafóricas. En realidad, la mayoría de las metáforas se usan convencionalmente, y existe la impresión, incorrecta, de que, por esta razón, son muertas. Sin embargo, según Kövecses (2010, p.11), incluso si las metáforas son muy convencionales y se utilizan sin esfuerzo, esto no significa que se haya perdido el proceso cognitivo y que estén muertas. La clasificación de las metáforas en términos de convencionalidad es un tema complicado y a veces confuso (Kövecses, 2010, p.33). Por eso, esta tesina no se centra en el lugar de ciertas metáforas en una escala potencial de convencionalidad, sino que se investigan varias metáforas convencionales y creativas. No se hace distinción entre los dos tipos, ya que los corpus utilizados como datos de entrenamiento también contienen textos literarios.

## 2.2 SIGNIFICADOS METAFÓRICOS

En consecuencia, se puede deducir que el vocabulario que surge de las metáforas conceptuales constituye una forma concreta de ambigüedad léxica. Las metáforas pueden producirse como formas léxicas, y sus diferentes significados pueden relacionarse entre sí de manera metafórica. Estos lexemas plantean un problema particular para su identificación y traducción automáticas, dado que los significados metafóricos dependen en gran medida del contexto. Los lemas con los que se trabajará en la investigación de la presente tesina forman parte del vocabulario que nace de las metáforas conceptuales generales y que forman parte de potenciales expresiones lingüísticas metafóricas.

Los lemas que físicamente vinculan los dos dominios conceptuales tienen en sí mismo entonces dos (o más) significados que pertenecen a diferentes dominios. Si se parte de la idea de que el significado que pertenece al dominio conceptual B (dominio conceptual fuente) es el significado literal del lema, se puede asumir que el otro significado pertenece al dominio conceptual A y que es metafórico. Para ilustrar la idea anterior se presentan los siguientes ejemplos:

(13) Metáfora conceptual:

*<Las emociones son temperaturas.>*

Expresiones lingüísticas metafóricas:

*<Busca en los amigos **el calor** que no encuentra en casa.>* (Clave, s. f.)

*<Discutían con **calor**, pero son buenos amigos.>* (Clave, s. f.)

(14) Metáfora conceptual:

*<La inteligencia es una fuente de luz.>*

Expresiones lingüísticas metafóricas:

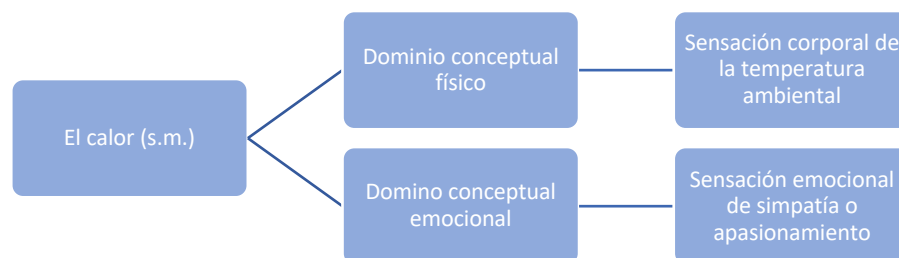
*<Su inteligencia **brilla** allá donde va.>* (Clave, s. f.)

*<Cam se había metido allí tratando de escapar por un momento de la obligación de **brillar** para los diversos invitados, todos ricos y poderosos.>* (SCAP)

En el ejemplo (13) se presenta la palabra *calor* como lema que ha nacido de la metáfora conceptual *<las emociones son temperaturas>*. En el diccionario explicativo español, Clave (s.



f.), se definen varios significados de esa palabra. Unos de los significados tienen que ver con la “sensación corporal de la temperatura” y otros con la “sensación emocional de simpatía o apasionamiento”. Partiendo de la idea de que en la metáfora conceptual las temperaturas ambientales son un concepto más concreto que las emociones humanas, se pueden distinguir entonces dos dominios: un dominio conceptual físico y un dominio conceptual emocional. En consecuencia, se puede deducir que el significado de la sensación emocional es efectivamente un significado metafórico de la palabra *calor*. La misma línea de argumentación se aplica al ejemplo (14). La figura siguiente demuestra esquemáticamente el hilo conductor de la palabra *calor* con sus significados metafóricamente vinculados.



*Figura (1): Significado literal y metafórico de la palabra “calor”*

Otra forma de analizar los significados metafóricos de las palabras es la que se presenta en Knowles y Moon (2004, pp.7 – 8). En esta forma de análisis es importante distinguir entre tres cosas: la palabra metafórica (vehicle/medio), su significado metafórico (topic/tema) y la conexión entre los dos (grounds/fundamento). Si este análisis se aplica al ejemplo (13), se puede llegar al siguiente resultado:

*Modelo: el calor*

*Tema: sensación emocional de simpatía o apasionamiento*

*Fundamento: idea de sensación agradable*

## 2.3 LA AMBIGÜEDAD LÉXICA

Los significados metafóricos de las palabras presentan problemas en el proceso de la comunicación y, en general, en el lenguaje. En consecuencia, estos significados metafóricos pueden presentar un obstáculo en la acción de la comunicación, ya que esta última solamente

se puede clasificar como realmente exitosa si el destinatario interpreta correctamente el mensaje del autor o hablante.

El fenómeno general de la ambigüedad léxica aparece en diferentes formas en el lenguaje, es decir de manera sintáctica y semántica. La ambigüedad léxica sintáctica, tal como se presenta en el ejemplo (15), se refiere a la ambigüedad generada por la categoría gramatical de las palabras, mientras que la ambigüedad léxica semántica, propuesta en el ejemplo (16), es generada por una palabra que tiene más de un significado. Es cierto que en español la ambigüedad léxica sintáctica es menos frecuente, ya que hay una clara diferencia entre las formas de las diferentes clases gramaticales (por ejemplo, verbos, sustantivos y adjetivos). Sin embargo, esto no excluye que pueda surgir una confusión en términos de ambigüedad léxica, como ilustrado en el ejemplo (15) que presenta una forma verbal de *cantar* o el sustantivo *canto*. (Small, Cottrell & Tanenhaus, 1988, p.4)

(15) <*Canto espléndido.*> (Internet)

(16) <*La novela realista busca ser un espejo de la sociedad.*> (Clave, s. f.)  
<*Fue al baño a mirarse en el espejo.*> (SCAP)

La ambigüedad léxica semántica tiene un cierto vínculo con las metáforas, como ilustrado en la figura (1). La presente tesina se limita a este tipo de ambigüedad léxica y solamente se ha trabajado con palabras con varios sentidos dentro de una misma categoría gramatical.

Teniendo en cuenta las explicaciones de Small, Cottrell y Tanenhaus (1988), la ambigüedad léxica se puede describir de la siguiente manera: todas las palabras que tienen más de un significado se incluyen en esta categoría. Sin embargo, todavía hay una distinción importante dentro de la categoría de la ambigüedad semántica: homonimia y polisemia. Esta distinción es particularmente necesaria porque la homonimia y la polisemia demuestran claramente la relación entre la ambigüedad léxica y las metáforas.

## 2.4 HOMONIMIA

El término *homonimia* se aplica a las palabras que se escriben o se pronuncian de la misma manera, pero tienen dos o más significados sin algún tipo de vínculo o relación semántica entre ellos. Para ilustrar este tipo de ambigüedad léxica, véanse los siguientes ejemplos:

- (17) <Tomaba el sol sentado en un banco del parque.> (Clave, s. f.)
- (18) <Fui al banco para abrir una cuenta corriente.> (Clave, s. f.)
- (19) <Un banco bacaladero.> (Clave, s. f.)

Los tres ejemplos presentan tres significados distintos incluidos en el diccionario explicativo español Clave (s. f.) del sustantivo masculino *banco*, refiriéndose a un asiento, una oficina bancaria y un conjunto de peces, respectivamente. Dado que estos tres significados no presentan ninguna relación entre ellos, se pueden calificar como homónimos.

(Small, Cottrell & Tanenhaus, 1988, p.4)

## 2.5 POLISEMIA

A diferencia del caso de la homonimia, las palabras polisémicas tienen significados que sí están relacionados entre sí (Small, Cottrell & Tanenhaus, 1988, p.4). Para ilustrar la idea de la polisemia se presentan de nuevo el ejemplo de la palabra *calor* (también ilustrado en el ejemplo (13) y la figura (1)):

- (20) <Si tienes calor, apaga el radiador.> (Clave, s. f.)
- <Discutían con calor, pero son buenos amigos.> (Clave, s. f.)

El ejemplo ilustra dos significados del sustantivo *calor*, uno de los significados siendo la sensación corporal y el otro la intensidad de la acción. La relación semántica entre los dos significados de la palabra *calor* son fáciles de distinguir. Esa situación de semejanza entre dos sentidos diferentes de una palabra se puede relacionar con la idea esencial de las metáforas según Lakoff y Johnson (1980, p.5), que implica que una metáfora es el entendimiento y la experiencia de una cosa en términos de otra. Así que el sentimiento de *calor* durante una discusión se entiende en términos de intensidad emocional o intensidad de la acción.

Como consecuencia se puede establecer un vínculo entre las metáforas y las palabras polisémicas. En realidad, las metáforas actúan como la fundación de la polisemia, ya que son proyecciones globales de diferentes dominios que influyen en la manera de hablar y pensar sobre cierto concepto (Lakoff & Johnson, 1980, p.5). La coherencia entre los diferentes conceptos semánticos es lo que crea un nuevo significado metafórico de una palabra. De esta manera, una palabra polisémica consigue un significado metafórico que se entiende en términos del sentido literal de la palabra en cuestión. (Bowdle & Gentner, 1999, p.91)

En el marco de la polisemia también se puede destacar la diferencia entre las metáforas creativas y las metáforas convencionales. Las metáforas nuevas se clasifican ante todo como comparaciones, porque la palabra en cuestión aún no tiene un significado metafórico. Como consecuencia, la palabra no es en sí misma una metáfora, sino que la comparación en su totalidad se considera como una metáfora.

Las metáforas convencionales también se pueden clasificar como comparaciones, pero pueden funcionar como metáforas en sí mismas. Conviene subrayar que las metáforas convencionales se clasifican como palabras que tienen dos sentidos diferentes, con al menos un significado metafórico. (Bowdle & Gentner, 1988, p.91)

Como se ha explicado anteriormente, en la presente tesina se trabaja solamente con las metáforas convencionales, ya que los lemas con los que se trabaja en la investigación tienen dos o más significados, entre los cuales figura al menos un significado metafórico. Las palabras se pueden clasificar como metafóricas, ya que no es necesario hacer una comparación para llegar a una expresión metafórica.

Finalmente, también cabe hacer hincapié en la polisemia metafórica y la polisemia metonímica, dos conceptos que, según algunos, no deben ser confundidos, aunque estén estrechamente relacionados. Según Lakoff y Johnson (1980, pp.35-40), la metonimia puede definirse como una parte de algo o alguien que representa la totalidad de una persona, una cosa o un grupo. El concepto de la metonimia se explica mediante el ejemplo de la palabra *cabeza*:

(21) <Los toros tienen cuernos en la cabeza.> (Clave, s. f.)

<Tenía un rebaño de doscientas cabezas.> (Clave, s.f.)

El primer ejemplo se refiere a la parte superior del cuerpo de un animal (o una persona), mientras que *las cabezas* en la segunda frase indican la totalidad de los animales del rebaño. El ejemplo (21) representa el tipo de metonimia que se clasifica como *la parte por el todo*, pero no es el único tipo. La metonimia se clasifica en los demás tipos:

a) *Una obra/un producto por su productor*

<Compró un Ford.>

b) *Un instrumento por el usuario*

<El saxo tiene la gripe hoy.>

c) *El controlador por el controlado*

<Nixon bombardeó Hanoi.>

d) *El lugar por los acontecimientos*

<Watergate cambió a nuestros políticos.>

e) *El lugar por la institución*

<La Casa Blanca no dice nada.>

f) *Una institución por la gente responsable*

<El Senado piensa que el aborto es inmoral.>

(Lakoff & Johnson, 1980, pp.38 – 39)

A primera vista, parece que Lakoff y Johnson (1980, p.36) distinguen las metáforas de las metonimias, argumentando que la metáfora se utiliza principalmente para comprender una cosa en términos de otra, mientras que la metonimia sirve para “conceptualizar una cosa en virtud de su relación con otra”.

Concretamente, esto significa que la metonimia forma una relación basada en la contigüidad, y no en similitud o correlación, como en el caso de las metáforas. La metáfora implica que los dos conceptos relacionados tienen cierta distancia entre ellos en nuestro sistema conceptual. Esa distancia se debe a la abstracción del dominio A y a la especificidad del dominio B. En la metonimia, en cambio, se presentan típicamente dos elementos que están estrechamente relacionados entre sí en un espacio conceptual. (Kövecses, 2010, p.175)

Como consecuencia, en el caso de la metonimia, no se trata de dominios conceptuales tan diferentes, o marcados por una distancia significativa. Esta es generalmente la mayor distinción entre las expresiones metafóricas y metonímicas.

Sin embargo, siguiendo la argumentación de Lakoff y Johnson (1980, p.36), se revelan algunas similitudes entre los dos conceptos. Por ejemplo, gracias a los conceptos metonímicos podemos conceptualizar una cosa a través de su relación con otra cosa (Lakoff & Johnson, 1980, p.36), lo que significa que la metonimia y la metáfora son dos fenómenos de origen conceptual. A través de la metonimia conceptual se forman expresiones lingüísticas metonímicas (Kövecses, 2010, p.173). Este argumento puede ser vinculado directamente con la definición presentada anteriormente de las metáforas. Por último, se explica que los conceptos metonímicos son capaces de estructurar nuestras acciones, ideas y actitudes. Además, están fundados en nuestras experiencias diarias, como las metáforas. (Lakoff & Johnson, 1980, p.39)

Kövecses (2010, p.184) promueve la idea de que metáforas y metonimias están vinculadas, e incluso añade que la metonimia forma la base de algunas metáforas conceptuales. Además, a veces resulta difícil establecer una línea divisoria clara entre expresiones metonímicas y metafóricas. En esta tesina, se considera la metonimia como una subcategoría de las metáforas, ya que la exclusión de algunos significados frecuentes metonímicos de los lemas puede influir en la identificación de sentidos no literales o en los resultados obtenidos. Sin embargo, esta opción pragmática no debe ser confundida con la alegación de que todas las metáforas conceptuales tienen una base metonímica.

### **3. LA DESAMBIGUACIÓN LÉXICA AUTOMÁTICA**

Como se ha comentado en los apartados anteriores, el lenguaje está lleno de ambigüedad. Para los seres humanos la desambiguación de los diferentes significados de una palabra es un fenómeno relativamente fácil, y muy a menudo inconsciente. Sin embargo, la desambiguación léxica sigue siendo un proceso difícil en la lingüística computacional, ya que el significado de una palabra ambigua depende en mayor parte del contexto. Analizando el contexto, un

sistema de desambiguación léxica automática implica que se determina computacionalmente el sentido de una palabra ambigua.

Por eso, el primer paso de cualquier sistema de desambiguación léxica automática es transformar el contexto en un formato legible y estructurado para el sistema. Esta tarea se divide en varios pasos:

1. *Tokenization*: el texto se fragmenta en partes, a menudo en palabras.  
<Ese / libro / es / una / basura.>
2. *Part-of-speech tagging*: se asigna una clase gramatical a cada parte que ha resultado la primera etapa.  
<Ese <sup>pron.dem.</sup> / libro <sup>sust.m.</sup> / es <sup>v.</sup> / una <sup>art.indeterm.</sup> / basura <sup>sust.f.</sup>>
3. *Lemmatization*: la palabra se transforma de nuevo en su forma básica.  
<Ese / libro / estar / una / basura.>
4. *Chunking*: dividir la frase, o el texto, en partes correlacionadas.  
<[Ese / libro] / [estar / una / basura].>
5. *Parsing*: hacer una composición esquemática de la oración.

Navigli (2010, p.11)

Luego, el sistema tiene que calcular y dotar a cada lema de un vector. Para ello, se eligen características (*features*) que representen el contexto. Las características sirven como una especie de indicación del significado correcto de la palabra. Estas características pueden ser identificadas mediante los pasos anteriores. Algunos ejemplos de características son: las palabras que rodean a la palabra ambigua, el texto o la frase en su conjunto o las características semánticas o sintácticas. (Navigli, 2010, p.12)

Esencialmente, un sistema de desambiguación léxica automática funciona como un clasificador que considera los diferentes sentidos de una palabra como las clases. En consecuencia, la clasificación automática asigna un ejemplo a una o más clases, basándose en el contexto directo de una palabra, pero también en fuentes de conocimiento externas, que pueden variar considerablemente. Las fuentes de conocimiento son necesarias para el procedimiento fundamental de cualquier sistema de desambiguación léxica. Por lo general, un sistema parte de una serie de palabras con múltiples significados. Para asignar el significado

correcto a cada palabra en función de su contexto, el sistema utiliza fuentes de conocimiento. Las fuentes de conocimiento pueden variar significativamente de un sistema a otro: hay sistemas cuya base consiste en corpus de textos anotados con los sentidos de las palabras, pero también existen sistemas que utilizan recursos más estructurados, como diccionarios. (Navigli, 2009, pp.2 – 4)

En función de la principal fuente de conocimiento utilizada para la definición de los diferentes sentidos de las palabras ambiguas, se pueden distinguir diferentes enfoques en la desambiguación léxica automática.

