

La detección de marcadores de negación en español: el sistema CLiC-Neg

Mónica González Manzano

TFM presentado para optar
al grado de **Máster**
en el programa de Máster de *Ciencia Cognitiva y Lenguaje*
Departamento de Lingüística
Universidad de Barcelona

bajo la supervisión de

Dra. Maria Antònia Martí Antonín
Universidad de Barcelona

Dra. Maria Taulé Delor
Universidad de Barcelona



UNIVERSITAT DE
BARCELONA

Julio de 2019

Agradecimientos

Este trabajo de máster no es un esfuerzo individual. Es el resultado del trabajo de todo un equipo que me ha enseñado, guiado y ayudado en todo momento.

En primer lugar, quiero agradecer el esfuerzo que han hecho mis codirectoras, M. Antònia Martí y Mariona Taulé, por su dedicación y por haberme dado la oportunidad de aprender a su lado. Muchas gracias también a todo el equipo del Servei de Tecnologies Lingüístiques de la UB. Ha sido un lujo compartir despacho con todos ellos. Especialmente a Montse Nofre por su ayuda y por enseñarme tanto y a Javier Beltrán. Todo el mérito del NEGES 2019 es suyo.

Gracias a mis compañeros del máster CCiL, con los que he compartido angustias y nervios durante el curso. Han hecho que este año fuera mucho más llevadero y, sobre todo, más alegre. A Marina Bolea, por compartir las reuniones sobre negociación, por ayudarme con mi TFM y dejarme seguir de cerca su TFG.

Por último, gracias, una vez más, a los de siempre. A los amigos reencontrados, por seguir ahí, sin importar ni el tiempo ni la distancia. A mi hermano, a Eva y a mis bichos, por ampliar la familia. A mis padres, por no rendirse nunca. Y, cómo no, a Esteban, por su apoyo incondicional. Este TFM es el resultado de unir un apasionado de la computación con una loca de la lingüística. Muchas gracias por descubrirme el mundo de la programación, de los Pythonistas y las posibilidades inagotables de combinar el lenguaje y la informática.

Resumen

La detección de la negación es un problema central en el ámbito del Procesamiento del Lenguaje Natural, ya que es fundamental que los sistemas de extracción de la información puedan distinguir si un concepto aparece negado o no. Por este motivo, se han desarrollado distintos algoritmos de detección de marcadores de negación. Los primeros, como NegEx, se basaron en reglas de tipo léxico y sintáctico. En cambio, en los últimos años ha habido un gran interés por aplicar técnicas de aprendizaje supervisado. En esta línea se han desarrollado diferentes sistemas, para el inglés y el español, que han obtenido resultados positivos.

En este trabajo final de máster presentamos nuestra aportación a la detección automática de marcadores de negación, el sistema CLiC-Neg. Se trata de un modelo basado en técnicas de aprendizaje automático supervisado que aplica un modelo estadístico, el Conditional Random Field (CRF). Ha sido entrenado y evaluado con el corpus SFU Review_{SP}-NEG, un corpus de reseñas en español. El sistema CLiC-Neg obtiene un valor-f del 84,09% en el corpus de evaluación y ha sido el primer clasificado en la subtarea A en el taller NEGES 2019.

Abstract

Negation detection is a pressing issue in Natural Language Processing, given that information extraction systems need to correctly identify whether a concept is negated. For this reason, several algorithms have been developed to improve negation cues detection. The first algorithms, such as NegEx, were based on lexical and syntactic rules. By contrast, during the last decade, researchers have been more interested in applying machine learning techniques, in English and Spanish, and have obtained positive results.

In this Master's thesis we describe our contribution to negation cues detection, the CLiC-Neg system. It is a model based on supervised machine learning techniques that uses a statistical model, the Conditional Random Field (CRF). We have trained and evaluated our system with the corpus SFU Review_{SP}-NEG, a corpus which contains reviews written in Spanish. The CLiC-Neg system obtains a F1 score of 84.09% and was the first classified in the subtask A in the Workshop NEGES 2019.

Índice

Capítulo 1. Introducción.....	1
1.1. La negación en biomedicina y análisis de sentimientos.....	1
1.2. Motivación y objetivos del trabajo.....	3
Capítulo 2. Tratamiento de la negación en PLN.....	3
2.1. Algoritmos de detección de marcadores de la negación basados en reglas.....	4
2.1.1. Algoritmos basados en reglas de tipo léxico.....	4
2.1.2. Algoritmos que incorporan conocimiento sintáctico.....	6
2.2. Sistemas de detección de la negación basados en aprendizaje automático.....	9
2.2.1. Corpus anotados con negación.....	9
2.2.1.1. Corpus en inglés anotados con negación.....	9
2.2.1.2. Corpus en español anotados con negación.....	10
2.2.2. Propuestas de modelos de detección de la negación basadas en aprendizaje automático.....	11
2.2.2.1. Aprendizaje automático para detectar marcadores de negación en inglés.....	11
2.2.2.2. Aprendizaje automático para detectar marcadores de negación en español: NEGES.....	13
Capítulo 3. Metodología para la detección automática de marcadores de negación.....	15
3.1. Marco del trabajo.....	16
3.1.1. El corpus SFU Review _{sp} -NEG.....	16
3.1.2. Formato del corpus.....	18
3.1.3. Etiquetas BIO para la identificación de marcadores de negación.....	20
3.2. Creación del sistema de detección de marcadores de negación CLiC-Neg.....	20
3.2.1. Base de referencia.....	20
3.2.2. El modelo CRF de detección de marcadores de negación.....	22
3.2.3. Análisis de errores del modelo CRF.....	24
3.2.4. El sistema CLiC-Neg.....	31
Capítulo 4. Resultados del NEGES 2019.....	34
Capítulo 5. Conclusiones.....	36
Apéndice. Lista completa de los marcadores de negación multpalabra incorporados al modelo CRF.....	36
Bibliografía.....	39

Índice de tablas

Tabla 1. Resultados de precisión y cobertura obtenidos por los principales algoritmos de detección de marcadores de negación basados en reglas.....	8
Tabla 2. Resultados de precisión y cobertura de los sistemas de aprendizaje automático más destacados.....	14
Tabla 3. Distribución de los valores de las estructuras negativas por dominio en el corpus SFU Review _{SP} -NEG.....	17
Tabla 4. Estadísticas del corpus SFU Review _{SP} -NEG.....	17
Tabla 5. Estadísticas del corpus de desarrollo y evaluación del taller NEGES.....	18
Tabla 6. Resultados de la base de referencia con el corpus de entrenamiento.....	21
Tabla 7. Resultados del modelo CRF con el corpus de desarrollo.....	24
Tabla 8. Valores-f de los diferentes sistemas evaluados con el corpus de desarrollo.....	33
Tabla 9. Precisión, cobertura y valor-f del sistema CLiC-Neg en el corpus de evaluación.....	34
Tabla 10. Valor-f obtenido por cada uno de los equipos participantes en el NEGES 2019 en el corpus de evaluación.....	35

1. Introducción

La detección de la negación es un problema central en el Procesamiento del Lenguaje Natural (PLN), puesto que es muy importante saber si un enunciado está negado o no. Muchos sistemas que reconocen entidades nombradas (del inglés *Named Entity Recognition*, NER) y técnicas de extracción de la información (del inglés *Information Extraction*, IE) son capaces de reconocer y extraer entidades, pero sin detectar si están negadas o no. Para entender la importancia de la negación, hay que considerar enunciados como (1):

1. El paciente **no** menciona sufrir **migrañas**.

Ante esta oración, los sistemas de PLN son capaces de extraer el concepto ‘migrañas’ y asociarlo al paciente. Sin embargo, no siempre detectan que este enunciado aparece negado. Esta información es imprescindible para saber si un paciente sufre o no migrañas y así facilitar el diagnóstico médico.

1.1. La negación en biomedicina y análisis de sentimientos

El tratamiento computacional de la detección de la negación y la modalidad es una área de investigación emergente en PLN (Morante & Sporleder, 2012:223). Ha habido dos campos en los que la detección de la negación ha sido especialmente fructífera: la biomedicina y el análisis de sentimientos.

En el campo de la biomedicina, detectar si una entidad está negada o no es fundamental para los sistemas de extracción de la información y de detección de entidades, como hemos mostrado en (1). De ahí que el primer algoritmo de detección de marcadores de negación, el NegEx (Chapman et al., 2001), se desarrollara con textos del ámbito clínico.

En el área del análisis de sentimientos, detectar la negación es básico para mejorar la minería de opiniones y analizar tendencias. Destacan en este campo las propuestas de Lifeng Jia et al. (2009), Wiegand et al. (2010), Alkorta et al. (2018) y Taboada et al. (2015), que con la detección de la negación buscan mejorar los sistemas para determinar la polaridad de los enunciados. De hecho, el corpus con el que hemos trabajado, el SFU Review_{SP}-NEG, contiene anotaciones sobre la polaridad y los cambios que esta experimenta a causa de la negación, con el objetivo de que sea útil para el análisis de sentimientos en español (Jiménez-Zafra et al., 2018a).

1.2. Motivación y objetivos del trabajo

El desarrollo de la detección automática de la negación ha sido relevante en inglés y en un registro muy específico, el de la biomedicina. De ahí que la comunidad científica sea cada vez más consciente de dos necesidades en este ámbito. La primera, la de ampliar y adaptar estas técnicas a otras lenguas además del inglés. La segunda, la de trabajar con corpus que contengan muestras lingüísticas que no pertenezcan a la biomedicina. De ahí la relevancia del corpus que hemos utilizado, el SFU Review_{SP}-NEG, ya que contiene reseñas escritas en español sobre diferentes productos extraídos de la web Ciao.es. En consecuencia, ofrece una buena

aproximación a la lengua real, además de incluir registros menos formales y un vocabulario que no es de especialidad.

El presente trabajo está motivado precisamente por esta necesidad de crear sistemas de detección automática de marcadores de negación con un corpus de temática más general en español. Además, hemos tenido en cuenta los resultados positivos que han obtenido las propuestas para la detección automática de marcadores de negación basadas en aprendizaje automático, tanto en inglés como en español. Partiendo de estos modelos, hemos desarrollado nuestro propio sistema de detección de marcadores de negación en español, el CLiC-Neg.

El sistema CLiC-Neg ha sido desarrollado para participar en la subtarea A del taller NEGES 2019 (Jiménez-Zafra et al., 2019). Dicha tarea consiste en diseñar un sistema de detección de marcadores de negación en español. Nuestro objetivo es resolver la tarea aplicando técnicas de aprendizaje automático supervisado. En concreto, probar un modelo basado en Conditional Random Field (CRF). Los objetivos del presente trabajo, por tanto, son los siguientes:

1. Desarrollar un sistema de detección automática de marcadores de negación en español.
2. Experimentar con técnicas de aprendizaje automático para poder evaluar si son un enfoque adecuado para nuestra tarea.
3. Aplicar un modelo basado en CRF para comprobar hasta qué punto es un enfoque válido para nuestra tarea.
4. Evaluar el impacto de añadir reglas y listas de marcadores de negación al modelo CRF.
5. Detectar líneas de investigación futuras sobre el tema de la negación.

El presente trabajo está organizado de la siguiente manera: el capítulo 2 describe el estado de la cuestión. Ofrece un resumen de las propuestas más relevantes en el ámbito de la detección de la negación y las clasifica en dos categorías: las que utilizan algoritmos basados en reglas y las que aplican técnicas de aprendizaje automático. El capítulo 3 describe el marco del trabajo y explica en qué consiste la subtarea A del taller NEGES 2019 y cómo se han procesado los datos. Este capítulo también describe la metodología utilizada. Esto es, describe el proceso que hemos seguido para elaborar nuestro sistema de detección de marcadores de negación, el CLiC-Neg. Exponemos los resultados de las diferentes versiones de nuestro sistema, basadas todas en CRF. El capítulo 4 describe los resultados obtenidos por el CLiC-Neg en el corpus de evaluación proporcionado por los organizadores del NEGES 2019. Asimismo, ofrece una comparativa entre los resultados de nuestro sistema y los del resto de participantes. Por último, en el capítulo 5 se presentan las conclusiones, destacando los resultados más relevantes del trabajo, así como las futuras líneas de investigación.

2. Tratamiento de la negación en PLN

Detectar automáticamente la negación plantea varios problemas de difícil solución. Los principales retos se circunscriben a tres tareas fundamentales: i) detectar los marcadores de negación, ii) delimitar su alcance e iii) identificar su foco.

Definir los marcadores de negación puede parecer una cuestión trivial. Sin embargo, un análisis de este fenómeno lingüístico plantea varios interrogantes. El primero es que el español es una lengua que presenta la posibilidad de tener marcadores discontinuos en ejemplos como (2). Crear un sistema que etiquete ‘nada’ correctamente como una continuación del marcador ‘no’ y no como un nuevo marcador de negación resulta complejo:

2. (...) **no** tengo **nada** en contra de Opel.

Además, existen enunciados que contienen marcadores de negación pero no presentan un verdadero significado negativo, como se puede ver en los siguientes ejemplos (Bolea, 2019):

3. El hotel está localizado en el medio de la **nada**.
4. El coche lo compré para viajar, **no**?
5. **No** me gusta **tanto** como lo otro.

Así, en (3) un detector de negación clasifica este enunciado como una negación por la presencia del marcador ‘nada’, pero en este contexto ‘nada’ es un sustantivo y la semántica de la oración es afirmativa. En (4), el marcador ‘no’ aparece en una interrogación retórica y tampoco el enunciado presenta un valor de negación. Por último, (5) expresa una comparación y, en realidad, una afirmación implícita. Esto es, en (5) el hablante afirma que un elemento no está al nivel de otro, pero no está negando que ese elemento le guste. Como se explica en Bolea (2019), los detectores automáticos de la negación presentan dificultades para identificar correctamente este tipo de enunciados y ejemplos como (3) a (5) son falsos positivos del sistema, esto es, el sistema los clasifica como enunciados que expresan negación y no debería.

Otro problema es definir qué tipo de información se tendrá en cuenta para definir los marcadores de negación. En general, los sistemas de tratamiento de la negación solo tienen en cuenta el nivel de la oración y, en general, no consideran el uso de prefijos de valor negativo, como pueden ser ‘i-’, ‘in-’, o ‘des-’, entre otros. Asimismo, la negación léxica, esto es, el uso de palabras que pueden expresar negación, como ‘apenas’ o ‘cero’, también juegan un papel importante en la definición de marcador de negación. No obstante, el sistema que hemos creado trata la negación básicamente a nivel de oración y, por tanto, no detecta ni la negación morfológica ni la léxica.

Tampoco es fácil definir el alcance de la negación. En el presente trabajo partimos de la idea de que el alcance se corresponde con todos los elementos afectados por la negación (Morante &

Daelemans, 2009:21). Sin embargo, no hay que confundir el alcance con el foco de la negación (Jiménez-Zafra et al., 2018b:7):

6. **No** pienso ir al concierto **ni** contigo **ni** con **nadie**¹.

En (6), el alcance del marcador de la negación ‘no’ es toda la oración. En cambio, puesto que el hablante niega que quiera ir acompañado al concierto (pero no niega que quiera ir al concierto) el foco es ‘ni contigo ni con nadie’. Estos son los elementos más afectados por la negación. Por tanto, el alcance siempre es un conjunto de palabras más extenso que el foco. De todas formas, definir el foco de la negación es complicado porque los criterios para detectarlo no son solo de tipo sintáctico, sino semántico y pragmático. En definitiva, se trata de una cuestión que supera los objetivos del presente trabajo y por eso no lo desarrollaremos.

En los siguientes apartados describimos los algoritmos y sistemas de detección automática de marcadores de negación más relevantes. Las propuestas que se han presentado desde el ámbito del PLN parten, fundamentalmente, de dos enfoques diferentes que no son incompatibles: el primero, como explicamos en 2.1, es el de utilizar algoritmos basados en reglas y el segundo, que describimos en 2.2, el de utilizar técnicas de aprendizaje automático.

2.1. Algoritmos de detección de marcadores de la negación basados en reglas

Los algoritmos basados en reglas presentan dos objetivos fundamentales: el primero, detectar los marcadores de negación y el segundo, definir el alcance de la negación. Esto es necesario para extraer correctamente qué entidades están realmente negadas en el discurso. Así, aunque nuestro sistema de detección de marcadores de negación no tiene en cuenta el alcance, muchos de los algoritmos creados para detectar marcadores sí tienen también como objetivo detectar el alcance de la negación.

A continuación, describimos los algoritmos de detección de marcadores de negación basados en reglas más utilizados. Este tipo de algoritmos se pueden dividir, a su vez, en algoritmos de tipo léxico (2.1.1) o de tipo sintáctico, que incorporan información generada por analizadores sintácticos (2.1.2).

2.1.1. Algoritmos basados en reglas de tipo léxico

En este apartado, describimos los algoritmos de detección de marcadores de negación basados en léxico más relevantes: NegEx, ConText y pyConTextNLO. NegEx es uno de los más populares. Desarrollado por Chapman et al. (2001), se diseñó para detectar qué términos médicos aparecen negados en informes de alta hospitalaria. Utiliza una lista de disparadores (*triggers*) que indican la presencia de negación.

En primer lugar, NegEx extrae términos léxicos y determina si estos están afectados por la negación léxica (esto es, por términos como *not*), si son una pseudonegación (‘no se detecta

1 Ejemplo extraído de la *Nueva Gramática de la Lengua Española* (Real Academia Española, 2009: 3638).

incremento’) o si son conjunciones (‘salvo’, ‘excepto’). Para detectar si una entidad o término está negado, los autores determinan qué disparadores deben ser considerados como marcadores de negación. En segundo lugar, para determinar el alcance de la negación, aplican dos expresiones regulares simples. NegEx obtiene, en un corpus formado por unos 1.000 informes y 1.235 conceptos extraídos manualmente por médicos, una especificidad (*specificity*)² del 94,51% y una precisión (*precision*) del 84,5%.