En primer lugar, existen los sistemas basados en conocimientos o diccionarios bilingües. Como su nombre lo implica, los conocimientos utilizados en este tipo de sistemas provienen de diccionarios, tesauros y bases de conocimiento léxicas. La gran ventaja de este tipo de sistemas es que se pueden emplear para la desambiguación de todas las palabras en un texto, y no solamente para la desambiguación de las palabras anotadas, como en los sistemas que se basan en datos de corpus. (Mihalcea, 2007, p.107)

En segundo lugar, existen tipos de enfoques que se basan en datos de corpus y que se pueden dividir en enfoques supervisados, semi-supervisados y no supervisados. Los métodos supervisados utilizan corpus anotados como datos de partida. Más precisamente, los sistemas supervisados pueden entrenarse utilizando una serie de ejemplos anotados manualmente y unas frases de ejemplo (*seed sentences*).

Este tipo de enfoque se centra en la desambiguación léxica en un conjunto específico de palabras. Sin embargo, en los sistemas supervisados puede surgir el problema de la escasez de datos. La razón es la amplitud del vocabulario de una lengua: es poco probable que se disponga de un conjunto de entrenamiento de tamaño que cubra todo el léxico de una lengua. (Navigli, 2009, p.4)

Los resultados de los sistemas supervisados también suelen estar afectados por errores humanos en los datos de entrenamiento. Especialmente cuando se clasifican los diferentes significados de una palabra en un contexto limitado, esto puede ser un problema.



Los enfoques no supervisados no clasifican los sentidos de palabras basándose en la anotación manual de datos de partida, sino en información encontrada en corpus no anotados. Este enfoque se basa generalmente en corpus monolingües y deduce que las palabras que suelen aparecer en un contexto similar tienen significados similares. También pueden basarse en corpus paralelos, o sea corpus con textos y sus traducciones. De esta manera se pueden identificar los significados recurrentes de una palabra en la lengua meta teniendo en cuenta el significado de una palabra en la lengua fuente. (Pedersen, 2007, pp.133 – 134) Generalmente, se deriva que los sistemas supervisados presentan resultados más precisos, como ha demostrado Agirre y Elmonds (2007). (Màrquez et al., 2007, p.168).

Últimamente, los enfoques semi-supervisados son una mezcla de los enfoques supervisados y no supervisados. Esto significa que un sistema utiliza tanto datos anotados como datos no anotados. En este caso, un sistema parte de un pequeño conjunto de datos anotados y aplica los conocimientos adquiridos sobre los significados de una palabra a datos nuevos. Los enfoques semi-supervisados son especialmente interesantes en la desambiguación léxica automática, ya que como en cualquier tarea de clasificación, se necesitan muchos datos anotados manualmente. Esto, a su vez, conduce a la escasez de datos, como es el caso en los enfoques supervisados. (Van Engelen & Hoos, 2020, p.374)

En la tarea de la desambiguación léxica automática específicamente, los enfoques semi-supervisados clasifican unos ejemplos nuevos basándose en los ejemplos anotados manualmente, que funcionan como frases semilla (*seed sentences*). Tras esta tarea de clasificación, el sistema añade los ejemplos en los que tiene un mayor nivel de certeza a la lista de datos anotados. Además, el sistema puede repetir el proceso de aprendizaje y aplicar los nuevos conocimientos adquiridos a nuevos datos. (Taghipour & Tou Ng, 2015, p.315)

### **3.1 DESAFÍOS EN LA DESAMBIGUACIÓN LÉXICA AUTOMÁTICA**

Desde el principio, la desambiguación léxica automática se ha añadido a la lista con los problemas computacionales de inteligencia automática (IA) completos, un término que se utiliza para describir un problema cuya solución implica la comprensión completa del lenguaje natural o del sentido común (Agirre & Elmonds, 2007, p.1).

El problema de la desambiguación automática resulta de varios factores, por ejemplo, del hecho de que los sistemas supervisados y semi-supervisados tienen que estar basados en datos anotados. De hecho, la anotación manual de recursos de conocimiento es un esfuerzo costoso y lento y debe repetirse cada vez que cambia algo en los sentidos de una palabra desambigua. Este problema muy importante en la desambiguación léxica automática se denomina “*knowledge acquisition bottleneck*” y es en gran medida la consecuencia de la ausencia de datos de corpus anotados. (Navigli, 2009, p.2)

Sin embargo, los métodos de aprendizaje no supervisados pueden superar este problema. Pero, a su vez, este enfoque también tiene un gran inconveniente según Navigli (2009, p.26). Los sistemas no supervisados no utilizan ningún tipo de datos de partida, y el proceso de clasificación a menudo no es preciso. Además, un sistema puede tardar mucho tiempo en aprender por sí mismo las características de las palabras ambiguas.

En Agirre y Elmonds (2007, pp.17 – 19) se resumen otros problemas recurrentes en la desambiguación léxica automática como el papel del contexto en tales sistemas y la manera de evaluar los sistemas de desambiguación léxica automática. La elección manual de sentidos de palabras ambiguas también implica una dificultad en la desambiguación automática. Este último problema implica que no hay una manera consistente de dividir los sentidos de una palabra.

Otro problema que se encuentra a menudo en los sistemas de desambiguación léxica automática es la dificultad de superar el umbral del sentido más frecuente (MFS). Un sistema basado en el valor MFS no distingue entre los diferentes significados de una palabra. En cambio, un tal sistema clasifica cada forma de la palabra según el significado más común de ella. Por ejemplo, en la lengua española el significado más común del sustantivo *espejo* se describe como siguiente por el diccionario Clave (s. f.): <*Lámina de vidrio cubierta de mercurio por la parte posterior que se refleje en ella lo que hay delante*>. Además, existe otro significado (Clave, s. f.): <*Lo que se toma como modelo o es digno de imitación*> (véase el apéndice). En consecuencia, un sistema que se base únicamente en el valor MFS asignará la palabra *espejo* en la siguiente frase al primer significado: <*Esa mujer es un espejo de bondad*> (Clave, s. f.). Dado que el significado más común es efectivamente correcto en muchos casos, el valor MFS es difícil de superar. Por lo general, es más fácil para los sistemas supervisados obtener

mejores resultados, mientras que los sistemas no supervisados no consiguen superar este umbral. Sin embargo, el problema suele surgir en el caso de que el valor de MFS sea alto, es decir, en el caso de que el segundo significado de una palabra aparece raramente en los corpus (y en general se usa raramente en una lengua específica). En los casos que el valor MFS es bajo, también es más fácil para un sistema supervisado de superarlo.

### 3.2 EL MODELO DE PROCESAMIENTO DE LENGUA BERT

Devlin et al. (2019) ha introducido un nuevo modelo de lengua pre-entrenado, BERT (Bidirectional Encoder Representations from Transformers), que puede ser utilizado para la ejecución de varias tareas de idioma. Para la desambiguación de significados, BERT utiliza el contexto directo de las palabras.

Cabe mencionar que uno de los desafíos más importantes de sistemas anteriores basados en “*word embeddings*” fue la ignorancia del hecho de que varios contextos pueden generar diferentes significados de una palabra. Al contrario, BERT es un modelo de procesamiento contextualizado, lo que significa que es capaz de calcular de una manera más adecuada los significados de palabras. Además, modelos como BERT pueden ser integrados fácilmente en sistemas de procesamiento del lenguaje natural (NLP).

BERT utiliza un transformador para calcular una representación modificada de una palabra en función del contexto. Esto significa que el texto dado como entrada en BERT se convierte en vectores que permiten realizar otros cálculos. Los vectores, también llamados “*word embeddings*”, representan el significado de una palabra. Inicialmente, los significados de las palabras se presentaban de forma estática, lo que significaba que no se tenía en cuenta el detalle de que diferentes contextos pueden hacer surgir un significado diferente de una palabra polisémica. BERT es una gran innovación precisamente porque puede calcular significados contextualizados, que a su vez generan representaciones dinámicas de los significados. Esto significa que para cada instancia o uso de una palabra se calcula un vector diferente.

Antes de la introducción de BERT los textos se analizaron efectivamente en un orden determinado, lo que no permitía al modelo de procesamiento aprender el contexto alrededor de una palabra específica. Para calcular el contexto se utiliza un elemento bidireccional que hace posible que el texto se analice simultáneamente y no estrictamente de la derecha a la izquierda y viceversa. (Devlin et al., 2019, p.2)

Concretamente, en la presente tesina se utiliza BERT para calcular los vectores de los significados desde frases semilla (*seed sentences*). Los valores de los diferentes vectores varían desde 0, que no representa ninguna similitud con el vector de la palabra en la frase semilla, a 1, que corresponde con una similitud completa.

## 4. METODOLOGÍA

Para la primera parte de la investigación de esta tesina se ha identificado una serie de palabras que son ambiguas en el sentido de que tienen al menos un significado literal (1) y metafórico (2).

- (1) *amasar*: Referido a una sustancia, hacer una masa con ella mezclándola con otros elementos y con algún líquido.

*<Para hacer pan hay que amasar harina, agua y levadura.>* (Clave, s. f.)

- (2) *amasar*: Referido esp. al dinero o a los bienes, reunirlos o acumularlos.

*<Los negocios le permitieron amasar una gran fortuna.>* (Clave, s. f.)

La identificación de esta serie de palabras se ha desarrollado de una manera casi arbitraria, utilizando criterios de frecuencia basados en SCAP. Más precisamente, para incluir las palabras en la muestra, solo se ha tenido en cuenta la frecuencia en los corpus en SCAP, para tener la certeza de disponer de suficientes ejemplos. También se ha vigilado por incluir palabras de diferentes categorías morfosintácticas (sustantivos, adjetivos y verbos). (Véase el apartado 4.1)

Después de la selección de palabras ambiguas, se ha redactado una ficha de significados y usos que parte de los contenidos del diccionario Clave (s. f.). Se ha sido utilizado el diccionario para establecer las definiciones concretas y las frases de ejemplo de los distintos significados de

una palabra. Esta información se ha utilizado para la anotación manual de cada palabra y sus diferentes significados (véanse los apartados 4.2 y 4.3).

Para la evaluación del sistema de identificación automática de palabras ambiguas, se han anotado manualmente *test sets* para cada palabra. Estas anotaciones manuales se han utilizado para establecer la diferencia entre los resultados generados por el sistema y las frases anotadas manualmente (véase el apartado 4.4).

También se ha realizado una segunda anotación manual para determinar si esto mejoraba los resultados de la primera fase de identificación. Para este fin, se anotaron frases en las que el sistema calculó una diferencia muy pequeña entre el significado literal y el metafórico. Es decir, se anotaron las frases en las que el sistema "dudaba" más (véase el apartado 4.5).

El propósito de este estudio es evaluar la precisión de un modelo automatizado de reconocimiento de usos. En concreto, el umbral que debe superar el modelo es la frecuencia del sentido más frecuente de la palabra (MFS).

#### **4.1 ELECCIÓN DE PALABRAS AMBIGUAS EN SCAP**

El *Spanish Corpus Annotation Project* (SCAP) es una herramienta de aprendizaje de vocabulario español que utiliza el Natural Language Processing (NLP) y metodologías del Análisis del Corpus (CA). (Goethals, 2018, p.1)

SCAP ofrece la posibilidad de seleccionar elementos específicos de vocabulario teniendo en cuenta el nivel de competencia lingüística española, las palabras relacionadas entre el español y la lengua materna del estudiante y especificidad semántica. Esta manera de aprendizaje de vocabulario también mejora la autonomía en el trayecto de aprendizaje, ya que SCAP también ofrece la posibilidad de elegir el vocabulario en corpus de dominios específicos (turismo, economía, literatura, etc.). Sin embargo, SCAP también tiene algunas deficiencias, y más concretamente en relación con la desambiguación léxica automática. El problema consiste en que, si la herramienta se utiliza para buscar palabras y ejemplos de construcciones con estas palabras, SCAP no distingue entre los diferentes significados de la palabra. Si se busca la

palabra *amasar* en los corpus, aparecerán todas las frases que contienen cualquier significado de *amasar*.

En el marco de este trabajo, SCAP se ha utilizado para seleccionar las palabras ambiguas. El único criterio delimitador ha sido el de la frecuencia en un corpus de noticias de Sanidad. El dominio en sí no tiene más relevancia para el estudio. Esta etapa se aplicó para acortar el tiempo dedicado a la selección de datos.

Para limitar la búsqueda aún más, se aplicó el filtro del nivel de aprendizaje. En la presente tesina solamente se han escogido sustantivos, adjetivos y verbos que pertenecen a los niveles más básicos (A).

## 4.2 SIGNIFICADOS DE CLAVE

Para el segundo paso de la investigación de esta tesina se ha recurrido a la versión digital del diccionario de uso del español actual, Clave (s. f.). En concreto, se utilizó el diccionario para obtener definiciones de los lemas que pudieran formar parte de la investigación. De este modo, también se estableció si la palabra encontrada en SCAP tenía un significado metafórico. Solamente las palabras que tenían al menos un significado metafórico, y al menos un significado literal se han elegido para esta investigación. Al final de esta etapa se habían elegido 20 sustantivos, 20 adjetivos y 20 verbos.

| Verbos     | Sustantivos | Adjetivos |
|------------|-------------|-----------|
| Amasar     | Basura      | Alegre    |
| Aterrizar  | Cabeza      | Ancho     |
| Atropellar | Calor       | Caliente  |
| Brillar    | Camino      | Cercano   |
| Cantar     | Cáncer      | Doble     |
| Desembocar | Cielo       | Dulce     |
| Disparar   | Droga       | Elástico  |
| Frenar     | Espejo      | Enorme    |
| Fumar      | Fiebre      | Estrecho  |

|           |          |            |
|-----------|----------|------------|
| Funcionar | Gota     | Fantástico |
| Helar     | Horno    | Feo        |
| Llorar    | Hueso    | Fresco     |
| Mojar     | Luz      | Frío       |
| Morder    | Mano     | Horrible   |
| Partir    | Miel     | Maduro     |
| Pesar     | Niebla   | Negro      |
| Querer    | Página   | Rico       |
| Sufrir    | Paraguas | Rubio      |
| Vencer    | Salud    | Santo      |
| Volar     | Vida     | Verde      |

*Tabla (1): Lista de lemas elegidos de SCAP después de haber consultado el diccionario Clave*

Tras elegir las palabras ambiguas, algunas de ellas se descartaron por las siguientes razones:

1. La palabra se divide en muchas significaciones en el diccionario. Este es el caso, por ejemplo, del verbo *volar*, que tiene 11 significados según el diccionario. Además, los significados de estas palabras suelen ser muy difíciles de distinguir entre sí y requieren mucho tiempo de anotación.
2. Una de las acepciones aparece poco (o ninguna vez) en los corpus de SCAP. En consecuencia, este tipo de palabras no son relevantes para la investigación.

Estos pasos han reducido la lista a unos 36 lemas.

#### **4.3 ANOTACIÓN MANUAL DE LOS SIGNIFICADOS**

Para cada lema se ha anotado manualmente una ficha con los significados literales y metafóricos. Para ilustrar esta etapa de la investigación, se muestra el ejemplo de la anotación manual del verbo *amasar*.

```

list_of_main_senses

main_sense → lit_vs_metaph → 1
ES → hacer una masa con sustancias y con algún líquido
NL → vergaren
label → literal
end_of_main_sense

main_sense → lit_vs_metaph → 2
ES → acumular
NL → verzamelen
label → metafórico
end_of_main_sense

list_of_subsenses

MWE > no
definition → Referido a una sustancia, hacer una masa con ella mezclándola con otros elementos y con algún líquido → Clave
example_sent → Para hacer pan hay que amasar harina, agua y levadura. → 5 → Clave
classification_main_sense → lit_vs_metaph → 1
end_of_subsense

MWE > no
definition → Referido esp. al dinero o a los bienes, reunirlos o acumularlos → Clave
example_sent → Los negocios le permitieron amasar una gran fortuna. → 4 → Clave
classification_main_sense → lit_vs_metaph → 2
end_of_subsense

```

Figura (2): Anotación manual del verbo 'amasar'

Cada ficha con las anotaciones manuales consiste en las siguientes partes:

1. Una lista de los sentidos de la palabra.

Para este paso, se ha hecho una distinción entre cada significado literal y metafórico de la palabra. Significados muy similares en el diccionario Clave (s. f.) se han considerado como un significado para facilitar la tarea de anotación y para obtener resultados más precisos. También se han excluido significados que solamente se utilizan regionalmente o que resultan muy poco frecuentes en los corpus en SCAP. Finalmente, se ha añadido a cada significado un etiquetado y un número para distinguir entre los diferentes significados.