NegEx es un algoritmo desarrollado para el inglés. Posteriormente se ha adaptado a otras lenguas. Chapman et al. (2013) fueron los primeros en adaptar este algoritmo al sueco, francés y alemán traduciendo la lista de disparadores del inglés a estas lenguas. Comprueban que, con la excepción de los marcadores de negación *no*, *not* y *without*, que son frecuentes en las cuatro lenguas, la lista de marcadores de negación y sus frecuencias de aparición son muy diferentes en cada traducción.

Los resultados obtenidos por las distintas adaptaciones de NegEx son positivos, aunque bastante variables. Así, la traducción al sueco obtiene una cobertura (*recall*) del 82% y una precisión del 75%, utilizando un corpus de entrenamiento formado por informes de autoevaluación extraídos del Stockholm EPR. En francés, NegEx obtiene mejores resultados (cobertura del 85% y precisión del 89%) con un corpus de apuntes sobre cardiología. Ambos, sin embargo, muestran una diferencia importante respecto al inglés, sobre todo en la precisión (la versión sueca del NegEx obtiene un 9,3% menos de precisión y la francesa un 4,4% menos). Los autores no evalúan su adaptación al alemán.

Cotik et al. (2016) proponen su propia traducción del NegEx al alemán y la evalúan con dos corpus, uno de apuntes clínicos y otro de informes de alta hospitalaria. Amplían la lista de disparadores de NegEx y añaden expresiones regulares para cubrir una mayor variedad de enunciados. Este modelo obtiene una precisión del 88% en los informes de alta y del 95% en los apuntes clínicos; una cobertura del 93% y 97%, respectivamente, y un valor-f del 91% y 96%. También evalúan la adaptación al alemán elaborada por Chapman et al. (2013) con los mismos corpus. El modelo de Chapman et al. (2013) obtiene un valor-f en los informes de alta del 72% y en los apuntes clínicos, del 52%. El rendimiento del sistema de Cotik et al. (2016) es, por tanto, superior.

Costumero et al. (2014) adaptan y traducen NegEx al español y lo usan para detectar la negación en textos clínicos. Traducen la lista de disparadores incluida en NegEx y la completan con una lista de marcadores de negación creada a partir de la anotación manual de un corpus. Su algoritmo está programado en Java y ha sido evaluado con 500 informes obtenidos del corpus SciELO (Packer 1999), con un total de 1.164.712 palabras, de las cuales 65.605 son diferentes, es decir, no repetidas. El sistema obtiene una precisión del 49,47%, una cobertura del 55,50% y un valor-f del 52,38% cuando utiliza la lista de términos negados como lista de marcadores de

2 Chapman et al. (2001:305) definen la especificidad como el resultado de dividir el número de términos no negados que NegEx acierta entre el número de términos no negados por el anotador humano.

negación. Los autores concluyen que las frecuencias de aparición de los disparadores de NegEx en inglés y en español son muy distintas y que sería necesario ampliar el corpus del español para encontrar más ejemplos de disparadores frecuentes en inglés que no han podido documentar en español.

ConText (Harkema et al., 2009) es una extensión del algoritmo NegEx diseñado para detectar no solamente marcadores de negación sino también otros modificadores semánticos que expresan nociones como la incertidumbre (ConText detecta si una afirmación es una hipótesis o no), el experimentador (esto es, si se refiere al propio paciente o no) y la referencia temporal (si se refiere a un hecho histórico o presente). ConText ha sido evaluado en un corpus de evaluación de 120 informes de seis tipos distintos con una precisión promedia del 94% y una cobertura del 92%. Se trata de un algoritmo que presenta una precisión similar a NegEx para detectar la negación y que, además, permite extraer otro tipo de información diferente, como la incertidumbre o la temporalidad.

Zhang (2016) propone una adaptación de ConText al español. Traduce los disparadores de negación y obtiene una lista de posibles disparadores mediante el etiquetado de 454 oraciones, extraídas del corpus SciELO (Parker, 1999), que contienen 1.897 disparadores, de entre los cuales 227 son verdaderos marcadores de negación. Este modelo obtiene una precisión del 83,8%, un valor-f del 86,1% y una exactitud³ del 96,1% en la detección de marcadores de negación.

Por último, pyConTextNLP es una adaptación del algoritmo ConText que convierte el código fuente, originalmente escrito en Java, a Python. Su principal ventaja es que permite añadir algunas expresiones regulares no disponibles para ConText.

2.1.2. Algoritmos que incorporan conocimiento sintáctico

Enunciados como (7) ilustran las limitaciones de un algoritmo que solo aplique reglas léxicas (Manimaran & Velmurugan, 2017:202):

7. There is **no** evidence of cervical lymph node enlargement.

En (7) no se niega que exista un nódulo linfático cervical. Por el contrario, al negar un aumento, se afirma implícitamente que existe el nódulo. Un algoritmo de tipo léxico como NegEx no es capaz de captar este matiz y detectaría *no* como un marcador de negación. Con el objetivo de evitar estas confusiones, varios investigadores incorporaron conocimiento sintáctico para enriquecer los algoritmos de detección de marcadores de negación.

3 Zang (2016:36) define la exactitud como el resultado de dividir los verdaderos positivos (VP) más los verdaderos negativos entre la suma de los VP y VN más los falsos positivos (FP) y falsos negativos (FN).

Una de las propuestas más destacadas es la de Mehrabi et al. (2015), que presenta el algoritmo DEEPEN (de DepEndency ParsEr Negation), que integra un analizador de dependencias basado en transiciones, concretamente el Standford Dependency Parser (SDP). DEEPEN extrae del corpus los conceptos que aparecen negados según NegEx y, posteriormente, el analizador sintáctico extrae información sobre la relación de dependencia entre los términos negados y el concepto. Si un término no es detectado como negado por NegEx nunca será analizado por DEEPEN. Los autores han probado este algoritmo en dos corpus diferentes y han obtenido una precisión del 89% y 97%, respectivamente, y una cobertura del 74% y 96%.

Gkotsis et al. (2016) desarrollan una aplicación de detección de la negación que utiliza la gramática no contextual probabilística (en inglés, la PCFG, de *Probabilistic Context-Free Grammar*). Entrenan su modelo con 6.000 frases anotadas manualmente procedentes de un corpus formado por informes de salud mental. Obtienen una precisión del 89,4%, una cobertura del 94,6% y un valor-f del 91,9%.

Por último, cabe destacar otros algoritmos de detección de marcadores de negación basados en reglas de tipo léxico pero que incluyen información sintáctica como son NegExpander (Aronow et al., 1999), NegFinder (Mutalik et al., 2001) y NegHunter (Gindl et al., 2008), todos ellos inspirados por el modelo propuesto en NegEx. NegFinder es un sistema desarrollado por Mutalik et al. (2001) que define marcadores léxicos y los combina con información sintáctica de gramáticas no contextuales cuando la distancia entre el marcador de negación y el final de la estructura negativa es demasiado grande, uno de los casos que NegEx no puede resolver por sí solo. NegFinder obtuvo una cobertura del 95.7% y una precisión del 91.8% al ser evaluado con 1869 conceptos extraídos de 10 documentos.

DepNeg (Sohn et al., 2012) utiliza el contexto sintáctico entre el marcador de negación y el final de la estructura negativa para mejorar la detección. Consigue una cobertura del 73.9% y una precisión del 96.6% en un corpus formado por 160 informes médicos de Mayo Clinic. Lo más destacable es que reduce el número de falsos positivos generados por NegEx.

La tabla 1 muestra los resultados de precisión y cobertura obtenidos por los principales algoritmos de detección de marcadores de negación, tanto de los basados en información léxica como de los que incorporan información sintáctica. No obstante, no se puede establecer una comparación válida entre los diferentes sistemas, ya que cada uno ha sido evaluado con corpus y metodologías distintas:

Algoritmo	Lengua	Corpus	Precisión	Cobertura
NegEx (Chapman et al., 2001)	inglés	1050 informes de alta hospitalaria	84,5%	--
NegEx (Chapman et al., 2013)	francés	apuntes sobre cardiología	89%	85%
NegEx (Chapman et al., 2013)	sueco	Stockholm EPR	75%	82%
NegEx (Chapman et al., 2013 (diseño); Cotik et al. (2016) (evaluación))	alemán	informes de alta hospitalaria (1) apuntes clínicos (2)	0,88% (1) 0,92% (2)	0,61% (1) 0,36% (2)
NegEx (Cotik et al., 2016)	alemán	informes de alta hospitalaria (1) apuntes clínicos (2)	0,88% (1) 0,92% (2)	0,93% (1) 0,97% (2)
NegEx (Costumero et al., 2014)	español	500 informes del corpus SciELO	83,37%	55,50%
ConText (Harkema et al., 2009)	inglés	120 informes médicos	94%	92%
ConText (Zang et al., 2016)	español	SciELO	83,8%	--
DEEPEN (Mehrabi et al., 2015)	inglés	159 informes médicos de Mayo Clinic	98%	99%
Gkotsis et al. (2016)	inglés	5000 expedientes de salud mental	89,4%	94,6%
NegFinder (Mutalik et al., 2001)	inglés	BioScope	91,8%	95,7%
DepNeg (Sohn et al., 2012)	inglés	160 informes médicos de Mayo Clinic	96,6%	73,9%

Tabla 1. Resultados de precisión y cobertura obtenidos por los principales algoritmos de detección de marcadores de negación basados en reglas

2.2. Sistemas de detección de la negación basados en aprendizaje automático

En el área del PLN, el aprendizaje automático ha tenido un crecimiento exponencial en la última década, ya que usar técnicas de este tipo ofrece la ventaja de construir modelos que son más extrapolables a otros conjuntos de datos y no tan diseñados *ad hoc* para unos datos y tareas en concreto, así como una mayor escalabilidad en comparación con los algoritmos basados en reglas.