2. Una lista con definiciones y frases semilla para cada significado.

Cada definición se ha mencionado tal y como aparece en Clave (s. f.). Si no se ha presentado ningún significado en el diccionario, se han utilizado otras fuentes de Internet. Además, se han añadido frases de ejemplo para cada significado. Estas frases sirven como datos de aprendizaje para el sistema y también aclaran el contexto en el que se utiliza el significado.

Después de la anotación manual, las 36 fichas han sido utilizadas para una primera prueba. El objetivo de esta prueba era comprobar si los diferentes significados de todas las palabras



aparecían suficientes veces en los corpus de SCAP para poder realizar investigaciones relevantes con ellos. Se encontró que los significados metafóricos de 6 lemas no se hallaban (casi) ninguna vez en los corpus de SCAP. Por ejemplo, la palabra *niebla* tiene dos significados en el diccionario Clave (s. f.):

1. Acumulación de nubes en contacto con la superficie terrestre. (significado literal)
2. Lo que dificulta el conocimiento de un asunto o su comprensión. (significado metafórico)

Sin embargo, el significado metafórico no aparece ninguna vez en los corpus de SCAP, por lo que la palabra tuvo que ser omitida. Este fue también el caso de las palabras *cantar* y *morder*. Además, también se han revisado los lemas en función de la frecuencia de cada significado en SCAP, lo que significa que existían casos dudosos para algunos lemas. Para ello, se han utilizado fichas en las que se han enumerado las frases más parecidas por significado. Si para un significado había pocas frases en las que apareciera el lema con su significado correcto, la probabilidad de que este significado no aparezca o apenas aparezca en el corpus SCAP era muy alta. Como consecuencia, se han omitido algunos significados, o en algunos casos, también lemas.

La figura siguiente ilustra algunas frases pertenecientes a los dos significados del adjetivo *ancho*. Como las frases son correctas en la mayoría de los casos, se ha mantenido el lema para la investigación.

| 1 con más espacio del necesario |      |                                                                                                                                                        |
|---------------------------------|------|--------------------------------------------------------------------------------------------------------------------------------------------------------|
| bejur_061542075[21, 21]         | 0.9  | Dejó caer todo su peso sobre la manivela hasta que la ranura entre las dos hojas del portón fue lo bastante ancha como para colarse entre ellas .      |
| binor_339549575[5, 5]           | 0.88 | La calzada era lo bastante ancha para habilitar dos carriles .                                                                                         |
| binor_175512235[8, 8]           | 0.88 | Su culo no tardaría mucho en ser tan ancho como largo es el mango de un hacha .                                                                        |
| bijurc_042522745[15, 15]        | 0.89 | Luego aparece su cara , asomando por el agujero , que apenas es lo bastante ancho para sus hombros .                                                   |
| binor_1865665[7, 7]             | 0.89 | Los hombros de Einar no eran más anchos que los de un muchacho .                                                                                       |
| bijurc_221599145[5, 5]          | 0.89 | Esta suddera es lo bastante ancha como para no se pueda distinguir la diferencia en la distancia .                                                     |
| bijurc_194554115[10, 10]        | 0.88 | De punta a punta , la cantera era casi tan ancha como la longitud de un estadio de fútbol .                                                            |
| bijurc_319528755[5, 5]          | 0.88 | La cabeza era casi tan ancha como la cintura del chico .                                                                                               |
| binor_303587545[19, 19]         | 0.88 | Un tipo de más de un metro noventa que pesaba más de noventa kilos , con los hombros tan anchos como él .                                              |
| bijurc_018549665[5, 5]          | 0.87 | Su sonrisa se hizo más ancha , y en sus ojos negros apareció un destello .                                                                             |
| bijurc_1145115265[11, 11]       | 0.89 | Con el tamaño que tenían ahora , el pasillo parecía tan ancho como un salón de baile .                                                                 |
| bijurc_138511895[5, 5]          | 0.89 | El tronco era lo suficientemente ancho como para ocultarle .                                                                                           |
| benor_19952265[12, 12]          | 0.87 | Me pasó por la cabeza que su silueta era el doble de ancha que cuando fue novio de mi hermana .                                                        |
| bijurc_260573855[18, 18]        | 0.87 | Desde la punta del burl al puente de la nariz de Zeus se abrió una hendidura lo bastante ancha para que saliera Atenea .                               |
| binor_0285137115[17, 17]        | 0.87 | En otro rincón de la celda , en el suelo , había una delgada colchoneta lo bastante ancha para tres personas apretadas .                               |
| binor_01053275[6, 6]            | 0.87 | Las caderas parecían un poco más anchas que antes , algo bastante natural después de dar a luz .                                                       |
| binor_1175103205[7, 7]          | 0.89 | Las ranuras del suelo son lo suficientemente anchas como para que la diarrea se cuele por ellas .                                                      |
| binor_234590135[10, 10]         | 0.87 | Sus ojos parecían más pequeños porque la mandíbula era más ancha .                                                                                     |
| binor_390534905[5, 5]           | 0.87 | Sus muñecas apenas eran más anchas que una caja de cerillas - .                                                                                        |
| bijurc_070512915[5, 5]          | 0.87 | La Arista es mucho más ancha y también más larga .                                                                                                     |
| benor_520544855[4, 4]           | 0.86 | El cuello era más ancho que la cabeza y las orejas más pequeñas que los ojos .                                                                         |
| benor_221541875[11, 11]         | 0.87 | Empieza con la segunda hoja , cuyo borde marchito es más ancho que el de la primera .                                                                  |
| binor_4165190225[6, 6]          | 0.86 | El cuello del entrenador era más ancho que su cabeza , y la cabeza era tan grande que los ojos parecían dos minúsculas mirillas .                      |
| bijurc_176537935[6, 6]          | 0.86 | La escalera era apenas lo bastante ancha para permitir el paso de una persona .                                                                        |
| binor_39756515[5, 5]            | 0.88 | El callejón es lo bastante ancho como para que pasen quizá dos coches pequeños , pero nada más .                                                       |
| 2 orgulloso                     |      |                                                                                                                                                        |
| binor_03256945[11, 11]          | 0.94 | No entiendo cómo podéis haberos enterado de esto y quedaros tan anchos .                                                                               |
| bijurc_1115112215[24, 24]       | 0.89 | La cosa no funcionó tan bien con Simon , que sabía que no me dedicaba a dominar a gente muerta y me quedaba tan ancha .                                |
| binor_2765141515[13, 13]        | 0.89 | ¿ Es que van a consentir que confiese y luego se quede tan ancho ?                                                                                     |
| bejur_069530135[13, 13]         | 0.89 | ¿ Me suelta toda esa mierda y ya está , se queda tan ancho ?                                                                                           |
| binor_323557405[22, 22]         | 0.86 | No puedes ganar todas las batallas , no puedes salirte siempre con la tuya , no puedes volverme loco y quedarte tan ancho » .                          |
| bijurc_025529265[17, 17]        | 0.87 | ¿ Qué clase de padre echa a su hijo de casa sin más y se queda tan ancho ?                                                                             |
| dreyin_035755[21, 21]           | 0.84 | Nuestra España es la misma que la de nuestros hijos , y ellos y nosotros la deseamos justa , dialogante y ancha .                                      |
| benor_049513875[17, 17]         | 0.88 | A este paso cualquier día dirás que Kim Novak era de la Atlántida y te quedarás tan ancho .                                                            |
| pjudi_49369595[11, 11]          | 0.81 | Y hacen esa calumniosa afirmación sin aportar ninguna prueba quedándose tan anchos y encantados de su magnífica ocurrencia .                           |
| pjudi_463915305[19, 19]         | 0.86 | El líder del Atlético acaba de declarar que la Liga española es una estafa y se ha quedado tan ancho .                                                 |
| psani_007945125[16, 16]         | 0.86 | Si por eso nos van a acusar ustedes de peseteros , pues háganlo y quédense tan anchos .                                                                |
| pviaj_00683565[22, 22]          | 0.8  | Para la historia del arte , aquel estio suponía el avance de algunos pasos en su carrera de fondo , universal y ancha .                                |
| dreyin_045355[126, 26]          | 0.82 | No es un pueblo fatigado e inerme el que esta noche , al finalizar un año , especialmente comprometido , espera un mañana mejor y más ancho para todos |
| pviaj_022025595[9, 9]           | 0.78 | Un lugar cerrado pero abierto , tan alto como ancho , vacío pero lleno de gente .                                                                      |
| benor_44958185[17, 17]          | 0.76 | Un hombre gordo , sonriente , con barba de tres días y bigotes como los suyos , anchos y descuidados como los que usaba entonces .                     |
| benor_471552875[8, 8]           | 0.79 | Su vida le parecía monótona , un tedio ancho y hondo se instalaba en su corazón .                                                                      |
| pviaj_009835265[18, 18]         | 0.78 | En el Moma , uno encuentra a alguien que conoció en Dublín , el mundo cierto que es ancho y ajeno , pero su metrópoli tiene algo de aldea .            |
| binor_06651695[8, 8]            | 0.78 | Me mira y yo le dedico una sonrisa ancha , alentadora , de buen padre .                                                                                |
| benor_453529155[21, 21]         | 0.8  | Si pudiera ver a su nieto , a su Gabriel , con esos poemas tan lindos que escribe , papá estaría ancho de orgullo .                                    |
| benor_4935143745[2, 2]          | 0.78 | Me sentía ancho como un pavo real mientras examinábamos en silencio las fotografías .                                                                  |
| benor_023544035[6, 6]           | 0.77 | Eran más bajos , oscuros , anchos y feos ; además , hablaban español .                                                                                 |
| benor_41151405[13, 13]          | 0.76 | Lo he aceptado tal cual es , y disfruto incluso con ese mundo ancho y ajeno que nos rodea .                                                            |
| binor_035554515[1, 1]           | 0.82 | Era ancha y sólida , de aluminio , y brillaba a la luz de los focos .                                                                                  |
| benor_39457035[17, 17]          | 0.84 | Al salir a la calle tuvo por primera vez la sensación de que el mundo era demasiado ancho , demasiado excitante .                                      |
| binor_146518405[9, 9]           | 0.77 | No le había parecido de los que llevan alianzas anchas .                                                                                               |

más tomar acomodó, el cliente recibe una carta con 14 aceites vírgenes-extra para mojar pan durante el aperitivo .  
 na Curra lo tiene claro : “ En momentos así de críticos , uno se tiene que mojar ” .  
 como si no le importara mojarlo , o como si la lluvia pudiera aclarar las ideas o borrar el miedo , eché a andar sin una dirección precisa .  
 asido un fiel animal , pero me temo que vas a tener que mojarlo bastante , y debo pedirte que me disculpes por usarlo así .  
 Se tiene que mojar con todas la consecuencias ” , insistió .  
 a mojar y a mí no ... , no me molestó el olor , de verdad .  
 Y encima no llegaste a mojar , hay que joderse ” , dijo un poco divertido por mi mala pata .  
 o : Esperáme aquí , así no volverás a mojarlo .  
 Has sido capaz de empaparte por perseguir un simple coche y no vas a ser capaz ahora de mojarlo por disfrutar de un paisaje ?  
 e da largas , excusas , pero se ha de mojar .  
 a los videojuegos , supongo ... Le dije que se iba a mojar , y ni caso .  
 me importaba mojarlo ; siempre hay algo alegre en un chaparrón .  
 olo me he mojado un poco , eso es todo – dijo él .  
 ver [ el día anterior a la entrevista ] cayó la de Dios y me mojé mucho , pero i qué pánico le tiene la gente a mojarse !  
 os está lloviendo a mares y me he mojado de todas formas .  
 Cómo no se te ocurre descalzarte antes de mojarlo todo ?  
 e acerqué tímidamente y me senté a cierta distancia para no mojarlo de nuevo .  
 i ya ni piensa en nada más que en mojar .  
 espús de tanto tiempo , ya solo con el olor de un hombre me estaba mojado .  
 e el fondo los científicos hablan como los políticos : No se quieren mojar en sus declaraciones para no quedar con el culo al aire .  
 a empresa prefiere no mojarse , pero la raíz del problema parece ser de nuevo la falta de fondos de la Comunidad .  
 l chico volvió a sonreír ; bueno , voy a mojarlo un poco , dijo .  
 si tendrás oportunidad de meter los pies en el agua sin llegar a mojarlo del todo – había dicho Melissa , con valentía suficiente para las dos .  
 me volví para increparlo porque nos estábamos mojado como estúpidos , pero no me escuchaba , estaba en otra parte , lejos de todas las penas terrenales .  
 o parecía que fuera a dejar de llover , así que iba a tener que mojarlo de nuevo cuando su ropa ni tan siquiera había terminado de secarse .  
 a cama  
 uando me dio mate con un alfil , sin poder evitarlo , mojé el calzoncillo .  
 onnor toma un buen trago , aunque la mayor parte se cae y le moja los pantalones .  
 enia lágrimas que lo demostraban , allí , en el plato de plástico , mojado los cubitos de queso .  
 e dijo que se retire y moja las galletas en la miel .  
 eaban el café y también mojabán en él los panecillos blancos .  
 ertió la ginebra en el vaso y me mojé los labios .  
 e la botella de vino ella apenas se mojó los labios y Pablo tomó varias copas .  
 o quería comer , solo tomaba agua , y con ella , de vez en cuando , mojaba los labios secos de su hijo .  
 upongo que sabrá lo que se hace , así que mojé el cepillo de dientes en el abrevadero .  
 e los años y la falta de costumbre hicieron que derramase una copa de agua que mojé el delicado mantel .  
 xplico que después extendí una funda de almohada y mojé el mango del cepillo cincelador en salsa para parrillada .  
 ncendí uno , sus dedos húmedos mojaron el papel .  
 o probó la empanada , apenas se mojó los labios con el vino .  
 ustapichico chorro de líquido salí y mojé la mejilla de Oddo , que soltó un grito y dejó caer todos sus fardos .  
 abel se compró unas cuantas con pájaros del Norte y despachó una carta que cerró ahí mismo , mojado la goma con la lengua .  
 e quedó en la orilla pensando en Florence , y en su distracción las olas le mojaron los zapatos .  
 u elección era « natural » , un mexicano bebía tequila en cualquier contexto ; sin embargo , apenas se mojó los labios .  
 n delgado chorro amarillo se deslizó por sus mustos y mojó la catasta .  
 eaba una rayadura o un agujero para que el agua entre en el asiento . Se acababan mojado los pantalones .  
 l agua del techo traspasó la rejilla y mojó los carbones ardientes .  
 l maestro lo midió de arriba abajo y con parsimonia se mojó los labios en el coñac de su copa .  
 l meste derretida traspasó las viejas botas de montar de Spencer y notó que se le mojabán los dedos de los pies .  
 ada vez detuvo la embarcación , le dio a beber a Nate un poco de agua y le mojó con cuidado el sudoroso cabello .  
 el sabor delicioso de la orina le mojó los mustos .  
 l sudor doró por mi pelo y mojé el trapo con el que lo había tapado .

ció en el caso del verbo *mojar* (figura (4)). Para el significado metafórico *a'*, casi no se hallaban frases correctas. En consecuencia, este lema no se definitiva.

En consecuencia, se han eliminado algunos lemas, lo que ha resultado en una lista definitiva de 29 lemas incluyendo 9 verbos, 8 adjetivos y 12 sustantivos:

| VERBOS     | SUSTANTIVOS | ADJETIVOS |
|------------|-------------|-----------|
| Amasar     | Basura      | Ancho     |
| Brillar    | Cabeza      | Bajo      |
| Desembocar | Calor       | Caliente  |

|          |          |          |
|----------|----------|----------|
| Disparar | Camino   | Estrecho |
| Frenar   | Cielo    | Fresco   |
| Llorar   | Espejo   | Frío     |
| Partir   | Fiebre   | Largo    |
| Sufrir   | Horno    | Negro    |
| Vencer   | Luz      |          |
|          | Miel     |          |
|          | Paraguas |          |
|          | Salud    |          |

Tabla (2): Lista definitiva de lemas después los pasos de filtración

#### 4.4 ANOTACIÓN MANUAL DE LOS TEST SETS

El siguiente paso ha consistido en anotar manualmente un conjunto de pruebas para cada lema. Se han seleccionado al azar unas cien frases para cada lema. A continuación, se ha añadido manualmente si la palabra de la frase tenía un significado metafórico o literal. Estas anotaciones se han utilizado para la parte de clasificación supervisada y semi-supervisada.