En los siguientes apartados abordamos tanto la descripción de los corpus anotados con negación más utilizados (apartado 2.2.1.), en inglés y en español, así como las propuestas más destacadas en el ámbito de la detección automática de la negación desde la perspectiva del aprendizaje automático, supervisado y no supervisado (apartado 2.2.2).

2.2.1. Corpus anotados con negación

La biomedicina es el ámbito que más corpus anotados con negación ha proporcionado. No obstante, existen corpus en otros ámbitos. La mayoría están disponibles para el inglés, pero cada vez se están anotando más corpus con negación en otras lenguas.

A continuación, describimos los corpus más relevantes, tanto para el inglés como para el español. Esta última es la lengua que más nos interesa, puesto que nuestra propuesta utiliza datos del español, pero es indudable que sin el trabajo previo que han hecho los investigadores en inglés no habría un punto de partida para el análisis de la negación en español.

2.2.1.1. Corpus en inglés anotados con negación

Los primeros textos anotados con marcadores de negación pertenecen al ámbito de la biomedicina. BioInfer (Pyysalo, 2007) cuenta con 1.100 frases extraídas de artículos de investigación biomédica y está anotado con entidades nombradas (*named entities*) y dependencias sintácticas. BioScope (Vincze et al., 2008) es otro corpus de biomedicina que incluye tres subcorpus: uno de documentos clínicos, otro de artículos científicos y otro de resúmenes de artículos. Está anotado a nivel de palabra con palabras clave que indican negación o especulación y su correspondiente alcance. Los autores extraen los textos de cuatro fuentes diferentes e incluyen informes médicos, en concreto, informes de radiología, para evitar que todos los datos del BioScope pertenezcan al ámbito de la biología. En total, contiene 20.924 oraciones anotadas manualmente, extraídas de un total de 3.236 documentos. La negación representa un 6.6% del total de las oraciones en los documentos clínicos, un 13,76% de los artículos científicos y un 13,45% de los resúmenes. BioScope es un corpus de libre acceso que se puede descargar gratuitamente para llevar a cabo trabajos de investigación desde su página web (www.inf.u-szeged.hu/rgai/bioscope).

También dentro de la biomedicina encontramos el corpus ChestX-ray (Peng et al., 2018), que incluye 900 informes de radiología que proceden de un corpus más amplio de un hospital de EE.UU. Este corpus ha sido creado para desarrollar la herramienta de detección de marcadores de negación NegBio.

El interés por la negación ha facilitado que equipos de investigación que trabajan en ámbitos muy distintos al biomédico también recopilen corpus y los anoten con información sobre negación. Así, Banjade et al. (2016) crean el corpus DT-Neg, un conjunto de datos que contiene las respuestas que proporcionan estudiantes de secundaria en el sistema de tutorías DeepTutor. Este sistema plantea problemas de física conceptuales que los alumnos deben responder de forma interactiva. El corpus contiene un total de 27.785 respuestas y 2.603 respuestas que contienen al menos un marcador de negación. La negación, por tanto, representa un 9,36% del corpus.

También cabe destacar el corpus ConanDoyle-Neg anotado con marcadores de negación (Morante & Daelemans, 2015) creado con el objetivo de aportar nuevos datos no relacionados con la biomedicina al estudio de la negación. Consta de un corpus de entrenamiento y otro de evaluación. Incluye dos novelas⁴ de Sir Arthur Conan Doyle anotadas con marcadores de negación y su alcance. El corpus presenta un total de 4.423 oraciones (3.640 en HB y 783 en WL), de las cuales 995 contienen negación (850 en HB y 145 en WL). Es decir, que la negación representa un 27,3% del corpus.

Por último, debemos mencionar el SFU Review Corpus (Konstantinova et al., 2013). Consta de 400 reseñas, sobre ocho temas diferentes (libros, coches, ordenadores, menaje, hoteles, películas, música y teléfonos). Presenta un total de 17.263 oraciones, de las cuales un 18% contienen negación. Es el precursor del SFU Review_{SP}-NEG.

2.2.1.2. Corpus en español anotados con negación

Recientemente, se han anotado corpus en lenguas diferentes al inglés, como el chino (Kang et al., 2017, para un corpus biomédico anotado para el análisis de la negación) y el español. Hay que destacar en esta última lengua el valor del corpus proporcionado para el taller NEGES, el SFU Review_{SP}-NEG, que describiremos en más profundidad en el apartado 3.

NewsCom (Taulé et al., en prensa) es un corpus desarrollado por el CLiC en la Universitat de Barcelona. Contiene comentarios de noticias, que se agrupan en nueve temáticas (inmigración, política, tecnología, terrorismo, economía, sociedad, religión, refugiados e inmobiliaria). Presenta un total de 1.500 comentarios, de los cuales 888 contienen al menos una estructura negativa. Ha sido manualmente anotado con los marcadores de negación, el alcance y

4 En concreto, se trata de *The Hound of the Baskervilles* (HB) y *The Adventure of Wisteria Lodge* (WL). Ambas obras son de libre acceso y están disponibles en el Proyecto Gutenberg (<http://www.gutenberg.org/browse/authors/d/#a37238>).

el foco. Es, por el momento, el único corpus en español que ofrece una anotación del foco de la negación.

2.2.2. Propuestas de modelos de detección de la negación basadas en aprendizaje automático

Los primeros modelos de detección de marcadores de negación basados en técnicas de aprendizaje automático las encontramos en trabajos que utilizan corpus en inglés (Morante et al., 2008; Morante & Daelemans, 2009; Goldin & Chapman, 2003, entre otros). Varios de estos trabajos se centraron en el análisis de corpus biomédicos, en especial de BioScope. Gracias a la organización en 2018 y 2019 del taller NEGES, organizado por la SEPLN en Sevilla (2018) y próximamente en Bilbao (2019), se pueden encontrar modelos de detección automática de marcadores de negación que usan aprendizaje automático con datos procedentes de un corpus en español que no está relacionado con la biomedicina.

En este apartado describimos algunos de los sistemas propuestos para detectar automáticamente los marcadores de negación y su alcance utilizando corpus en inglés (2.2.2.1) y, a continuación, revisamos los resultados obtenidos por aproximaciones al aprendizaje automático aplicado al español (2.2.2.2). En concreto, describimos dos propuestas presentadas en el NEGES 2018, una basada en Conditional Random Field (CRF) y otra en redes neuronales.

2.2.2.1. Aprendizaje automático para detectar marcadores de negación en inglés

Morante et al. (2008) aplican métodos de aprendizaje automático para la detección del alcance de la negación en textos biomédicos. Desarrollan dos modelos para detectar marcadores y el alcance de la negación. Su detector de marcadores es un clasificador que predice si un elemento es marcador de negación o no y, si lo es, si es el primer elemento o la continuación de un marcador de negación.

Como primer paso, Morante et al. (2008) desarrollan dos bases de referencia (*baseline*). La primera es muy sencilla y consiste en aplicar la etiqueta NEG a los lemas *no* y *not*. Solo con esta técnica, consiguen reconocer el 72.80% de los casos. Para la segunda base de referencia, añaden una lista más extensa de expresiones negativas e incluyen términos como *neither*, *unable* o *without* y consiguen reconocer el 85.85% de los marcadores. Los resultados de estas bases de referencia son elevados. La primera obtiene una precisión del 90,42%, una cobertura del 72,80% y un valor-f del 80,66%. La segunda obtiene un 89,77%, un 93,48% y un valor-f del 91,54%. Por tanto, la segunda base de referencia mejora la cobertura y el valor-f de la primera, lo que resulta lógico, puesto que utiliza un vocabulario más extenso. No obstante, la primera obtiene mayor precisión.