#### 4.5 ANOTACIONES ADICIONALES

Tras obtener los resultados basados en la primera anotación, se ha realizado una segunda anotación manual sobre un número de frases para cada lema. Para ello, se han utilizado frases en las que el sistema ha calculado una diferencia muy pequeña entre el significado literal y el metafórico de una palabra dentro de una frase determinada. Estas anotaciones se han realizado con la esperanza de limitar las "dudas" del sistema y mejorar los resultados.

Para ilustrar este paso, se vuelve a utilizar el ejemplo del verbo *amasar*. El sistema ha calculado una diferencia del valor F muy pequeña entre los significados de *amasar* para la siguiente frase:

*<No amaso dólares en Central Park West como algunos, pero me arreglo.>*

De hecho, según el sistema, la diferencia entre el significado literal y el metafórico era igual a 0,0012.

Esta frase se ha etiquetado manualmente con el significado correcto (metafórico) y se ha introducido de nuevo en el sistema como frase de ejemplo.

## 5. RESULTADOS Y ANÁLISIS DE ERRORES

En esta parte de la tesina se presentan los resultados más importantes de la investigación sobre la clasificación de los significados literales y metafóricos de los lexemas. Todos los resultados se basan en el MFS y en el valor F que el sistema ha calculado para cada significado. Esto significa que todos los resultados que, utilizando las anotaciones manuales, representan una mejor puntuación que el MFS, son interesantes en la presente investigación.

El sistema aprende la desambiguación de las palabras en varios pasos. Primero, el sistema se dota con las fichas anotadas manualmente. Para esta parte de la clasificación, el sistema utiliza solamente las frases de semilla (*seed sentences*). Por lo tanto, los resultados obtenidos en esta parte se etiquetan como “resultados supervisados”, ya que la tarea de clasificación está totalmente supervisada.

Además, se presentan los resultados de la tarea de clasificación semi-supervisada. La parte semi-supervisada consiste concretamente en dos etapas. En primer lugar, se han añadido automáticamente algunas frases a la parte supervisada para igualar el número de frases por cada significado. Esto significa que, si para el significado literal de una palabra se han anotado dos frases ejemplos y para el significado metafórico solo una frase, en esta etapa se añade automáticamente una frase de ejemplo al significado metafórico. En el resto la presentación de los resultados, este paso se denomina "igualar el número de frases". La segunda etapa de la parte semi-supervisada consiste en añadir de nuevo automáticamente algunas frases para que el sistema pueda aprender autónomamente (“añadir frases”).

Sin embargo, como se ha comentado en el apartado 4.5, se han efectuado anotaciones adicionales después de la obtención de los resultados que han salido de la etapa supervisada y semi-supervisada presentada anteriormente. Al añadir esta anotación, se ha intentado eliminar los casos dudosos en la clasificación por el sistema. De hecho, se ha añadido una pequeña sección supervisada entre el paso de igualar el número de frases y el de añadir frases automáticamente. La sección supervisada consiste en las frases para las que el sistema ha

calculado una diferencia mínima entre el significado literal y el metafórico. En la presentación de los resultados esta parte se denomina “añadir las anotaciones adicionales”.

Por lo tanto, los resultados del estudio se analizan en dos partes. En primer lugar, se presentan los resultados globales de la parte en la que no se utilizaron anotaciones adicionales. A continuación, se presentan también los resultados de la segunda parte del estudio en la que se han utilizado las anotaciones adicionales. Además, los resultados se comparan con el valor del sentido más frecuente (MFS).

## 5.1 SUSTANTIVOS

El valor del MFS de los sustantivos es en general bastante alto (0,74). Lo que se desprende del valor de la media del valor F es que esta puntuación no aumentó después de la parte supervisada y semi-supervisada. Solo después de haber añadido las anotaciones adicionales el valor F ha aumentado considerablemente a 0,84.

La puntuación más alta del MFS es igual a 0,98 para la palabra *espejo* y la puntuación más baja es de 0,57 para la palabra *luz*. En general, un valor MFS elevado también significa que es más difícil desarrollar un sistema que supere el umbral. En la primera parte del experimento (sin añadir anotaciones adicionales), siete de doce lemas han sido clasificados mejor que el MFS (*basura, cabeza, camino, cielo, fiebre, miel, paraguas*). De estos siete casos, solamente tres han mostrado una diferencia considerable (*camino, fiebre, paraguas*). La diferencia entre la puntuación del MFS y la puntuación F después de la clasificación del resto de los lemas es como máximo igual a 0,08.

La tabla siguiente muestra todos los datos obtenidos después de la tarea de clasificación de los sustantivos. Se ha añadido una segunda tabla para ilustrar la mejora o la disminución del valor F después de añadir las anotaciones de las frases difíciles para el sistema. También se presentan los resultados después de la tarea de añadir automáticamente frases durante la última etapa semi-supervisada.

| LEMA        | MFS  | SUPERVISADO | SEMI-SUPERVISADO<br>(igualar el número de frases + añadir frases) |      | ANOTACIÓN ADICIONAL | AÑADIR FRASES |
|-------------|------|-------------|-------------------------------------------------------------------|------|---------------------|---------------|
| BASURA      | 0,78 | 0,77        | 0,77                                                              | 0,79 | 0,87                | 0,87          |
| CABEZA      | 0,7  | 0,72        | 0,72                                                              | 0,72 | 0,73                | 0,74          |
| CALOR       | 0,94 | 0,46        | 0,46                                                              | 0,5  | 0,83                | 0,76          |
| CAMINO      | 0,72 | 0,92        | 0,92                                                              | 0,92 | 0,91                | 0,91          |
| CIELO       | 0,83 | 0,89        | 0,89                                                              | 0,9  | 0,94                | 0,94          |
| ESPEJO      | 0,98 | 0,73        | 0,73                                                              | 0,76 | 0,97                | 0,97          |
| FIEBRE      | 0,78 | 0,9         | 0,89                                                              | 0,89 | 0,87                | 0,87          |
| HORNO       | 0,92 | 0,71        | 0,72                                                              | 0,73 | 0,89                | 0,89          |
| LUZ         | 0,57 | 0,57        | 0,5                                                               | 0,47 | 0,57                | 0,6           |
| MIEL        | 0,66 | 0,71        | 0,71                                                              | 0,71 | 0,71                | 0,71          |
| PARAGUAS    | 0,87 | 0,98        | 0,98                                                              | 0,98 | 0,95                | 0,97          |
| SALUD       | 0,97 | 0,91        | 0,71                                                              | 0,55 | 0,95                | 0,92          |
| VALOR MEDIO | 0,74 | 0,76        | 0,75                                                              | 0,74 | 0,84                | 0,84          |

Tabla (3): Datos obtenidos de la tarea de clasificación de los sustantivos

El valor del MFS es bastante alto por la mayor parte de los lemas, resultando en una media de 0,74. En algunos casos (*calor, espejo, horno, salud*) el sistema no tenía casi ningún margen de error. Esto se ilustra más en detalle con el ejemplo de la palabra *salud*, que tiene un valor del MFS igual a 0,97. Un sistema que pueda superar este valor, debe ser casi perfecto. La tabla y el gráfico debajo ilustran la evolución del valor F durante las diferentes etapas de la tarea de clasificación.

|     | SUPERVISADO | IGUALAR<br>EL NÚMERO DE<br>FRASES | AÑADIR FRASES | ANOTACIÓN<br>ADICIONAL | AÑADIR FRASES |
|-----|-------------|-----------------------------------|---------------|------------------------|---------------|
| MFS | 0,97        | 0,97                              | 0,97          | 0,97                   | 0,97          |
| F   | 0,91        | 0,71                              | 0,55          | 0,95                   | 0,92          |

Tabla (4): Evolución del valor F del lexema 'salud'

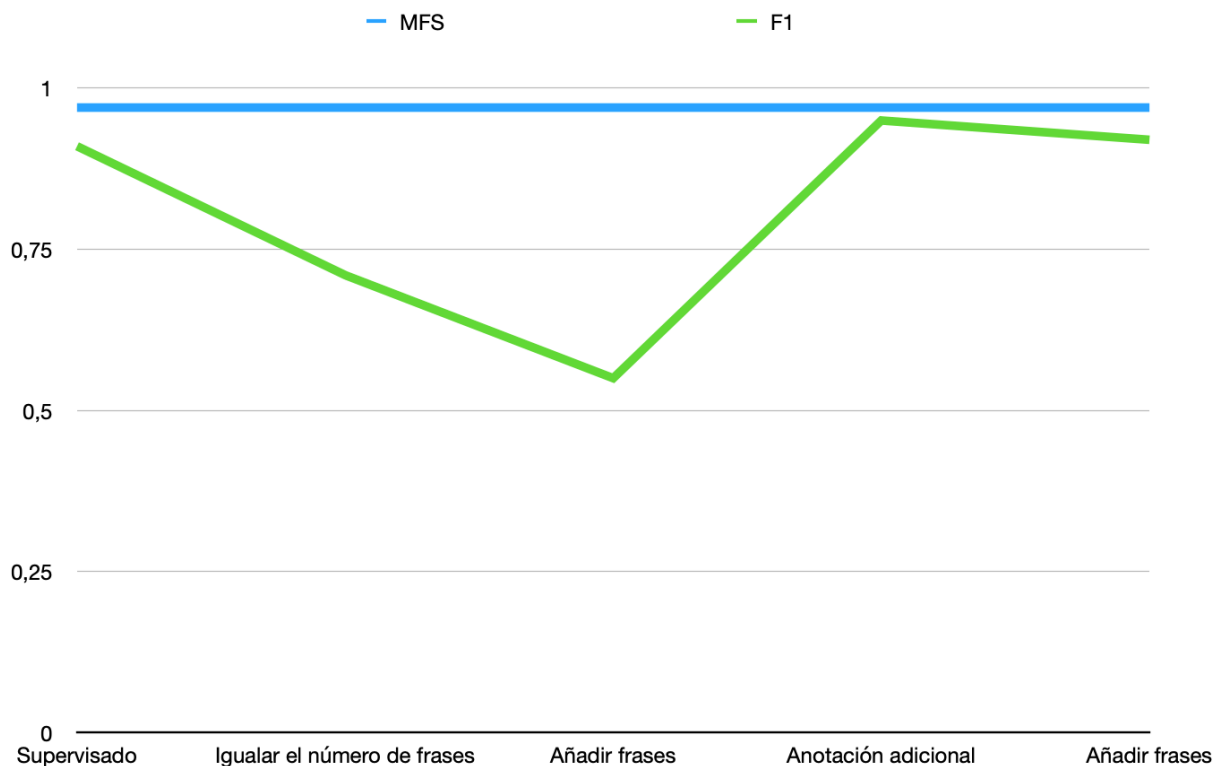


Figura (5): Evolución del valor F del lexema 'salud'

Como muestra la figura (5), la puntuación F empeora durante la primera parte del experimento y pasa de una puntuación alta, que es igual a 0,91 (tras la etapa supervisada), a una puntuación bastante baja igual a 0,55 (tras la etapa semi-supervisada). Se trata de una diferencia de 0,42, por lo que se puede concluir que el sistema considera la clasificación de esta palabra como difícil.

Un breve análisis de errores muestra que el sistema tiene dificultades para identificar el significado literal de la palabra, y considera que el significado es metafórico en muchos contextos. Para hablar de ello con más detalle, se enumeran los significados literales y metafóricos (Clave, s. f.) utilizados para la anotación manual de la ficha de la palabra *salud*. Véase el apéndice para consultar una clasificación más detallada de significados y las frases de ejemplo para cada significado.

1. Significado literal: estado en el que un organismo vivo realiza normalmente todas sus funciones.

*<Contaminar el medio ambiente es un delito contra la salud pública.>*

2. Significado literal: condiciones físicas en que se encuentra el organismo de un ser vivo en un determinado momento.

*<Desde hace algún tiempo su salud no es buena porque padece de los nervios.>*

3. Significado metafórico: buen estado o buen funcionamiento de algo.

*<Nadie duda de la salud de la economía de este país.>*

El sistema no ha clasificado incorrectamente ninguna frase en la que aparece el tercer significado. Esto se puede deber en gran parte al hecho de que el MFS es tan alto. De hecho, de estos resultados se puede destacar que el significado metafórico no es muy común en los corpus utilizados.

Sin embargo, tanto en la parte supervisada como en la semi-supervisada, el sistema ha cometido errores en la clasificación correcta de los dos sentidos literales. Esto se debería a que la diferencia entre los dos significados literales es bastante abstracta en este caso. Los contextos en los que se producen estos dos significados pueden, en muchos casos, ser muy similares, lo que puede generar dudas y dificultades durante la clasificación automática, pero también manual. Es decir que, durante las anotaciones manuales, se han encontrado dificultades durante la clasificación de los significados en algunos casos.

El análisis de errores también muestra claramente que el número de errores cometidos es casi cinco veces mayor después de la parte semi-supervisada que después de la parte supervisada. El patrón de error se mantuvo: no se ha cometido ningún otro error que el de clasificar una frase en la que aparece el primer significado como una frase en la que aparece el segundo significado. Esto demuestra que el sistema podría haber obtenido un mejor resultado si los dos primeros significados se hubieran considerado como uno solo al anotar manualmente las fichas. La siguiente tabla muestra algunos ejemplos de frases clasificadas incorrectamente.

| SIGNIFICADO<br>CORRECTO | SIGNIFICADO<br>PREDICHO | FRASE                                                                                           |
|-------------------------|-------------------------|-------------------------------------------------------------------------------------------------|
| 1                       | 2                       | Que en la <i>salud</i> influye más el código postal que el código genético es algo bien sabido. |



|   |   |                                                                                                                                                                    |
|---|---|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1 | 2 | En la calle Pintores, frente al centro de <i>salud</i> , ha desaparecido una franja de espacio verde de casi 2.000 metros cuadrados.                               |
| 1 | 2 | El sistema de <i>salud</i> estadounidense, con sus millones de personas desatendidas, escandaliza al resto del mundo, pero no tanto a los propios norteamericanos. |
| 1 | 2 | “Dinamitar el sistema de <i>salud</i> español supone un retroceso de 30 años en el VIH y el sida”.                                                                 |

Tabla (5): Análisis de errores del lexema ‘salud’

Sin embargo, hay una clara diferencia en los resultados tras las anotaciones adicionales de las frases sobre las que el sistema ha calculado pequeñas diferencias entre los diferentes significados. Este paso ha consistido específicamente en anotar solo seis frases. La tabla debajo demuestra algunas frases con los vectores calculados por cada significado literal de la palabra *salud*. La diferencia entre los dos vectores calculados por la primera frase es mínima (igual a 0,0001), lo que dificulta que el sistema clasifique el significado de la palabra *salud* en la categoría correcta. En otros casos, como en la segunda frase, el sistema ha llegado a calcular una diferencia igual a 0.

| SIGNIFICADO 1 | SIGNIFICADO 2 | FRASE                                                                                                                         |
|---------------|---------------|-------------------------------------------------------------------------------------------------------------------------------|
| 0,776         | 0,7759        | Un servicio que representa el 16% del gasto total de la Generalitat en la contratación privada de servicios de <i>salud</i> . |
| 0,8405        | 0,8450        | En este caso, el derecho a la salud del menor se impone a una posible autorización dada por los adultos.                      |

Tabla (6): Frases dudosas del lexema ‘salud’

En lugar de que los resultados sean cada vez peores, las anotaciones de este tipo de frases han aumentado el valor F casi al mismo nivel del valor MFS. Las anotaciones adicionales han alcanzado el nivel del valor F a un nivel de precisión del 95%. El análisis de errores muestra

que el sistema sigue cometiendo los mismos errores que antes, pero lo hace en muchas menos frases.

Por ejemplo, después de las anotaciones adicionales, el sistema ha clasificado correctamente la frase *<En la calle Pintores, frente al centro de salud, ha desaparecido una franja de espacio verde de casi 2.000 metros cuadrados>*, pero sigue clasificando incorrectamente la frase *<Que en la salud influye más el código postal que el código genético es algo bien sabido>*.

Después de la última parte semi-supervisada, todavía hay una diferencia entre MFS y el valor F, pero se trata de una pequeña diferencia, que es igual a 0,05.

Además, el sistema sí ha conseguido mejorar el valor de algunos lemas con un valor MFS elevado, como el de la palabra *cielo*. El valor del MFS en este caso (0,83) es inferior al de la palabra *salud*, pero sigue siendo uno de los valores MFS más altos de los sustantivos.