A continuación, desarrollan un sistema, basado en técnicas de aprendizaje automático, para comparar sus resultados con las bases de referencia. Este modelo obtiene un valor-f del 94.40%, una cobertura del 94,59% y una precisión del 94,21% en la detección de marcadores de

negación. Esto supone una mejora respecto a las bases de referencia de casi un 13% respecto a la primera y algo más de un 2% respecto a la segunda. No obstante, los resultados varían mucho según el marcador de negación. Los menos frecuentes obtienen peores resultados que los que muestran una frecuencia más elevada.

Morante & Daelemans (2009) parten del modelo propuesto en Morante et al. (2008). Utilizan el corpus BioScope para diseñar dos clasificadores distintos, uno para la identificación de marcadores de negación y otro para detectar el alcance. Para ello, utilizan técnicas de aprendizaje automático supervisado. El sistema mejora los resultados de la base de referencia y obtiene un valor-f del 99,37% en el primer corpus, del 97,81% en el segundo y de un 99,03% en el tercero. La cobertura es de un 98,75%, un 95,72% y un 98,09%, respectivamente.

Goldin & Chapman (2003) aplican técnicas de aprendizaje automático para detectar si un término médico aparece negado por el marcador *not* o *no*. Utilizan un corpus de 207 frases, extraídas de informes médicos, que contienen negación. Como métodos, aplican los árboles de decisión y *Naive Bayes* y alcanzan un valor-f del 90%.

Agarwal & Yu (2010) crean un sistema de detección de marcadores de negación y su alcance basado en CRF. Obtienen un valor-f del 98% para la detección de marcadores y de un 95% para la del alcance en textos clínicos y una cobertura del 97% y 85%, respectivamente, en artículos de investigación biomédica. Estos resultados son superiores a los de NegEx y a los del sistema creado por Morante & Daelemans (2009).

Cruz et al. (2016) elaboran un sistema de detección de los marcadores de negación basado en métodos de aprendizaje automático. Este sistema detecta también valores especulativos. Para entrenar su modelo, utilizan una parte del corpus BioScope. Crean un modelo que, como el de Morante et al. (2008) y Morante & Daelemans (2009), realiza dos tareas de clasificación consecutivas. Prueban dos técnicas de aprendizaje automático diferentes. Implementan una Support Vector Machine (SVM) para la detección de los marcadores de especulación y utilizan un árbol de decisiones C4.5, una técnica basada en *Naive Bayes*, para detectar marcadores de negación. Obtienen una precisión del 96,5%, una cobertura del 98% y un valor-f del 97,3% en la detección de marcadores de negación.

Por último, comparan los resultados de su sistema con los de Morante & Daelemans (2009) y los de NegEx porque son los únicos de libre acceso. Hay que destacar que NegEx obtiene un valor-f del 65,6%, bastante por debajo de Morante & Daelemans (2009) y de Cruz et al. (2016). La conclusión final es que los modelos basados en técnicas de aprendizaje automático obtienen mejores resultados incluso en comparación con algoritmos basados en reglas tan sofisticados como NegEx.

2.2.2.2. Aprendizaje automático para detectar marcadores de negación en español: NEGES

Con motivo de la organización de la Tarea 2 en el contexto del taller NEGES 2018, se presentaron propuestas para abordar la detección automática de los marcadores de negación en español. Las dos propuestas presentadas aplicaron métodos de aprendizaje automático (redes neuronales y CRF) para resolver la tarea.

Fabregat et al. (2018) obtuvieron el segundo mejor resultado desarrollando un sistema basado en redes neuronales recurrentes (RNN, del inglés *Recurrent Neural Network*), concretamente en una LSTM (*Long Short-Term Memory*). Para construir su sistema, los autores utilizan la librería Keras de Python y Tensor Flow. Plantean la tarea como un problema de etiquetado de secuencias y utilizan las etiquetas BIO para clasificar las palabras como inicio de un marcador de negación, continuación o elemento que queda fuera del marcador. Para evaluar mejor el modelo, diseñan una base de referencia que simplemente comprueba si las palabras están incluidas en una lista de marcadores de negación. El sistema obtiene una precisión promedia del 91,47%, una cobertura del 82,17% y un valor-f del 86,44%, unos resultados muy próximos a los del equipo UPC-TALP y supera en todos los dominios los obtenidos por la base de referencia.

Loharja et al. (2018) desarrollan un sistema basado en aprendizaje automático, pero con un enfoque diferente, basado en CRF. Con esta técnica, pretenden mejorar la detección de marcadores discontinuos, que es precisamente uno de los grandes desafíos de la tarea. En estos casos, además de identificar los marcadores de negación, el sistema debe ser capaz de reconocer cuál es el elemento inicial y cuál la continuación del mismo marcador y no identificarlos como marcadores independientes. Los autores desarrollan un clasificador que decide si cada palabra que procesa forma parte o no de un marcador de negación y si la respuesta es positiva, la etiqueta como inicio (“B”, *begin*) o continuación (“I”, *inside*). Si es negativa, la etiqueta como “O” (*outside*).

Utilizan la librería NLTK para Python 3. En concreto, usan un módulo de NLTK para aplicar un etiquetado de CRF utilizando como núcleo de este etiquetador Python CRF Suite, ya que su aproximación a la tarea se basa en el reconocimiento de entidades nombradas. En una primera etapa, entrenan este etiquetado CRF con la parte de entrenamiento del corpus de desarrollo. En una segunda etapa, ponen a prueba el etiquetador con el corpus de evaluación. Obtienen una precisión del 91,48%, una cobertura del 82,18% y un valor-f del 86,45% en el corpus de evaluación, que de hecho es el valor-f más elevado de entre todas las propuestas presentadas al NEGES 2018. La cobertura resulta bastante más baja que el valor-f y, en algunos ámbitos, es inferior al 80%. Esto se debe a un número relativamente elevado de falsos negativos ocasionados por las dificultades que presenta el modelo para detectar el inicio y la continuación de los marcadores negativos en los casos de discontinuidad.

La tabla 2 muestra una comparativa de los resultados de precisión y cobertura de los diferentes sistemas de aprendizaje automático (supervisado y no supervisado) descritos en este apartado. Nuevamente, no podemos establecer una verdadera comparación, puesto que cada sistema ha sido evaluado con corpus y métodos diferentes:

Algoritmo	Lengua	Corpus	Precisión	Cobertura
Morante et al. (2008)	inglés	textos biomédicos	94,21%	94,59%
Morante & Daelemans (2009)	inglés	BioScope (tres subcorpus)	100% (en los tres corpus)	98,75% (1) 95,72% (2) 98,09% (3)
Agarwal & Yu (2010)	inglés	artículos de investigación biomédica	97%	98%
Cruz et al. (2016)	inglés	BioScope	96,5%	98%
Fabregat et al. (2018)	español	SFU Review _{SP} -NEG	91,47%	82,17%
Loharja et al. (2018)	español	SFU Review _{SP} -NEG	91,48%	82,18%

Tabla 2. Resultados de precisión y cobertura de los sistemas de aprendizaje automático más destacados

Los resultados tan positivos logrados por los sistemas basados en aprendizaje automático (supervisado y no supervisado) nos han hecho decantarnos por estos métodos a la hora de desarrollar nuestro sistema de detección de marcadores de negación. En concreto, nos basamos en los modelos que utilizan el CRF, como es el caso de Agarwal & Yu (2010) y Loharja et al. (2018), puesto que es un enfoque idóneo para las tareas de etiquetado. La prueba de que es una metodología útil es que el CLiC-Neg ha quedado como primer clasificado en la subtarea A del taller NEGES 2019.

3. Metodología para la detección automática de marcadores de negación

El modelo de detección automática de la negación CLiC-Neg, ha sido desarrollado para competir en la subtarea A del taller NEGES 2019. Esta tarea, descrita pormenorizadamente en Jiménez-Zafra et al. (2019) consiste en crear un sistema para detectar de manera automática los marcadores de negación presentes en el SFU Review_{SP}-NEG.

Los resultados de los sistemas presentados a la competición son evaluados con las medidas de precisión, cobertura y valor-f. Para obtener estas medidas, se utilizan los *scripts* de evaluación del *SEM2012 Shared Task -Resolving the Scope and Focus of Negation (Morante and Blanco, 2012), proporcionados por la propia organización del NEGES 2019 (Jiménez-Zafra et al., 2018b:35).

Los criterios de evaluación consideran como verdadero positivo (VP) los casos en los que el sistema detecta que una oración contiene negación e identifica correctamente todos los marcadores de negación. Como falsos positivos (FP), se consideran los casos en los que el sistema etiqueta erróneamente los marcadores de negación. Por último, un falso negativo (FN) se da cuando el sistema no detecta un marcador de negación. Morante & Blanco (2012) añaden los siguientes criterios a la evaluación:

- Se ignoran los signos de puntuación.
- Las coincidencias parciales no se consideran un FP, sino que se clasifican como FN. De esta manera, no se penaliza más la identificación parcial que la no identificación de un marcador.

A continuación, exponemos la metodología utilizada para obtener los datos y el proceso que hemos seguido para desarrollar y entrenar el modelo de detección de marcadores de negación CLiC-Neg.