En las fichas anotadas manualmente se han incluido dos significados (Clave, s. f.) de la palabra:

1. Significado literal: bóveda celeste o atmosfera.

*<Esta noche se ven más estrellas en el cielo.>*

2. Significado metafórico: paraíso.

*<Las almas justas gozan del cielo.>*

En la primera parte del experimento se observa ya una mejora de 0,06 del valor F, pero esta mejora no ha subido mucho después de la etapa semi-supervisada. El sistema ha superado el umbral del valor de MFS inmediatamente después de la primera etapa supervisada. Lo que llama la atención es que, aunque se ha superado el umbral, los resultados no han mejorado mucho en la parte semi-supervisada.

|            | SUPERVISADO | IGUALAR<br>EL NÚMERO DE<br>FRASES | AÑADIR FRASES | ANOTACIÓN<br>ADICIONAL | AÑADIR FRASES |
|------------|-------------|-----------------------------------|---------------|------------------------|---------------|
| <b>MFS</b> | 0,83        | 0,83                              | 0,83          | 0,83                   | 0,83          |
| <b>F</b>   | 0,89        | 0,89                              | 0,90          | 0,94                   | 0,94          |

Tabla (7): Evolución del valor F del lexema 'cielo'

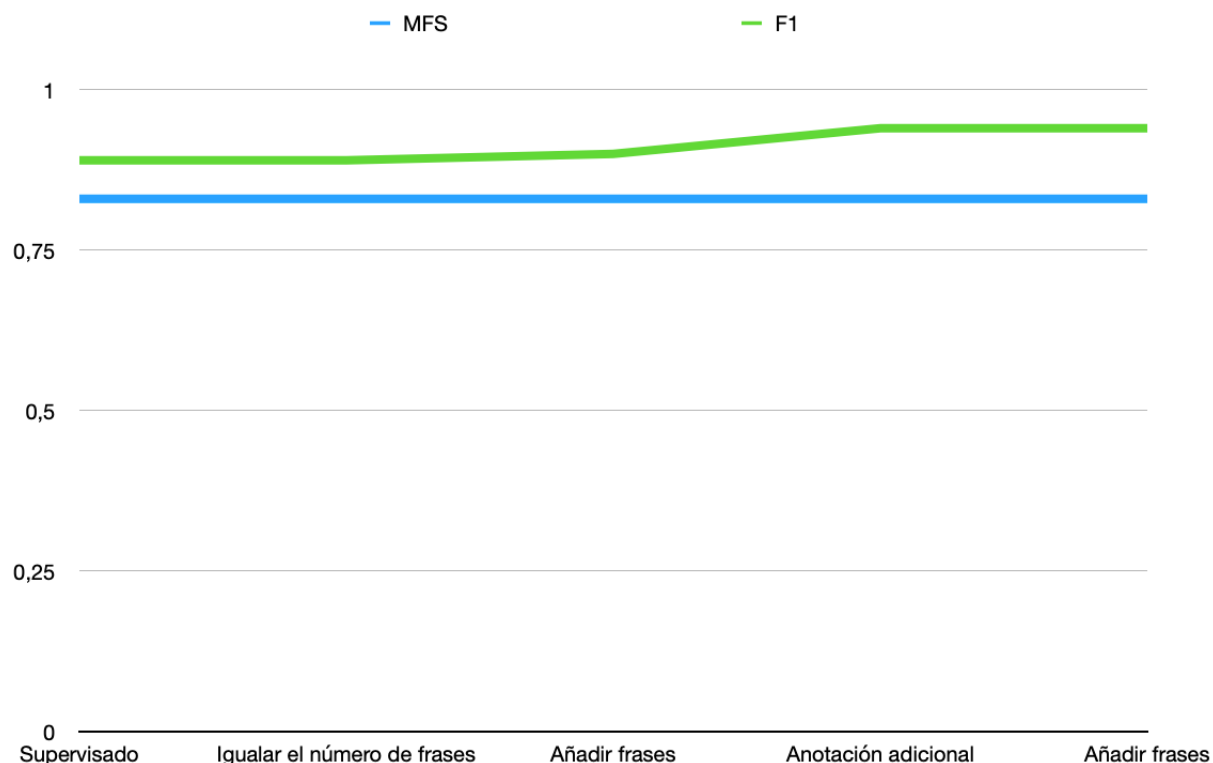


Figura (6): Evolución del valor F del lexema 'cielo'

El sistema no ha cometido muchos errores antes de las anotaciones adicionales, pero los que ha cometido se deben en gran medida a que en algunos casos las frases han sido clasificadas conforme al significado metafórico, mientras que el significado correcto de la palabra *cielo* es en realidad literal. También ha ocurrido lo contrario, pero con mucha menos frecuencia. En realidad, solo una frase fue clasificada de este modo (la tercera frase del cuadro siguiente). La siguiente tabla muestra algunos ejemplos de frases clasificadas incorrectamente.

| SIGNIFICADO CORRECTO | SIGNIFICADO PREDICHO | FRASE                                                                                   |
|----------------------|----------------------|-----------------------------------------------------------------------------------------|
| 1                    | 2                    | En dirección contraria al <i>cielo</i> viven los campesinos que trabajan estas tierras. |
| 1                    | 2                    | El horizonte era una zona fronteriza donde no había <i>cielo</i> ni tierra.             |
| 2                    | 1                    | Es obvio que han dejado de mirar al <i>cielo</i> .                                      |

Tabla (8): Análisis de errores del lexema 'cielo'

La anotación adicional de seis frases ha mejorado el valor F obtenido después de la primera parte supervisada y semi-supervisada, pero el análisis de errores muestra que se siguen cometiendo los mismos errores en algunos contextos. Al final, el sistema consiguió reducir la tasa de error en un 65% (frente al valor MFS).

*<Es obvio que han dejado de mirar al cielo>* es una frase particular que el sistema clasifica incorrectamente. A primera vista, también para los seres humanos parece una frase literal, pero al mejor observarla se puede decidir que la palabra *cielo* significa *paraíso*, *Dios* o cualquier otra denominación religiosa. Una breve búsqueda de esta frase en Internet ha demostrado que el contexto más amplio de la frase, y de la palabra *cielo*, sí tiene que ver con la religión. De hecho, la expresión *<mirar al cielo>* tiene una connotación específica y espiritual en la fe cristiana. Dado que la religión o la creencia es un concepto abstracto, incluso para el cerebro humano, este significado metafórico de la palabra *cielo* no es nada fácil de explicar. Esto también demuestra que, para algunos significados metafóricos, un solo cálculo del vector y el contexto directo en el que se encuentra esa palabra no es suficiente para que un sistema la clasifique correctamente.

## 5.2 ADJETIVOS

El valor medio del MFS para los adjetivos es igual a 0,72, que es un valor bastante alto. El valor F medio es igual a 0,74 tras la tarea de clasificación supervisada. Tras la etapa semi-supervisada, esta media se ha reducido en 0,01 y ha pasado a ser igual a 0,73. De esto se puede concluir que el valor F no superó considerablemente el valor del MFS después de la primera parte del experimento, es decir sin añadir anotaciones adicionales.

Sin embargo, tras la adición de unas anotaciones más, se observa un aumento de hasta 0,1 (0,83) respecto a la media del MFS. El valor de 0,83 se mantiene en el último paso, es decir, después de añadir frases adicionales para completar la tarea semi-supervisada.

| LEMA        | MFS  | SUPERVISADO | SEMI-SUPERVISADO<br>(igualar el número de frases + añadir frases) |      | ANOTACIÓN ADICIONAL | AÑADIR FRASES |
|-------------|------|-------------|-------------------------------------------------------------------|------|---------------------|---------------|
| ANCHO       | 1    | 1           | 1                                                                 | 1    | 1                   | 1             |
| BAJO        | 0,2  | 0,5         | 0,52                                                              | 0,51 | 0,66                | 0,69          |
| CALIENTE    | 0,9  | 0,85        | 0,87                                                              | 0,88 | 0,93                | 0,93          |
| ESTRECHO    | 0,76 | 0,79        | 0,79                                                              | 0,79 | 0,83                | 0,85          |
| FRESCO      | 0,48 | 0,69        | 0,69                                                              | 0,69 | 0,72                | 0,73          |
| FRÍO        | 0,7  | 0,72        | 0,72                                                              | 0,7  | 0,75                | 0,75          |
| LARGO       | 0,34 | 0,71        | 0,71                                                              | 0,72 | 0,79                | 0,8           |
| NEGRO       | 0,8  | 0,71        | 0,58                                                              | 0,6  | 0,84                | 0,81          |
| VALOR MEDIO | 0,72 | 0,74        | 0,73                                                              | 0,73 | 0,83                | 0,83          |

Tabla (9): Datos obtenidos de la tarea de clasificación de los adjetivos

El lema *negro* ilustra muy claramente que las anotaciones adicionales han aportado muchas mejoras. De hecho, el lema tiene un valor MFS bastante alto, lo que ya hace difícil que el sistema lo supere. En la parte supervisada y semi-supervisada del experimento, el sistema tampoco ha conseguido mejorar este valor, ya que se observa una diferencia de 0,21 entre el MFS y el nuevo valor F calculado.

|     | SUPERVISADO | IGUALAR<br>EL NÚMERO DE<br>FRASES | AÑADIR FRASES | ANOTACIÓN<br>ADICIONAL | AÑADIR FRASES |
|-----|-------------|-----------------------------------|---------------|------------------------|---------------|
| MFS | 0,8         | 0,8                               | 0,8           | 0,8                    | 0,8           |
| F   | 0,71        | 0,58                              | 0,6           | 0,84                   | 0,81          |

Tabla (10): Evolución del valor F del lexema 'negro'

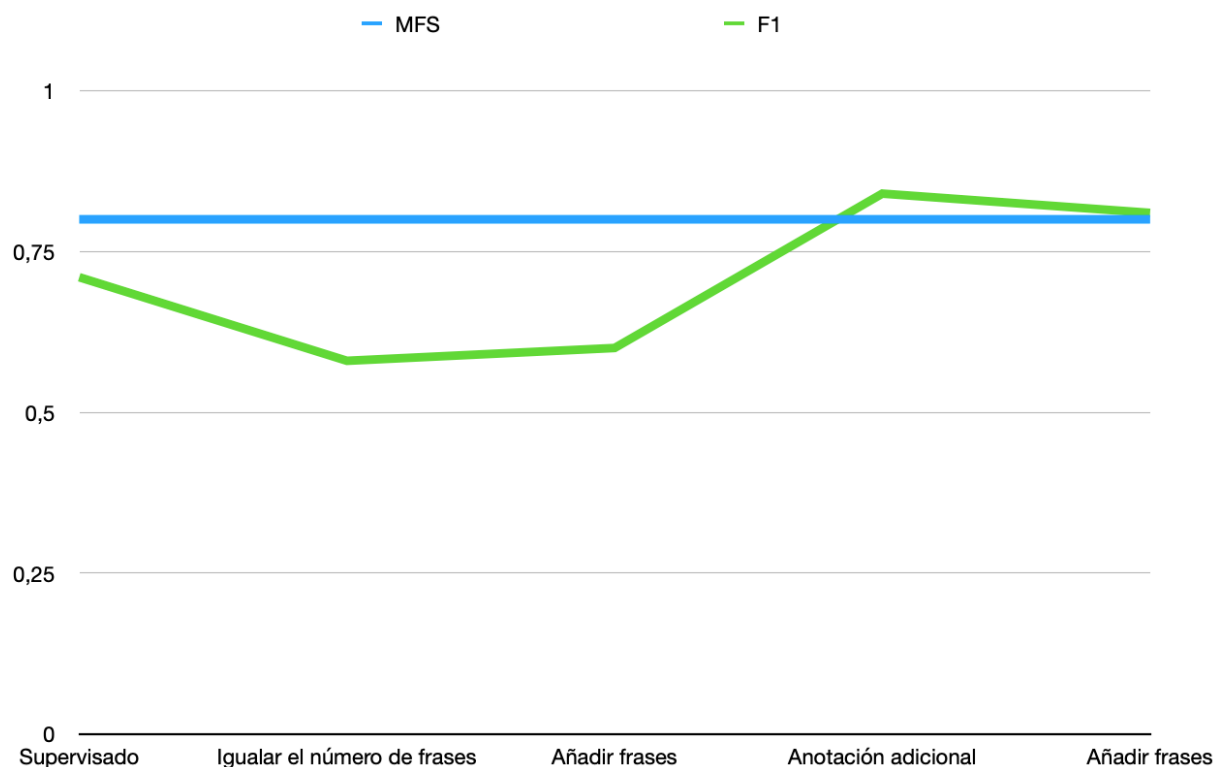


Figura (7): Evolución del valor F del lexema 'negro'

Para la parte supervisada se han anotado cuatro significados de este adjetivo según las definiciones del diccionario Clave (s. f.):

1. Significado literal: el color.  
<¿Has visto un gato negro?>
2. Significado literal: oscuro.  
<Va a caer un chaparrón, porque las nubes están negras.>
3. Significado metafórico: emoción de tristeza.  
<He tenido un día negro.>
4. Significado literal: grupo étnico.  
<Era hija de un criollo cubano y de su esclava negra, que había muerto en el parto.>

El análisis de errores muestra que durante esta etapa el sistema clasifica muchas de las palabras en la cuarta anotación (grupo étnico) mientras que la anotación correcta pertenecía al primer significado (color). La tabla debajo muestra algunas frases clasificadas erróneamente por el sistema. De las 51 frases mal clasificadas, a 45 se les ha asignado un significado incorrecto, mientras que el significado correcto en todos esos casos fue el del color.

| SIGNIFICADO<br>CORRECTO | SIGNIFICADO<br>PREDICHO | FRASE                                                                                                                                  |
|-------------------------|-------------------------|----------------------------------------------------------------------------------------------------------------------------------------|
| 1                       | 4                       | Yo vestía con pantalones cortos <i>negros</i> , una camiseta negra de tirantes y unas botas de verano desabrochadas.                   |
| 1                       | 4                       | Alguien me cubre la cabeza con una capucha <i>negra</i> y me la ata al cuello.                                                         |
| 1                       | 4                       | Está tumbado con los pantalones negros y la camisa negra; la gorra de béisbol está en la cómoda y las botas <i>negras</i> en el suelo. |
| 3                       | 4                       | Esa nube <i>negra</i> que traía con él, ese constante recordatorio de la muerte de Viruca, se ha despejado.                            |
| 2                       | 4                       | Pero consigue enderezar el rumbo y seguir avanzando en la noche <i>negra</i> estallada de refucilos.                                   |

Tabla (11): Análisis de errores del lexema 'negro'

Sin embargo, tras añadir doce anotaciones adicionales, se observa un aumento significativo del valor F y, por tanto, un fuerte descenso en el número de errores cometidos. La diferencia de los vectores calculados para el adjetivo *negro* en las frases dudosas también ha sido muy pequeña en este caso, oscilando entre 0 y un máximo de 0,0002. Después de haber eliminado las dudas mediante anotaciones manuales, el valor F se ha fijado en 0,84, lo que significa que el sistema ha superado el umbral del MFS.