En primer lugar, presentamos el marco del trabajo (apartado 3.1), esto es, todos los aspectos relacionados con la organización del taller NEGES 2019: el corpus proporcionado (apartado 3.1.1), algunos aspectos del preprocesado del corpus, como son el formato en el que este ha sido presentado (apartado 3.1.2) y, por último, el sistema de etiquetado que hemos usado para identificar los marcadores de negación (apartado 3.1.3).

Por último, explicamos las etapas en el proceso de creación del sistema CLiC-Neg (apartado 3.2), desde la elaboración de una base de referencia (3.2.1), hasta la creación del modelo basado en CRF (3.2.2), el análisis de errores generados por este sistema (3.2.3) y las posteriores versiones de este modelo, que añaden reglas y listas de marcadores de negación. Estas dos últimas versiones reciben el nombre del sistema CLiC-Neg. Al final del apartado 3.2.4. se

muestra una comparativa de los resultados obtenidos por las diferentes versiones del sistema con la parte de evaluación del corpus de desarrollo.

3.1. Marco del trabajo

En el contexto del NEGES 2019, los organizadores proporcionan el corpus de desarrollo, en formato CoNLL, para que los participantes desarrollen sus sistemas. Finalmente, evalúan las propuestas enviadas con un corpus de evaluación que los sistemas no han visto nunca anteriormente. A continuación, describimos los aspectos relacionados con la subtarea A del NEGES 2019 y que constituyen los datos con los que hemos creado nuestro sistema.

3.1.1. El corpus SFU Review_{SP}-NEG

El corpus proporcionado para la tarea es el SFU Review_{SP}-NEG. Este corpus fue desarrollado para mejorar el tratamiento computacional de la negación en español (Jiménez-Zafra et al., 2018a:242) y ha sido anotado también con elementos de polaridad para facilitar el análisis de sentimientos. Sus contenidos son similares a los del SFU Review Corpus (véase el apartado 2.2 para más detalles), aunque a diferencia de la versión en inglés, que contiene anotaciones sobre los marcadores de negación y especulación con su correspondiente alcance, este ha sido anotado con marcadores de negación, el alcance, el evento y cómo se ve afectada la polaridad de las palabras que quedan dentro del alcance de la negación.

El corpus consta de 400 comentarios de coches, hoteles, lavadoras, libros, teléfonos móviles, ordenadores, música y películas que han sido extraídos de la web Ciao.es. Cada uno de estos dominios contiene 25 reseñas positivas y 25 negativas. Cada reseña ha sido anotada con el lema y la categoría morfosintáctica y, al nivel de la oración, con los marcadores de negación, el alcance, el evento y la polaridad.

Los datos estadísticos del corpus (Jiménez-Zafra et al., 2018a:244) muestran que el 96,75% de los documentos del SFU Review_{SP}-NEG contienen negación. Esto es, 387 de las 400 reseñas contienen al menos una estructura negativa y solamente 13 no presentan ninguna negación.

El corpus contiene un total de 9.455 oraciones, de las cuales 3.022 contienen al menos una estructura negativa y un total de 4.329 estructuras negativas. En total, estas estructuras suman 221.866 palabras. Han sido anotadas manualmente y clasificadas por su valor semántico como oraciones negativas (“neg”), de contraste u oposición entre dos términos (“contrast”), de comparación (“comp”) y de no negación (“noneg”). La tabla 3 muestra la distribución de estos cuatro valores:

3. Metodología para la detección automática de marcadores de negación

Dominio	neg	contrast	comp	noneg	Total
Libros	806 (88,09%)	65 (7,10%)	4 (0,44%)	40 (4,37%)	915 (100%)
Coches	324 (91,79%)	12 (3,40%)	7 (1,98%)	10 (2,83%)	353 (100%)
Teléfono móviles	501 (94,98%)	6 (1,14%)	1 (0,19%)	20 (3,79%)	528 (100%)
Ordenadores	305 (93,27%)	13 (3,97%)	1 (0,31%)	8 (2,45%)	327 (100%)
Hoteles	361 (90,93%)	8 (2,02%)	2 (0,50%)	26 (6,55%)	397 (100%)
Películas	936 (89,93%)	48 (4,61%)	10 (0,96%)	48 (4,60%)	1.042 (100%)
Música	373 (92,33%)	15 (3,71%)	3 (0,74%)	13 (3,22%)	404 (100%)
Lavadoras	337 (92,84%)	8 (2,20%)	2 (0,55%)	16 (4,41%)	363 (100%)
Total	3.943 (91,08%)	175 (4,04%)	30 (0,70%)	181 (4,18%)	4.329 (100%)

Tabla 3. Distribución de los valores de las estructuras negativas por dominio en el corpus SFU Review_{SP}-NEG⁵

Para la anotación del corpus, se utilizó la herramienta de anotación AncoraPipe (Martí et al., 2007). La tabla 4 muestra las estadísticas del corpus SFU Review_{SP}-NEG, con el número de oraciones por documento y de palabras por oración:

Dominio	#Documentos positivos	#Documentos negativos	#Oraciones	Promedio de oraciones por documento	#Palabras	Promedio de palabras por oración
Libros	25	25	1.840	36,80	42.171	22,92
Coches	25	25	756	15,12	18.696	24,73
Teléfonos móviles	25	25	1.022	20,44	23.285	22,78
Ordenadores	25	25	652	13,04	16.554	25,39
Hoteles	25	25	859	17,18	19.223	22,38
Películas	25	25	2.472	49,44	59.677	24,14
Música	25	25	953	19,06	23.928	25,11
Lavadoras	25	25	901	18,02	18.322	20,35
Total	200	200	9.455	23,64	221.866	23,47

Tabla 4. Estadísticas del corpus SFU Review_{SP}-NEG

5 Extraído de Jiménez-Zafra et al. (2018b:25).

Por último, los organizadores del NEGES 2019 proporcionan, como corpus de desarrollo, solamente una parte del corpus SFU Review_{SP}-NEG. Un 80% del corpus de desarrollo lo usamos para entrenar el sistema (corpus de entrenamiento) y un 20% para evaluar nuestros diferentes sistemas de detección de marcadores de negación (corpus de evaluación). Así podemos hacernos una idea del funcionamiento de los sistemas que elaboramos.

Una vez se envía el sistema de detección de marcadores a competición, los organizadores evalúan los sistemas de todos los participantes con un nuevo corpus de evaluación que no ha sido visto por los sistemas durante el desarrollo. La tabla 5 muestra el número de oraciones y de estructuras negativas en cada uno de estos conjuntos de datos:

Corpus	Reseñas	Oraciones	Con negación	Sin negación
Desarrollo (entrenamiento)	264	7863	3862	4001
Desarrollo (evaluación)	56	1314	423	891
Evaluación	80	2203	627	1576

Tabla 5. Estadísticas del corpus de desarrollo y evaluación del taller NEGES

3.1.2 Formato del corpus

La organización del taller NEGES 2019 facilita el corpus de desarrollo en el formato CoNLL (Farkas et al., 2010). En este formato, cada línea contiene una palabra no repetida (un *token*) y varias columnas que incluyen información gramatical. Esta incluye el lema, los rasgos morfológicos y la anotación referente a la negación, en el caso de la parte de entrenamiento del corpus de desarrollo, pero no en el de evaluación. En la figura 1, se muestra una línea del corpus en formato CoNLL correspondiente a la oración ‘y no conllevan ninguna molestia al llevarlo en el bolsillo de los pantalones o en el bolso’, que contiene dos marcadores de negación:

3. Metodología para la detección automática de marcadores de negación

moviles_yes_4_1	36	14	y	y	cc	coordinating	-	-	-
moviles_yes_4_1	36	15	no	no	rn	negative	no	-	-
moviles_yes_4_1	36	16	conlle van	conlle var	vmip3p0	main-	-	-	-
moviles_yes_4_1	36	17	ninguna	ninguno	di0fs0	indefinite	ninguna-	-	-
moviles_yes_4_1	36	18	molestia	molestia	ncfs000	common	-	-	-
moviles_yes_4_1	36	19	a	a	sps00	preposition	-	-	-
moviles_yes_4_1	36	20	el	el	da0ms0	article	-	-	-
moviles_yes_4_1	36	21	lle var	lle var	vmn0000	main	-	-	-
moviles_yes_4_1	36	22	lo	lo	pp3cna00	personal	-	-	-
moviles_yes_4_1	36	23	en	en	sps00	preposition	-	-	-
moviles_yes_4_1	36	24	el	el	da0ms0	article	-	-	-
moviles_yes_4_1	36	25	bolsillo	bolsillo	ncms000	common	-	-	-
moviles_yes_4_1	36	26	de	de	sps00	preposition	-	-	-
moviles_yes_4_1	36	27	los	el	da0mp0	article	-	-	-
moviles_yes_4_1	36	28	pantalones	pantalón	ncmp000	common	-	-	-
-	-	-	-	-	-	-	-	-	-
moviles_yes_4_1	36	29	o	o	cc	coordinating	-	-	-
moviles_yes_4_1	36	30	en	en	sps00	preposition	-	-	-
moviles_yes_4_1	36	31	el	el	da0ms0	article	-	-	-
moviles_yes_4_1	36	32	bolso	bolso	ncms000	common	-	-	-
moviles_yes_4_1	36	33	.	.	fp	-	-	-	-

Figura 1. Ejemplo de una oración en formato CoNLL

El primer elemento de la frase es el identificador del documento (columna 1), ‘moviles_yes_4_1’, encabezado por el título del dominio (en este caso, «Móviles»). A continuación, en la columna 2 aparece el índice de la frase (36) y en la columna 3 la posición que ocupa la palabra en la frase. En el caso de expresiones multipalabra, como, por ejemplo, ‘por supuesto’, aparecen escritas con guion bajo y solo tienen un índice:

```
películas_no_2_2013    1    Por_supuesto    por_supuesto    rg    - - - -
```

En la columna 4, aparece la palabra, seguida en las columnas 5 y 6 por el lema y la etiqueta que sintetiza la información referente a la categoría morfosintáctica (ncfs000, vmp00sm), incluyendo los rasgos más generales, y, por último, en la columna 7, la categoría gramatical, como artículo, auxiliar o verbo principal. El marcador de negación ‘no’ aparece anotado en el corpus de desarrollo como *negative* en la columna 8, pero en cambio ‘ninguna’ aparece solo con la etiqueta de ‘indefinido’. Esta es la manera de marcar, en formato CoNLL, que ‘ninguna’ es la continuación del marcador de negación y que ‘no’ es el inicio del marcador de negación.

Los organizadores del taller NEGES (2018 y 2019) facilitan como corpus de desarrollo una parte del SFU Review_{SP}-NEG en formato CoNLL que solo incluye la información referente al lema y la categoría morfosintáctica. Frases como las que hemos mostrado en formato CoNLL, son el *input* de nuestro sistema de detección de marcadores de negación durante la etapa de

entrenamiento. Para la evaluación del sistema, usamos como *input* la parte de evaluación del corpus de desarrollo que, a diferencia del ejemplo en formato CoNLL que hemos visto, no tiene anotados los marcadores de negación con la etiqueta *negative*.

3.1.3. Etiquetas BIO para la identificación de marcadores de negación

Hemos anotado los marcadores de negación con las etiquetas BIO (Ramshaw & Marcus, 1999:6). El elemento inicial de un marcador complejo o un marcador simple será etiquetado como “B” (*begin*, es decir, inicio), el resto de elementos que aparecen dentro de la misma expresión negativa serán etiquetadas como “I” (*Inside*, dentro) y el resto de elementos que no forman parte de ninguna expresión negativa serán identificados como “O” (*outside*, fuera). Podemos ver en (8) la anotación del ejemplo anterior con etiquetas BIO:

8. no conllevan ninguna molestia al llevarlo en el bolsillo.
['B', 'O', 'I', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O']

Este es el *output* interno del sistema de detección de marcadores de negación. El sistema de etiquetado BIO ayuda a predecir mejor los marcadores de negación y diferenciarlos entre sí para marcar qué elemento es el inicial y cuáles son la continuación. El corpus SFU Review_{SP}-NEG solo especifica qué marcadores aparecen en una oración, pero esta información se ha podido traducir directamente al formato de etiquetas BIO. Por eso, este *output* interno nos resulta muy útil para evaluar el sistema.

El *output* que genera el sistema y se envía al NEGES es una oración en formato CoNLL que presenta los mismos atributos que en el corpus de desarrollo. Esto es, texto en formato CoNLL en el que los marcadores de negación se identifican como *negative*.

3.2. Creación del sistema de detección de marcadores de negación CLiC-Neg

Decidimos aplicar técnicas de aprendizaje automático para elaborar nuestro propio sistema de detección de marcadores de negación. Asimismo, incorporamos algunos cambios para comprobar si combinar el aprendizaje automático con reglas y listas de marcadores de negación podría obtener mejores resultados.

A continuación, describimos todas las etapas en el desarrollo de nuestro sistema de detección de marcadores de negación. Explicamos cómo diseñamos una base de referencia con la que comparar resultados (apartado 3.2.1) y, posteriormente, elaboramos un modelo basado en CRF (apartado 3.2.2). A continuación, presentamos un análisis de los errores generados por este modelo (apartado 3.2.3) y cómo intentamos mejorarlo añadiendo reglas y listas de marcadores de negación (apartado 3.2.4). Estas últimas versiones han sido definidas como el sistema CLiC-Neg. Por último, mostramos los resultados obtenidos por las diferentes versiones del sistema (modelo CRF, sistema CLiC-Neg más reglas y sistema CLiC-Neg más listas) en el corpus de desarrollo.

3.2.1. Base de referencia

El primer paso para empezar a desarrollar el sistema CLiC-Neg consistió en definir una base de referencia muy sencilla con la que poder comparar los resultados de nuestro sistema. Diseñamos una base de referencia que simplemente utilizase un diccionario de marcadores de negación extraído del corpus de desarrollo.

Para obtener este diccionario, desarrollamos un clasificador binario para detectar si una oración contiene negación o no⁶. Para ello, aplicamos la técnica de la regresión logística, una extensión de la regresión lineal que ha obtenido buenos resultados en la resolución de tareas de clasificación. Este sistema obtuvo una precisión del 98% utilizando como rasgo solamente el lema de las palabras. Consideramos, por tanto, que era un sistema suficientemente exigente para utilizarlo como base de referencia y compararlo con el sistema CLiC-Neg.

Aislamos todas las oraciones negativas y extrajimos los doce marcadores de negación más frecuentes: ‘sin’, ‘no’, ‘nunca’, ‘ni’, ‘tampoco’, ‘nada’, ‘ni_siquiera’, ‘nadie’, ‘ninguno’, ‘no_sólo’, ‘jamás’, ‘falto’. Los incorporamos a un diccionario y diseñamos un modelo muy básico que utilizara esta lista como único rasgo. El modelo obtuvo un valor-f promedio del 73.08%. La tabla 6 muestra los resultados obtenidos por la base de referencia, en cada uno de los dominios del corpus, así como el promedio de la precisión, la cobertura y el valor-f:

Dominio	Precisión	Cobertura	valor-f
Coches	70,27%	55,32%	61,91%
Hoteles	85,71%	59,02%	69,9%
Lavadoras	90,62%	64,44%	75,32%
Libros	82,93%	70,83%	76,4%
Teléfono móviles	89,13%	75,23%	81,59%
Música	72%	69,23%	70,59%
Ordenadores	80,49%	63,46%	70,97%
Películas	89,41%	69,09%	77,95%
Promedio	82,57%	65,82%	73,07%

Tabla 6. Resultados de la base de referencia con el corpus de entrenamiento

Estos resultados confirman, por tanto, las conclusiones de Cruz et al. (2016:1404) en lo que se refiere al uso de listas de marcadores de negación. Esto es, que solo incluyendo algunos de los marcadores más frecuentes se puede cubrir un porcentaje muy elevado de ejemplos que contengan negación.

6 Para más información sobre este método, cómo se elaboró y cómo se evaluó este detector de oraciones que contienen negación, véase Bolea (2019).

Sin embargo, nuestra base de referencia solo es capaz de detectar los marcadores simples, esto es, solo puede utilizar o bien la etiqueta “B” cuando detecta un marcador o bien “O” para el resto de casos, pero no puede aplicar la etiqueta “I”. Los casos de marcadores discontinuos, por tanto, no pueden ser correctamente identificados. En consecuencia, este será uno de los objetivos del sistema CLiC-Neg.

3.2.2. El modelo CRF de detección de marcadores de negación

A continuación, escogimos un modelo de aprendizaje para mejorar los resultados obtenidos por la base de referencia. Primero probamos una técnica de aprendizaje automático no supervisado, siguiendo la propuesta de Fabregat et al. (2018) para el NEGES 2018. Concretamente, implementamos una red neuronal de tipo LSTM (*Long Short-Term Memory*), una clase de red neuronal recurrente (o RRN, del inglés *Recurrent Neural Network*), ya que es una de las técnicas más utilizadas en *Deep Learning*.

La principal ventaja de esta técnica es que, a diferencia de otras redes neuronales, esta puede almacenar la información que va procesando en una memoria para poder utilizarla posteriormente. Las redes neuronales recurrentes utilizan estructuras circulares que les permiten obtener nuevos datos, almacenarlos y volver a utilizarlos. Estas redes funcionan como si contuvieran varias redes neuronales que van procesando y almacenando información en una especie de reacción en cadena. Este tipo de redes plantean, por contra, el problema de que presentan dificultades para manejar dependencias a larga distancia.