El aumento del valor F también es visible en el análisis de errores, ya que el número de errores cometidos se reduce en casi 1/3. De 51 frases mal clasificadas, el sistema solo ha cometido 14 errores. El sistema clasifica correctamente más frases en las que el adjetivo *negro* se refiere al color, pero sigue cometiendo errores en los demás significados. Por ejemplo, todavía se han cometido errores en cuanto al significado metafórico *triste*. En la tabla se presentan varias frases que no han sido clasificadas correctamente por el sistema tras las anotaciones adicionales.

| SIGNIFICADO<br>CORRECTO | SIGNIFICADO<br>PREDICHO | FRASE                                                                                                                                          |
|-------------------------|-------------------------|------------------------------------------------------------------------------------------------------------------------------------------------|
| 4                       | 1                       | <i>Negros</i> es como se supone que tienes que llamarlos.                                                                                      |
| 2                       | 1                       | Muy cerca, tal vez a un metro de distancia, había una corriente de aguas <i>negras</i> .                                                       |
| 3                       | 1                       | Según dijo, de lo contrario, el futuro es "muy <i>negro</i> ".                                                                                 |
| 1                       | 2                       | Solberino regresó de nuevo al velero clavado en el arrecife, mirando de reojo las sombras <i>negras</i> y grasientas de las sirenas inquietas. |

Tabla (12): Análisis de errores del lexema 'negro'

Por otro lado, entre los adjetivos, se encuentran algunos ejemplos que ya han sido clasificados muy bien por el sistema en la parte supervisada. Estos ejemplos tienen valores MFS bajos, como por ejemplo los lexemas *largo* (0,34), *bajo* (0,2) y *fresco* (0,48). El adjetivo *bajo* tiene el valor del MFS más bajo. En consecuencia, después de la parte supervisada y después de las anotaciones adicionales, se ha producido una mejora igual a 46% en comparación con el valor del MFS.

|            | SUPERVISADO | IGUALAR<br>EL NÚMERO DE<br>FRASES | AÑADIR FRASES | ANOTACIÓN<br>ADICIONAL | AÑADIR FRASES |
|------------|-------------|-----------------------------------|---------------|------------------------|---------------|
| <b>MFS</b> | 0,2         | 0,2                               | 0,2           | 0,2                    | 0,2           |
| <b>F</b>   | 0,5         | 0,52                              | 0,51          | 0,66                   | 0,69          |

Tabla (13): Evolución del valor F del lexema 'bajo'



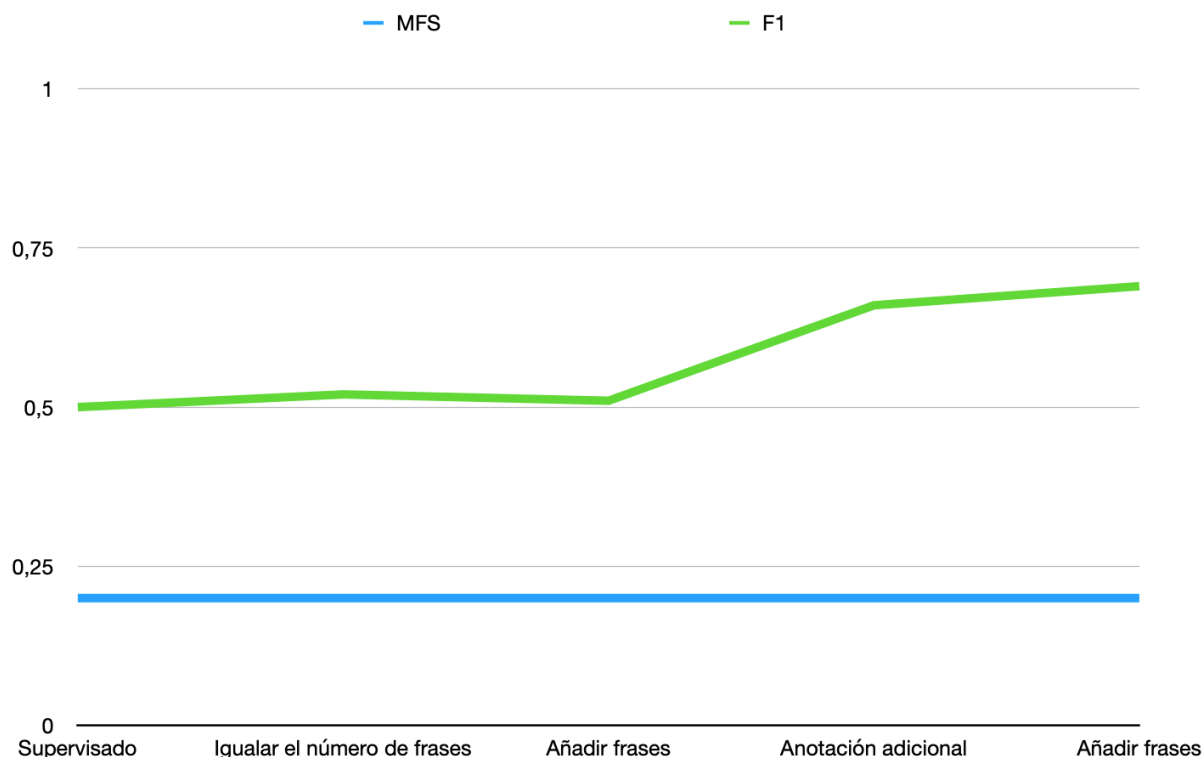


Figura (8): Evolución del valor F del lexema 'bajo'

Aunque el valor F ha mejorado mucho con respecto al MFS, sigue siendo inferior al de otros lemas. Una de las razones puede ser que *bajo* tiene muchos significados, y que se han utilizado 5 significados (Clave, s. f.) para la anotación de las fichas, respectivamente:

1. Significado literal: de poca altura.  
*<Al lado de un jugador de baloncesto pareces muy bajo.>*
2. Significado literal: a poca altura.  
*<Las tierras bajas se inundan con mucha facilidad.>*
3. Significado metafórico: inferior.  
*<En verano la presión atmosférica es más baja.>*
4. Significado metafórico: inclinado hacia abajo.  
*<Siempre me habla con los ojos bajos.>*
5. Significado metafórico: referido a un sonido.  
*<Su hermano tiene una voz baja, ronca y hueca.>*

Podría haber una correlación entre la capacidad del sistema para clasificar correctamente los lexemas y el número de significados. Si consideramos que los lexemas para los que se han

elegido más de dos significados en las anotaciones manuales son lexemas con muchos significados, se puede establecer que en el caso de los adjetivos hay tres palabras con muchos significados: *bajo* (cinco significados), *fresco* (tres significados) y *negro* (cuatro significados). Sin embargo, la idea de que la cantidad de significados por palabra afecta a la puntuación final del valor F, no puede aceptarse plenamente si se analizan estas tres palabras. Más concretamente, aunque *bajo* y *fresco* son efectivamente unos de los adjetivos con más de dos significados, no se puede determinar con certeza que sus bajas puntuaciones se deban al número de significados utilizados para las anotaciones manuales. *Negro*, por ejemplo, tiene cuatro significados, y es una de las palabras mejor clasificadas por el sistema. El valor F ha llegado incluso a 0,81. Por el contrario, *bajo* y *fresco* son los dos adjetivos con el valor F más bajo de los adjetivos, incluso después de las anotaciones adicionales.

Los errores cometidos por el sistema no están vinculados a un significado concreto para la palabra *bajo*, como ha ocurrido con los significados del adjetivo *negro*. Por lo tanto, no existe un patrón consistente que pueda aclarar dónde el sistema encontró dificultades. Tras la adición de anotaciones adicionales, el valor F ha aumentado, pero sigue siendo significativamente inferior al valor F de los otros adjetivos.

| SIGNIFICADO<br>CORRECTO | SIGNIFICADO<br>PREDICHO | FRASE                                                                                                                    |
|-------------------------|-------------------------|--------------------------------------------------------------------------------------------------------------------------|
| 1                       | 5                       | Es una mujer de mediana edad, <i>baja</i> y rechoncha, con forma de manzana, con el pelo rubio rizado con la permanente. |
| 5                       | 4                       | No con suspiros altos no con suspiros <i>bajos</i> : con suspiros a ras de suelo.                                        |
| 3                       | 5                       | “El médico me dijo que no circuncidar a mi hijo era algo sucio, de clase <i>baja</i> ”.                                  |

Tabla (14): Análisis de errores del lexema ‘bajo’

Un caso particular es el del adjetivo *ancho*. Tanto el valor MFS como el valor F a lo largo de los experimentos se mantuvieron iguales a 1. Un valor F alto indica que el sistema realiza una buena clasificación de los significados, pero como el valor MFS también fue igual a 1, significa

que no se detectó (casi) ningún significado metafórico. Al seleccionar los significados de la palabra *ancho*, también quedó claro que el significado metafórico no estaba presente (véase el apartado 4.3), excepto en la forma de la colocación *quedarse ancho*. Esta colocación aparece con bastante frecuencia en los corpus, pero no parece tener ningún valor añadido para la tarea de clasificación.

### 5.3 VERBOS

La categoría de los verbos presenta un valor medio del MFS igual a 0,71, lo que es similar al de los adjetivos. El valor F medio ha mejorado en 0,05 tras la primera parte supervisada del experimento, pero se ha reducido de nuevo a 0,73 tras la parte semi-supervisada. Las anotaciones adicionales han producido una gran mejora en este caso, y han aumentado el valor F medio a 0,84, que se ha mantenido después de añadir frases adicionales.

| LEMA           | MFS  | SUPERVISADO | SEMI-SUPERVISADO<br>(igualar el número de<br>frases + añadir frases) |      | ANOTACIÓN<br>ADICIONAL | AÑADIR<br>FRASES |
|----------------|------|-------------|----------------------------------------------------------------------|------|------------------------|------------------|
| AMASAR         | 0,33 | 0,93        | 0,93                                                                 | 0,87 | 0,76                   | 0,9              |
| BRILLAR        | 0,94 | 0,76        | 0,76                                                                 | 0,62 | 0,94                   | 0,85             |
| DESEMBOLCAR    | 0,47 | 0,43        | 0,43                                                                 | 0,4  | 0,73                   | 0,77             |
| DISPARAR       | 0,69 | 0,78        | 0,78                                                                 | 0,78 | 0,81                   | 0,81             |
| FRENAR         | 0,36 | 0,74        | 0,74                                                                 | 0,77 | 0,77                   | 0,77             |
| LORAR          | 0,98 | 0,97        | 0,95                                                                 | 0,94 | 0,96                   | 0,96             |
| PARTIR         | 0,53 | 0,91        | 0,91                                                                 | 0,9  | 0,9                    | 0,9              |
| SUFIR          | 0,99 | 0,69        | 0,48                                                                 | 0,41 | 0,88                   | 0,81             |
| VENCER         | 0,54 | 0,63        | 0,63                                                                 | 0,6  | 0,82                   | 0,79             |
| VALOR<br>MEDIO | 0,71 | 0,76        | 0,74                                                                 | 0,73 | 0,84                   | 0,84             |

Tabla (15): Datos obtenidos de la tarea de clasificación de los verbos

Esta mejora de la puntuación F gracias a las anotaciones adicionales queda bien ilustrada por el verbo *desembocar*.

|            | SUPERVISADO | IGUALAR<br>EL NÚMERO<br>DE FRASES | AÑADIR<br>FRASES | ANOTACIÓN<br>ADICIONAL | AÑADIR FRASES |
|------------|-------------|-----------------------------------|------------------|------------------------|---------------|
| <b>MFS</b> | 0,47        | 0,47                              | 0,47             | 0,47                   | 0,47          |
| <b>F</b>   | 0,43        | 0,43                              | 0,4              | 0,73                   | 0,77          |

Tabla (16): Evolución del valor F del lexema 'desembocar'

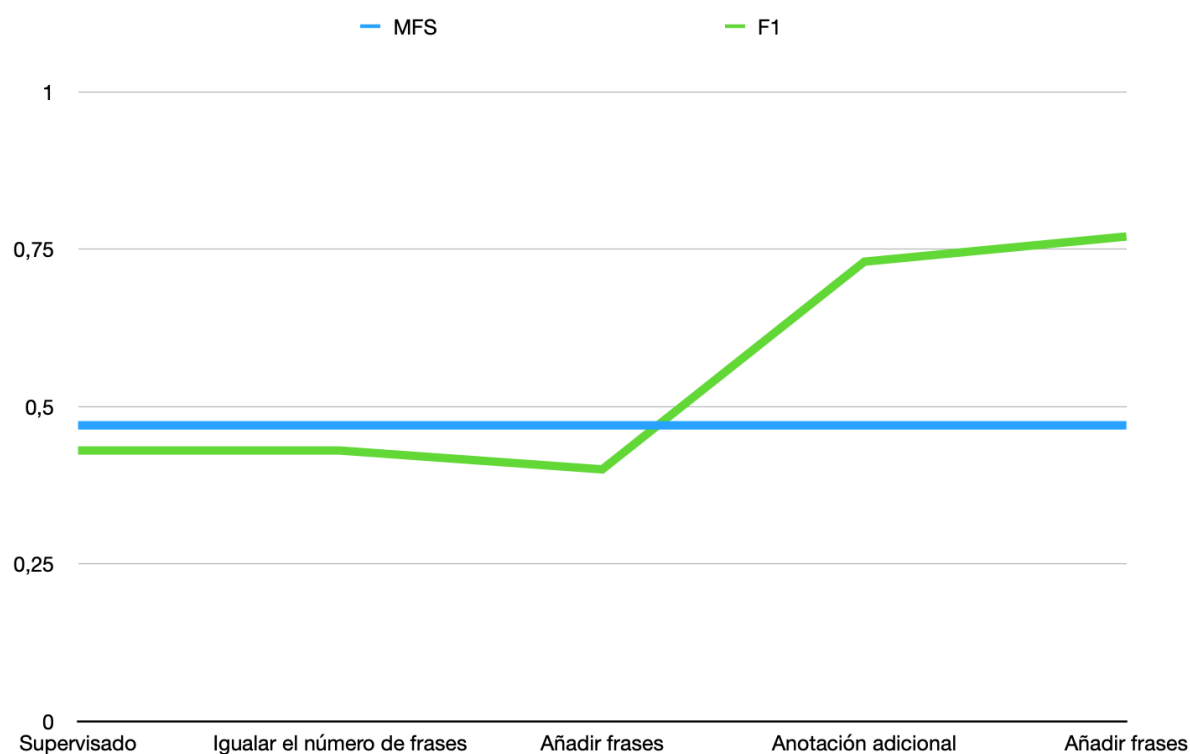


Figura (9): Evolución del valor F del lexema 'desembocar'

En la primera parte del experimento, el sistema no ha superado el valor MFS, que es bastante bajo, igual a 0,47. De hecho, los resultados han empeorado ligeramente después de la parte semi-supervisada. La ficha anotada contiene tres significados del verbo mencionados en el diccionario Clave (s. f.), que son:

1. Significado literal: desaguar.  
<El Ebro desemboca en el mar Mediterráneo.>
2. Significado literal: acabar o tener salida (referido a una calle).

<En esta plaza desembocan cuatro calles.>

3. Significado metafórico: Concluir o terminar (referido a una discusión, por ejemplo).

<La discusión desembocó en una pelea callejera.>

Los errores cometidos por el sistema se asocian principalmente a los dos primeros significados. Esto puede ser debido al hecho de que los primeros significados son bastante similares y que aparecen en contextos parecidos. Aquí, la diferencia entre el significado metafórico y los significados literales es evidente, por lo que el sistema no cometió muchos errores al clasificar el significado metafórico.

| SIGNIFICADO CORRECTO | SIGNIFICADO PREDICHO | FRASE                                                                                                                                             |
|----------------------|----------------------|---------------------------------------------------------------------------------------------------------------------------------------------------|
| 2                    | 1                    | El segundo desvío a la derecha <i>desembocaba</i> en Alcalá.                                                                                      |
| 2                    | 1                    | Calles con vistas, como Dom Pedro V de Lisboa, que <i>desemboca</i> en el mejor mirador de la ciudad.                                             |
| 2                    | 1                    | Se detienen al final del camino de entrada, que <i>desemboca</i> en una avenida vacía.                                                            |
| 3                    | 2                    | Estas añaden, sin embargo, que es muy pronto para saber en qué <i>desembocarán</i> unos procesos judiciales que podrían tardar años en dirimirse. |

Tabla (17): Análisis de errores del lexema 'desembocar'

Las anotaciones adicionales han aclarado mejor el contexto del segundo significado, lo que se refleja en los resultados. El valor se ha incrementado en 0,34 en comparación con el paso anterior del experimento supervisado y semi-supervisado, por lo que se ha logrado una mejora total de 0,29 en comparación con el valor MFS.

De estos datos se desprende que el sistema clasifica de forma errónea muchos menos casos. De hecho, el análisis de errores indica que el número de frases clasificadas incorrectamente se reduce en un 47,8%.

Después de haber clarificado los dos primeros significados en las anotaciones adicionales, los errores que aún se cometen, están en este caso relacionados con el significado metafórico. El

sistema frecuentemente elige la segunda interpretación del significado en vez del tercer significado.

| SIGNIFICADO<br>CORRECTO | SIGNIFICADO<br>PREDICHO | FRASE                                                                                                                                                              |
|-------------------------|-------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 3                       | 2                       | Cuando De la Rosa sale de prisión inicia una serie de contactos con Mario Conde que <i>desembocan</i> en los llamados pactos de la Salceda, la finca del banquero. |
| 3                       | 2                       | Lo que parecía una rutinaria tarea de escolta había <i>desembocado</i> en una surrealista conversación telefónica con el Presidente de los Estados Unidos.         |
| 3                       | 2                       | Estas añaden, sin embargo, que es muy pronto para saber en qué <i>desembocarán</i> unos procesos judiciales que podrían tardar años en dirimirse.                  |

Tabla (18): Análisis de errores del lexema 'desembocar'

La mayor mejora en la clasificación de significados se ha producido en el caso del verbo *amasar*. Como el valor MFS es muy bajo, el sistema ha logrado muy buenos resultados inmediatamente después de la parte supervisada. De hecho, el valor F es igual a 0,93 después de la parte supervisada, lo que significa simultáneamente que hay una mejora de 0,6 respecto al valor MFS. En este caso, el valor de la parte supervisada es el más alto. Esto significa que incluso después de añadir anotaciones adicionales, el valor de 0,93 no aumentó. En realidad, después de las anotaciones adicionales, el valor F ha disminuido ligeramente en 0,03 en comparación con la parte supervisada.

|     | SUPERVISADO | IGUALAR<br>EL NÚMERO<br>DE FRASES | AÑADIR<br>FRASES | ANOTACIÓN<br>ADICIONAL | AÑADIR<br>FRASES |
|-----|-------------|-----------------------------------|------------------|------------------------|------------------|
| MFS | 0,33        | 0,33                              | 0,33             | 0,33                   | 0,33             |
| F   | 0,93        | 0,93                              | 0,87             | 0,76                   | 0,87             |

Tabla (19): Evolución del valor F del lexema 'amasar'

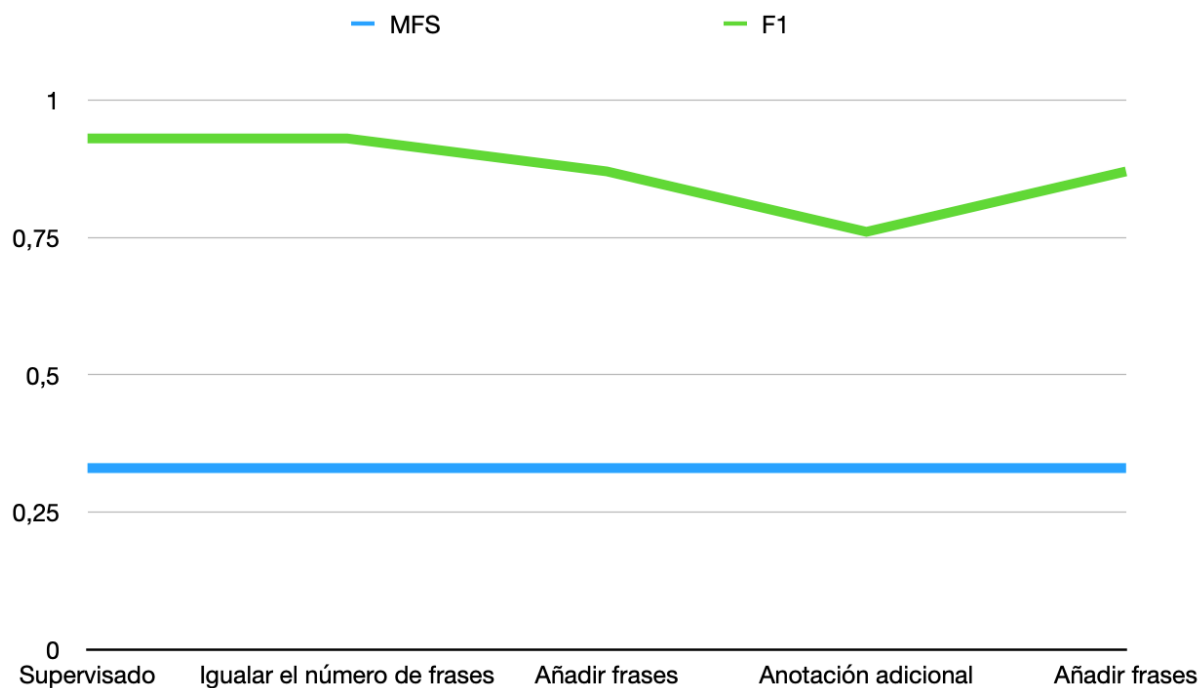


Figura (10): Evolución del valor F del lexema 'amasar'

La ficha anotada que se ha utilizado en la parte supervisada contiene solo dos significados del verbo *amasar*, lo que puede explicar los pocos errores. Los errores cometidos se limitan, obviamente, a estos dos significados (Clave, s. f.).

1. Significado literal: referido a una sustancia, hacer una masa con ella mezclándola con otros elementos y con algún líquido.

*<Para hacer pan hay que amasar harina, agua y levadura.>*

2. Significado metafórico: referido esp. al dinero o a los bienes, reunirlos o acumularlos.

*<Los negocios le permitieron amasar una gran fortuna.>*

| SIGNIFICADO CORRECTO | SIGNIFICADO PREDICHO | FRASE                                                                                                                                                    |
|----------------------|----------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2                    | 1                    | Cuando mi padre se retiró de la junta directiva, vendí la maldita fábrica de inmediato y <i>amasé</i> más dinero que Dios — dijo riendo entre dientes —. |

|   |   |                                                                                                                                            |
|---|---|--------------------------------------------------------------------------------------------------------------------------------------------|
| 1 | 2 | Soltó una risotada mientras su mano reptaba hacia el pecho de Irene y empezaba a manosearlo sobre el vestido como quien <i>amasar</i> pan. |
|---|---|--------------------------------------------------------------------------------------------------------------------------------------------|

Tabla (20): Análisis de errores del lexema 'amasar'

El caso de los verbos también demuestra que un gran número de significados por palabra no significa que el sistema sea peor en la tarea de clasificación. Tres de los nueve verbos utilizados en este estudio tenían más de dos significados, a saber, *desembocar* (tres significados), *partir* (tres significados) y *vencer* (cuatro significados) (véanse los significados correspondientes en el apéndice). Tras la tarea de clasificación, se han obtenido valores F iguales a 0,77 y 0,79 para *desembocar* y *vencer*, respectivamente. Estos valores son inferiores a la media, pero al mismo tiempo no son inferiores a la puntuación más baja de las palabras con solo dos significados (un valor F de 0,77 para la palabra *frenar*). Sin embargo, se demuestra lo contrario: el verbo *partir* tiene una puntuación de 0,9 como resultado final. La puntuación es una de las más altas de la lista e incluso equivale a la de *amasar*, para la que solo se anotaron dos significados.

## 6. CONCLUSIÓN Y DISCUSIÓN

Los capítulos anteriores han indicado que el problema de la clasificación automática de palabras polisémicas sigue siendo relevante en el campo de la lingüística computacional. El estudio de la clasificación automática de 29 lexemas con al menos un significado literal y otro metafórico mostró que un enfoque supervisado y semi-supervisado produce resultados pertinentes. El sistema utilizado en este estudio clasificó 16 de los 29 lexemas mejor que la línea de base MFS desde la fase inicial del método supervisado. Es un resultado positivo que no deja de sorprender. Los resultados obtenidos tras la parte semi-supervisada no siempre mejoraron el punto de partida supervisado. Lo que se observa, en cambio, es que en casi todos los casos los resultados sí han mejorado tras añadir anotaciones adicionales de las frases dudosas. Se trata de un paso adicional en un método supervisado.

En cuanto a las categorías morfológicas, no se puede decir que los lexemas de una categoría se hayan clasificado significativamente mejor que el resto. Sin embargo, se puede deducir que los adjetivos mostraron los resultados más bajos.



También se analizó brevemente si el número de significados por palabra ha afectado a los resultados de la tarea de clasificación. Aunque a primera vista parece que las palabras con solo dos significados se han clasificado mejor por el sistema, un análisis más profundo de los resultados no ha respaldado esta idea. Los lexemas para los que se han anotado tres o más significados se han clasificado generalmente tan bien o incluso mejor que los lexemas para los que solo se han anotado dos significados.

En cuanto al enfoque, se puede concluir que el sistema funciona mejor especialmente después de la parte supervisada. Esto se aplica tanto a la primera parte del experimento como a la parte en la que se añaden las anotaciones adicionales. La adición automática de frases por el propio sistema no ha producido una gran mejora en la media del valor F.

Para comprender de forma más crítica los resultados, conviene comentar algunas de las dificultades encontradas al efectuar el estudio.

En primer lugar, es necesario problematizar la idea de que dividir los usos de una palabra según sus significados es una tarea sencilla y conveniente para todas las tareas de clasificación. Aunque dividir las palabras en diferentes significados es una solución muy buena y, en muchos casos, muy efectiva para la desambiguación léxica automática, no siempre es factible en la práctica. Las anotaciones manuales, que forman la base de la investigación, no siempre son fáciles de compilar y ejecutar. Y no solo se trata de un problema práctico, sino también conceptual.

Kilgarrieff (1997) presentó este problema en su trabajo, *I don't believe in word senses*, en el que sostiene que los investigadores de la desambiguación léxica automatizada se basan demasiado en los significados de las palabras presentados en diccionarios. Al contrario, según el autor, no hay significados de palabras en el lenguaje natural que sean relevantes para investigaciones computacionales tal y como aparecen en un diccionario. Además, demuestra que el contexto en el que aparecen las palabras es mucho más complejo de lo que se presenta en los diccionarios. De hecho, Kilgarrieff (1997) sostiene que los significados de las palabras que aparecen en los diccionarios no son siempre correctos, ya que estos significados se eligen en función de una serie de parámetros. En primer lugar, subraya que todos los significados

elegidos son una elección subjetiva de los lexicógrafos. Además, todos los diccionarios se escriben para un grupo objetivo concreto, por lo que los significados elegidos pueden diferir también de un diccionario a otro. Por último, añade que un enfoque muy diferente de la división de significados podría tener consecuencias económicas y culturales para un editorial. Un diccionario completamente diferente se vendería mal y no sería aceptado como algo común. La consecuencia de esto para la desambiguación léxica automática es que los significados de las palabras siempre se definen solo con respecto a los intereses que se acaban de enumerar. El conjunto de significados definidos en un diccionario no siempre corresponde al conjunto relevante para una aplicación de NLP, como la identificación y clasificación de significados metafóricos, por ejemplo. De hecho, algunos significados metafóricos de algunas palabras no son suficientemente interesantes para incluirlos en un diccionario. Una consecuencia, por supuesto, es que esos significados no se incluirán en estudios como este.

En segundo lugar, se plantea el problema de la anotación manual de los diferentes significados de las palabras. En el aprendizaje automático supervisado en general, se ha observado con mayor frecuencia el problema del error humano en la anotación. También en este estudio, este fue el caso para un número limitado de anotaciones. En algunos casos, como en el de las palabras con muchos significados o en el de las palabras cuyos significados eran muy cercanos, los diferentes significados de una palabra se superponían, haciendo que la anotación fuera cuestionable. A continuación, se muestran algunos ejemplos de anotaciones que se consideran ahora erróneas (véanse los significados correspondientes en el apéndice).

| SIGNIFICADO<br>ANOTADO | SIGNIFICADO<br>CORRECTO | SIGNIFICADO<br>PREDICHO | FRASE                                                                                                                                               |
|------------------------|-------------------------|-------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------|
| 1                      | 2                       | 2                       | La <i>salud</i> del general es un factor latente desde que el juez Baltasar Garzón ordenó su arresto y el Gobierno español solicitó su extradición. |
| 1                      | 2                       | 2                       | Vamos a <i>llorar</i> el fallecimiento de Edward, no la muerte de la bestia.                                                                        |

Tabla (21): Ejemplos de anotaciones erróneas

Los resultados han mostrado entonces que es mejor considerar algunos significados como uno solo. Esto sería especialmente ventajoso en el caso de palabras cuyos significados pueden confundirse fácilmente, como en el caso de la palabra *salud*. Esto demuestra una vez más que los significados encontrados en un diccionario no siempre son los ideales para este tipo de investigación y que se necesita una manera consistente para la división de los significados. De hecho, la tarea de dividir y fusionar los diferentes significados fue una de las más difíciles del estudio. Esto, por supuesto, también implica el hecho de que los significados han tenido que ser elegidos cuidadosamente en función de su frecuencia en los corpus.

Otra dificultad a la hora de distinguir los significados metafóricos y literales es que se requiere el conocimiento de la estructura metafórica subyacente. Lakoff y Johnson (1980) han propuesto varios ejemplos, como *<el tiempo es dinero>*, del que pueden derivarse varias palabras polisémicas metafóricas como *ahorrar* o *gastar*. Estas formas básicas de las metáforas conceptuales podrían facilitar la clasificación tanto humana como automática de sus metáforas lingüísticas correspondientes. Sin embargo, algunas estructuras de metáforas conceptuales son más difíciles de reconocer que otras. Además, esta tarea también sería una tarea de clasificación en la que se pueden cometer errores.

Por ejemplo, la estructura de la metáfora conceptual básica de la palabra polisémica *calor*, *<las emociones son temperaturas>*, es fácil de reconocer para deducir metáforas lingüísticas como *<Discutían con calor, pero son buenos amigos.>* (Clave, s. f.). En cambio, aplicar la misma idea a la palabra *horno* es mucho más difícil. Aquí hay una similitud entre un aparato de cocina y una habitación donde la temperatura es alta, pero no es ideal tomar *<los aparatos de cocina son habitaciones>* como idea básica de metáfora conceptual para deducir metáforas lingüísticas como *<En verano, esta habitación es un horno.>* (Clave, s. f.).

Estos diferentes puntos de vista sobre la anotación de palabras con múltiples significados demuestran que hay muchos enfoques para esta tarea. Dado que la elección de los diferentes significados de una palabra es crucial en la desambiguación léxica automática, cabe esperar que un enfoque diferente, o una clasificación diferente de los significados, también hubiera producido resultados diferentes.

## 7. BIBLIOGRAFÍA

- Agirre, E., & Edmonds, P. (Eds.). (2007). *Word sense disambiguation: Algorithms and applications* (Vol. 33). Springer Science & Business Media.
- Arrese, Á. (2015). Las metáforas de la crisis del euro en la prensa española. *Communication & Society*, 28(2), 19-38.
- Bowdle, B. F., & Gentner, D. (1988). The career of metaphor. *Psychological review*, 112(1), 193.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of Deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Fundación Santa María-Ediciones SM (s. f.). *Clave Diccionario de Uso del Español Actual*. Clave. <http://clave.smdiccionarios.com/app.php>
- Goethals, P. (2018). Customizing vocabulary learning for advanced learners of Spanish. In T. Read, B. Sedano Cuevas, & S. Montaner-Villalba (Eds.), *Technological innovation for specialized linguistic domains : languages for digital lives and cultures, proceedings of TISLID'18* (pp. 229–240). Presented at the Third conference on Technological Innovation for Specialized Linguistic Domains (TISLID'18), Éditions Universitaires Européennes.
- Kilgarrieff, A. (1997). I don't believe in word senses. *Computers and the Humanities*, 31(2), 91–113.
- Knowles, M., & Moon, R. (2004). *Introducing metaphor*. Routledge.
- Kovecses, Z. (2010). *Metaphor: A practical introduction*. Oxford University Press.
- Lakoff, G., Johnson, M. (1980). *Metaphors we live by*. Chicago: The University of Chicago Press.
- Màrquez, L., Escudero, G., Martínez, D., & Rigau, G. (2007) Supervised Corpus-Based Methods for WSD. En Agirre, E., & Edmonds, P. (Eds.). (2007). *Word sense disambiguation: Algorithms and applications* (Vol. 33) (pp. 167 – 216). Springer Science & Business Media.
- Medina Montero, J. F. (2007). La metáfora en el léxico futbolístico: el caso de los participantes en español, y sus posibles equivalentes en italiano. En *Luque Toro, L.(ed.), Léxico Español Actual, Actas del I Congreso Internacional de Léxico Español*

- Actual, Venecia-Treviso, 14-15 de marzo de 2005, Venezia: Libreria Editrice Cafoscarina, 2007, pp. 197-239. Venezia, Libreria Editrice Cafoscarina.*
- Mihalcea, R. (2007) Knowledge-Based Methods for WSD. En Agirre, E., & Edmonds, P. (Eds.). (2007). *Word sense disambiguation: Algorithms and applications* (Vol. 33) (pp. 107 – 131). Springer Science & Business Media.
- Navigli, R. (2009). Word sense disambiguation: A survey. *ACM computing surveys (CSUR)*, 41(2), 1-69.
- Pedersen, T. (2007) Unsupervised Corpus-Based Methods for WSD. En Agirre, E., & Edmonds, P. (Eds.). (2007). *Word sense disambiguation: Algorithms and applications* (Vol. 33) (pp. 133 – 166). Springer Science & Business Media.
- Pérez, R. G. (2010). *Estudio cognitivo-contrastivo de las metáforas del cuerpo: análisis empírico del corazón como dominio fuente en inglés, francés, español, alemán e italiano* (Vol. 64). Peter Lang.
- Small, S. L., Cottrell, G. W., & Tanenhaus, M. K. (Eds.). (1988). *Lexical Ambiguity Resolution: Perspective from Psycholinguistics, Neuropsychology and Artificial Intelligence*. Elsevier.
- Taghipour, K., & Ng, H. T. (2015). Semi-supervised word sense disambiguation using word embeddings in general and specific domains. En *Proceedings of the 2015 conference of the North American chapter of the association for computational linguistics: human language technologies* (pp. 314-323).
- Van Engelen, J. E., & Hoos, H. H. (2020). A survey on semi-supervised learning. *Machine Learning*, 109(2), 373-440.

## 8. APÉNDICE

LISTA ALFABÉTICA DE LEXEMAS UTILIZADOS CON SU SIGNIFICADOS LITERALES Y METAFÓRICOS<sup>1</sup>

### ADJETIVOS

#### ANCHO

1. Que tiene más anchura o mide más horizontalmente de lo que es necesario o habitual.  
*<Los pantalones de campana son muy anchos por abajo.>*
2. Amplio, espacioso o con más espacio del necesario  
*<Cuando solo vamos tres en el coche, el que va en el asiento de atrás va muy ancho.>*
3. **Orgulloso, satisfecho, ufano y contento.**  
*<No entiendo cómo podéis haberos enterado de esto y quedaros tan anchos.>*  
**quedarse (tan) ancho**  
*<Tiró la colilla a la arena y se quedó tan ancho.>*

#### BAJO

1. Que tiene menos altura de la que se considera normal.  
*<Al lado de un jugador de baloncesto pareces muy bajo.>*  
*<El cielo estaba cubierto por nubes bajas.>*
2. Referido a un terreno, que está a poca altura sobre el nivel del mar.  
*<Las tierras bajas se inundan con mucha facilidad.>*
3. **Que tiene un valor o una intensidad inferiores a los normales.**  
*<En verano la presión atmosférica es más baja.>*  
**Que está en un lugar inferior o que ocupa una posición inferior.**  
*<La planta baja está destinada a locales comerciales.>*
4. **Inclinado o dirigido hacia abajo.**

---

<sup>1</sup> Fundación Santa María-Ediciones SM (s. f.). *Clave Diccionario de Uso del Español Actual*. Clave.  
<http://clave.smdiccionarios.com/app.php>

*<Siempre me habla con los ojos bajos.>*

- 5. Referido a un sonido, a una voz o a un tono musical, que tienen una frecuencia de vibraciones pequeña.**

*<Su hermano tiene una voz baja, ronca y hueca.>*

## CALIENTE

1. Con temperatura elevada.

*<No planches este vestido con la plancha muy caliente.>*

- 2. Acalorado, vivo o apasionado.**

*<Los ánimos estaban calientes y una tontería desencadenó la pelea.>*

**Conflictivo.**

*<Será un invierno caliente por la crisis del Gobierno.>*

**en caliente**

*<Si no lo hiciste en caliente, ya no merece la pena que te enfrentes a él.>*

## ESTRECHO

1. Que tiene poca anchura o menos anchura de la normal.

*<Fuimos por una carretera muy estrecha.>*

Ajustado o apretado.

*<Esta camisa me queda estrecha.>*

- 2. Rígido, riguroso o estricto.**

*<El policía los sometió a una estrecha vigilancia.>*

**Referido esp. a una relación, que es muy íntima o que se asienta en fuertes vínculos.**

*<Trabajamos en estrecha colaboración.>*

## FRESCO

1. Que tiene una temperatura moderada o agradablemente fría.

*<En mi casa, el agua del grifo sale fresca.>*

2. Referido esp. a un alimento, que acaba de ser obtenido o que no está curado.

*<Hay lechugas frescas de la huerta.>*

Referido a un alimento, que no ha sido congelado.

*<Aunque sea más cara, prefiero que compres la merluza fresca.>*

Referido esp. a una pintura, que aún no se ha secado.

*<No toques la puerta porque la pintura está fresca y te vas a manchar.>*

**3. Referido esp. a un acontecimiento, que acaba de suceder o que es inédito.**

*<Me pasé por allí y traigo noticias frescas.>*

**FRÍO**

**1. Con temperatura inferior a la normal o a la conveniente.**

*<En verano, desayuno leche fría.>*

**2. Referido a una persona, que es poco afectuosa o que se muestra indiferente ante estímulos y sensaciones.**

*<Es una mujer fría, distante y muy seria.>*

**Sin pasión o sin mostrar emoción o afecto.**

*<No me gusta este pintor porque sus cuadros son fríos y cerebrales.>*

**LARGO**

**1. Que tiene mucha longitud o más de la normal o de la necesaria.**

*<Después de la curva hay una recta muy larga.>*

*<Tiene una camisa de manga larga.>*

*<El abrigo me queda largo.>*

Referido a una prenda de vestir, que llega hasta los pies.

*<En la recepción las mujeres iban con vestidos largos.>*

*<Antes, los hombres llevaban pantalones largos y los niños, cortos.>*

**2. Dilatado o extenso.**

*<Fue una película demasiado larga.>*

*<El relato me resultó largo y aburrido.>*

*<Ayer no pude acabar la novela porque era muy larga.>*

**Referido a una cantidad, que es más de lo que indica.**

*<Vivió en aquel pueblo treinta años largos.>*

*<Ledesma está a 30 kilómetros largos de Salamanca.>*

**NEGRO**



1. Del color del carbón o de la oscuridad absoluta.

*<¿Has visto un gato negro?>*

De color más oscuro en relación con algo de la misma especie o clase.

*<El olor del desayuno con huevos frescos, pan negro, gachas y jamón la asaltó nada más entrar en la habitación.>*

2. **Oscuro, oscurecido o deslucido.**

*<Va a caer un chaparrón, porque las nubes están negras.>*

3. **Triste, desgraciado o poco favorable.**

*<He tenido un día negro.>*

4. Referido a una persona, que pertenece al grupo étnico caracterizado, entre otros rasgos, por el color oscuro de su piel y por la forma rizada y tiesa del cabello.

*<Era hija de un criollo cubano y de su esclava negra, que había muerto en el parto.>*

## SUSTANTIVOS

### BASURA

1. Conjunto de desperdicios y de cosas que no sirven y se tiran.

*<Bilbao y San Sebastián superan la media nacional de producción de basura.>*

2. **Lo que se considera de mala calidad, mal hecho o de poco valor.**

*<Esa novela tan mal escrita es una basura.>*

### CABEZA

1. En una persona y en algunos animales, parte superior o anterior del cuerpo en la que se encuentran algunos órganos de los sentidos.

*<Los toros tienen cuernos en la cabeza.>*

En una persona y en algunos mamíferos, parte que comprende desde la frente hasta el cuello, excluida la cara.

*<Como tenía piojos, le raparon la cabeza.>*

2. **En un reparto o en una distribución, persona.**

*<En el reparto tocaron a tres por cabeza.>*

**Animal cuadrúpedo de ciertas especies domésticas y de algunas salvajes.**

*<Tenía un rebaño de doscientas cabezas.>*

**3. Pensamiento, imaginación o capacidad intelectual humana.**

*<Al verte me vino a la cabeza que te debo dinero.>*

**Inteligencia, sensatez o agilidad mental.**

*<Le estoy dando vueltas a la cabeza, pero no encuentro la solución.>*

**4. En algo que se considera un todo, parte más importante o más visible.**

*<La cabeza de la Iglesia católica es el papa.>*

**CALOR**

**1. Sensación que experimenta el cuerpo animal con una subida de temperatura.**

*<Si tienes calor, apaga el radiador.>*

Temperatura ambiental elevada.

*<En verano hace calor.>*

**2. Afecto, simpatía y adhesión.**

*<Busca en los amigos el calor que no encuentra en casa.>*

**Entusiasmo o apasionamiento.**

*<Discutían con calor pero son buenos amigos.>*

**CAMINO**

**1. Vía por donde se transita habitualmente, esp. si es de tierra apisonada y sin asfaltar.**

*<No hay ningún camino que atraviese el bosque.>*

Trayecto o itinerario que se sigue para ir de un lugar a otro.

*<El peregrino hizo el camino por etapas.>*

de camino

*<Si quieres te llevo, porque tu casa me pilla de camino.>*

camino de cabras

*<No intentes meter el coche por ese camino de cabras.>*

camino de

*<Camino de casa, me encontré con unos amigos.>*

*<Van camino de su pueblo.>*

**2. Dirección que se sigue para llegar a un lugar o para conseguir algo.**

*<Las conversaciones van por buen camino y pronto se conseguirá un acuerdo.>*

**Medio para hacer o conseguir algo.**

*<El mejor camino para aprobar es estudiar.>*

**abrirse camino**

*<La vida es dura y solo tú puedes abrirte camino.>*

**CIELO**

1. Espacio en el que se mueven los astros y que, visto desde la Tierra, parece formar sobre ella una cubierta arqueada.

*<Esta noche se ven más estrellas en el cielo.>*

Capa gaseosa que rodea la Tierra.

*<Mira qué nubes se están formando en el cielo.>*

2. **En algunas tradiciones religiosas, morada de Dios.**

*<Cuando era pequeño, me decían que si era bueno iría al cielo.>*

**En algunas religiones, goce eterno que se disfruta en presencia de Dios.**

*<Las almas justas gozan del cielo.>*

**ESPEJO**

1. Lámina de vidrio cubierta de mercurio por la parte posterior para que se refleje en ella lo que hay delante.

*<Fue al baño a mirarse en el espejo.>*

espejo retrovisor

*<Chocó contra uno de los coches aparcados de refilón, lo abolló, destrozó su espejo retrovisor y finalmente logró controlar el vehículo.>*

2. **Lo que reproduce algo fielmente.**

*<La novela realista busca ser un espejo de la sociedad.>*

**Lo que se toma como modelo o es digno de imitación.**

*<Esa mujer es un espejo de bondad.>*

**FIEBRE**

1. Aumento anormal de la temperatura del cuerpo, que es síntoma de algún trastorno o enfermedad.

*<Sufre una pulmonía grave y tenía fiebre muy alta.>*

Enfermedad infecciosa cuyo síntoma fundamental es un aumento anormal de la temperatura.

*<Si vas a viajar a África, debes vacunarte contra las fiebres.>*

fiebre aftosa

*<Cataluña se ha puesto en marcha contra la fiebre aftosa.>*

fiebre amarilla

*<Los expertos recomiendan vacunarse contra la fiebre amarilla.>*

fiebre tifoidea

*<Se inmunizó contra la fiebre tifoidea por medio de una pastilla oral.>*

**2. Ansiedad o agitación con que se lleva a cabo una actividad.**

*<Esa fiebre por ganar dinero te puede llevar a meterte en negocios sucios.>*

## HORNO

1. Aparato o construcción de albañilería fabricados para generar calor, que se utilizan para caldear, cocer o fundir una materia que se introduce en su interior.

*<Encendé el horno y poné las milanesas congeladas.>*

En una cocina, aparato que se usa para asar los alimentos.

*<Entramos a la cocina y preparé el pescado para ponerlo en el horno y le ofrecí una cerveza.>*

horno crematorio

*<Se proponía llevar el cadáver hasta un cercano horno crematorio.>*

**2. Lugar donde hace mucho calor.**

*<En verano, esta habitación es un horno.>*

## LUZ

1. Forma de energía que ilumina y hace posible la visión.

*<La luz se propaga por medio de los fotones.>*

*<De noche nos iluminamos con luz eléctrica.>*

Claridad o destello que despiden algunos cuerpos.

*<La luz de las llamas es roja y amarilla.>*

Corriente eléctrica.

*<Les cortaron la luz porque no pagaban.>*

luz cenital

*<Una luz cenital lo enfoca sólo a él que, sentado en un taburete, se enfrenta al público.>*

luz natural

*<Estamos en un sótano, así que no podemos calcular cuánto tiempo queda de luz natural.>*

**2. Aparato que sirve para alumbrar, o dispositivo que pone en marcha este aparato.**

*<Se ha fundido la luz.>*

*<Trae una luz, que no veo nada.>*

**3. Modelo que sirve de ejemplo o de guía.**

*<Mi mujer es la luz de mi vida.>*

**4. Aclaración o ayuda.**

*<Estos documentos arrojan nueva luz sobre el asunto.>*

**Claridad mental.**

*<Demuestras tener pocas luces si te has creído esa mentira.>*

**MIEL**

**1. Sustancia viscosa, amarillenta y muy dulce que producen las abejas.**

*<Buth le tendió una taza de té con miel.>*

**2. Satisfacción o sensación agradable proporcionadas por el éxito.**

*<Tienes que entrenarte a fondo si quieres saborear las mieles del triunfo.>*

**PARAGUAS**

**1. Utensilio portátil compuesto por un bastón y por unas varillas plegables cubiertas por una tela impermeable, que se utiliza para protegerse de la lluvia.**

*<En el exterior está cayendo una tormenta de esas que hacen que te mojes aunque vayas con el paraguas.>*

**2. Ámbito de acción o de influencia.**

*<Esta crisis bélica debe resolverse dentro del paraguas de la ONU.>*

**SALUD**

**1. Estado en el que un organismo vivo realiza normalmente todas sus funciones.**

*<Contaminar el medio ambiente es un delito contra la salud pública.>*

Condiciones físicas en que se encuentra el organismo de un ser vivo en un determinado momento.

*<Desde hace algún tiempo su salud no es buena porque padece de los nervios.>*

**2. Buen estado o buen funcionamiento de algo.**

*<Nadie duda de la salud de la economía de este país.>*

## **VERBOS**

### **AMASAR**

1. Referido a una sustancia, hacer una masa con ella mezclándola con otros elementos y con algún líquido.

*<Para hacer pan hay que amasar harina, agua y levadura.>*

**2. Referido esp. al dinero o a los bienes, reunirlos o acumularlos.**

*<Los negocios le permitieron amasar una gran fortuna.>*

### **BRILLAR**

1. Despedir rayos de luz, propia o reflejada.

*<El sol brilla en el firmamento.>*

**2. Sobresalir o destacar de manera que despierta admiración.**

*<Su inteligencia y belleza brillan allá donde va.>*

### **DESEMBOCAR**

1. Referido especialmente a un río, verter sus aguas.

*<El Ebro desemboca en el mar Mediterráneo.>*

2. Referido especialmente a una calle, acabar o tener salida.

*<En esta plaza desembocan cuatro calles.>*

**3. Concluir o terminar.**

*<La discusión desembocó en una pelea callejera.>*

### **DISPARAR**

1. Referido esp. a un arma, hacer que lance un proyectil o lanzarlo.

*<La ametralladora dispara con rapidez.>*

*<Apreté el gatillo y disparé.>*

*<Se le disparó el arma y hubo un herido.>*

Referido a un objeto, esp. a un proyectil, lanzarlo con violencia o salir despedido.

*<Los arqueros dispararon sus flechas.>*

**2. Crecer o aumentar rápidamente y sin moderación.**

*<La subida del petróleo ha disparado los precios de la gasolina.>*

*<Con el descontento social y la crisis, se ha disparado la violencia.>*

**FRENAR**

**1. Referido a un vehículo, moderar su marcha o pararlo con el freno.**

*<El conductor no pudo frenar el coche y se salió de la curva.>*

*<Frena un poco, que vamos demasiado deprisa.>*

**2. Referido especialmente a una persona o a sus actos, moderarlos, contenerlos o detenerlos.**

*<Frénate y no le digas nada, que es el jefe.>*

*<Con la edad ha aprendido a frenar sus impulsos.>*

**LLORAR**

**1. Derramar lágrimas.**

*<Esta vez lloraba de alegría.>*

*<Con tanto humo, me lloran los ojos.>*

**2. Referido a un suceso desgraciado, lamentarlo o sentirlo profundamente.**

*<Toda la ciudad lloró la muerte de la alcaldesa.>*

**PARTIR**

**1. Referido a un todo, dividirlo o separarlo en varias partes.**

*<Partió la cuerda en dos trozos.>*

*<Parte el jamón en tacos pequeños.>*

Referido a una parte, cortarla y separarla de un todo.

*<Me partió un trozo de queso demasiado grande.>*

Referido a un todo, repartirlo o distribuirlo entre varios.

*<Partió todas sus posesiones entre sus familiares más allegados.>*

**2. Ponerse en marcha.**

*<El tren partió de la estación a la hora prevista.>*

**3. Arrancar, proceder o tener el origen.**

*<¿De quién partió la idea de organizar la fiesta?>*

**a partir de**

*<A partir de ese día, no volví a verlo.>*

**SUFRIR**

1. Referido a un daño moral o físico, experimentarlos, sentirlos o vivirlos con intensidad.

*<Sufre continuos dolores de cabeza.>*

Referido a un daño moral o físico, recibirlos con resignación o aceptarlos sin quejas.

*<Sufría en silencio aquellos desplantes.>*

**2. Referido especialmente a un peso, sostenerlo o resistirlo.**

*<Los cimientos sufren el peso del edificio.>*

**VENCER**

1. Referido a un enemigo, sujetarlo, someterlo o derrotarlo.

*<Julio César venció a los galos.>*

En una competición o en una comparación, aventajar, superar, exceder o resultar preferido.

*<El equipo español vence por cuatro puntos al final de la primera parte.>*

**2. Referido a un plazo o a un término de tiempo, cumplirse o pasar.**

*<El plazo de matrícula vence el próximo día 15.>*

**3. Referido a una persona, rendirla, fatigarla o poderla.**

*<Me voy a la cama porque me vence el sueño.>*

**4. Referido a una dificultad o a un obstáculo, superarlos obrando contra ellos.**

*<Debes vencer la falta de medios y presentar el trabajo como sea.>*