Las redes de tipo LSTM solucionan esta desventaja, ya que, a diferencia del resto de RNN, pueden aprender esta clase de dependencias. Esto hace que puedan aprender y almacenar información que puede ser reutilizada en dependencias a larga distancia. Gracias a esta ventaja, son una técnica útil para una tarea de etiquetado como lo es la detección de los marcadores de negación. Sin embargo, este método no nos resultó útil porque a causa del número limitado de oraciones negativas que aparecen en el corpus facilitado para la tarea⁷ nuestra red neuronal LSTM no estaba aprendiendo ningún modelo relevante para detectar marcadores de negación.

Por este motivo, decidimos cambiar de enfoque y adaptar la propuesta de Loharja et al. (2018), que utiliza técnicas de aprendizaje automático supervisado basadas en CRF. Este modelo fue el primer clasificado en la Tarea 2 del NEGES 2018. También seguimos el trabajo de Agarwal & Yu (2010) que, como expusimos en el apartado 2.2.2.2, obtuvo resultados muy positivos. Estos trabajos demuestran que el CRF es especialmente útil para predecir secuencias de etiquetas, dado que es capaz de tener en cuenta las etiquetas no solo de la palabra que analiza, sino también de las anteriores y las posteriores. Es decir, que en lugar de manejar solamente probabilidades locales, puede estimar la probabilidad de una secuencia completa. Además, esto hace suponer que se necesitan menos datos para obtener buenos resultados (Lafferty et al., 2001).

⁷ Véase el apartado 3.1.1 para más información sobre el número total de oraciones y negaciones contenidas en el corpus SFU Review_{SP}-NEG.

La principal ventaja del CRF en comparación con modelos generativos, como por ejemplo los modelos ocultos de Markov (HMMs, del inglés *Hidden Markov Models*) y las gramáticas estocásticas, que han sido ampliamente aplicadas en ciertas tareas de PLN como el etiquetado de categorías morfosintácticas, es que estos últimos necesitan enumerar todas las probabilidades locales observadas para poder determinar una distribución conjunta de probabilidades. Puesto que esta es una tarea demasiado compleja, al final los modelos generativos se apoyan en lo que observan en un momento dado y no tienen en cuenta las observaciones pasadas o futuras. Con el CRF, es posible flexibilizar la hipótesis de que cada observación debe ser independiente y así se permite tener en cuenta el contexto mirando las etiquetas de las palabras vecinas (esto es, las inmediatamente anteriores y posteriores). Dado que se puede considerar identificar un marcador de negación como una tarea de etiquetado, donde la etiqueta de un determinado marcador dependerá de la de las palabras anteriores y posteriores, el modelo basado en CRF nos permitió predecir las etiquetas BIO con la ayuda del contexto.

Una vez elegido el modelo, desarrollamos el sistema CLiC-Neg usando Python y la librería especializada en tareas de aprendizaje automático Scikit-learn. Escogimos la librería CRF Suite para implementar nuestro modelo CRF. Usamos los *wrappers* para Python, puesto que CRF Suite está escrita en C y C++. En concreto, usamos los *wrappers* para la librería Scikit-learn, que funciona exclusivamente en Python. Elegimos Scikit-learn y no NLTK porque consideramos que era mejor opción, ya que no necesitábamos realizar ninguna tarea de preprocesado del corpus, como *tokenización* o etiquetado morfológico, tareas para las que resulta especialmente útil NLTK. En cambio, consideramos que Scikit-learn era una mejor opción para implementar modelos de aprendizaje automático.

Con todos estos elementos, procedimos con el entrenamiento y evaluación del modelo. Para el entrenamiento, nuestro modelo de CRF aplicó una serie de rasgos. El rasgo más sencillo que se tuvo en cuenta fue el mismo que se aplicó en la base de referencia, esto es, la lista con los doce marcadores negativos más frecuentes. Para mejorar resultados, se añadieron otros rasgos para poder recoger más información sobre el contexto. Así, añadimos, para cada frase, una representación de cada palabra que variaría según las palabras vecinas, contenidas en un intervalo de [-6, 1] palabras, es decir, las seis palabras que aparecen antes y la palabra posterior. Por cada una de estas palabras en este intervalo, el modelo almacena también, además de la palabra, su categoría morfosintáctica. Esta es la lista completa de todos los rasgos que usamos para diseñar nuestro modelo basado en CRF:

- PALABRA: la palabra en sí misma.
- POS: la categoría morfosintáctica (*Part of Speech*) de la palabra observada.
- BIGRAMANTERIOR: un bigrama de hasta 6 palabras anteriores a la palabra observada.
- BIGRAMAPOSTERIOR: un bigrama de hasta 1 palabra posterior a la palabra observada.
- PREVIOPOS: la información sobre la categoría morfosintáctica de hasta 6 palabras anteriores a la palabra observada.

- POSPOS: la información sobre la categoría morfosintáctica posterior a la palabra observada.

Además de estos rasgos, también aplicamos los parámetros por defecto del modelo CRF incluidos en la librería Scikit-learn. Así, diseñamos un modelo CRF que entrenamos con la parte de entrenamiento del corpus de desarrollo. A continuación, evaluamos el modelo CRF con la parte de evaluación del corpus de desarrollo. Los resultados se muestran en la tabla 7, donde se puede comprobar que el valor-f promedio alcanzado es del 84,18%:

Dominio	valor-f
Coches	75,86%
Hoteles	85,18%
Lavadoras	86,42%
Libros	85,2%
Teléfono móviles	90,29%
Música	81,48%
Ordenadores	81,25%
Películas	87,81%
Promedio	84,18%

Tabla 7. Resultados del modelo CRF con el corpus de desarrollo

El modelo basado en CRF mejora los resultados de la base de referencia en todos los ámbitos, sin excepción. Nótese que esta mejora es más significativa en algunas áreas, como por ejemplo «Coches», donde el valor-f es casi un 12% mejor que en la base de referencia y se sitúa en torno al 10% en casi todos los dominios. Los resultados son mejores en algunos temas, como por ejemplo en «Teléfonos Móviles» (donde el valor-f es del 90.29%) o «Libros» (con un valor-f del 86.42%) que en otras áreas como «Coches», donde el valor-f se sitúa en un 75%. Sería interesante, para futuras investigaciones, analizar si existen rasgos específicos de cada temática, relacionados con los distintos géneros textuales, que puedan ayudar a comprender esta disparidad en la precisión del modelo. Por ejemplo, una variable que podría tener impacto en estos resultados es la puntuación, es decir, si los comentarios presentan una puntuación más o menos normativa, algo que podría dificultar la detección de marcadores discontinuos.

3.2.3. Análisis de errores del modelo CRF

Una vez vistos los resultados de la evaluación de nuestro modelo basado en CRF, llevamos a cabo un exhaustivo análisis de los errores generados para definir mejor cómo reajustar el

modelo. Llegamos a la conclusión de que los errores podían clasificarse, fundamentalmente, en cuatro categorías:

1. Errores causados porque el sistema detecta como negativa cualquier oración que contenga al menos un marcador, aunque semánticamente no exprese una verdadera negación: como explican Jiménez-Zafra et al. (2018b:12-15) se deben marcar como no negativas las expresiones siguientes:

1.1. Estructuras con marcadores de negación retóricas: esto es, marcadores de negación que aparecen en interrogaciones retóricas, como en (9), o contenidos en expresiones multipalabra como ‘hasta_que_no’, como en (10). En estos casos, estas oraciones no expresan una verdadera negación:

- 9. El coche lo compré para viajar, **no**?⁸
- 10. No pienso irme hasta que **no** vengas.

1.2. Expresiones multipalabra que contienen marcadores de negación: algunas locuciones en español contienen marcadores de negación pero no expresan un verdadero valor negativo, como en (11) y (12):

- 11. Eso **no_hace_más_que** denotar su estado de pobreza absoluta.
- 12. **No_veas** los pifostios que se montaban.

1.3. Estructuras de contraste: en español se pueden utilizar estructuras que contraponen dos elementos y que no presentan un valor verdaderamente negativo, aunque contengan marcadores de negación. Estas estructuras se pueden usar para añadir una corrección, como en (13) o para añadir nueva información, como en (14):

- 13. **No** vinieron 2 soldados, **sino** 6.
- 14. Es un coche pensado **no_solo** para su uso, **sino** para su disfrute.

1.4. Estructuras comparativas: las estructuras comparativas pueden incluir marcadores de negación para expresar negación en relación con otro elemento. Pueden expresar comparación, como en (15), o grado superlativo, como en (16). Son estructuras discontinuas y se deben considerar comparativas y no como negación:

- 15. **No** me gusta **tanto como** el otro.
- 16. El motor **no** es **todo** lo potente que debería.

Nuestro modelo basado en CRF no detecta este tipo de casos y los clasifica como estructuras negativas. En consecuencia, etiqueta los marcadores con las etiquetas “B” e “I”,

8 Los ejemplos de (9) a (16) han sido extraídos de Jiménez-Zafra et al. (2018b:12-15).

mientras que el resultado debería haber sido etiquetar todas las palabras de estas oraciones como “O”, como puede verse en los ejemplos (17) a (20):

17. Ya estaba casi todo, **no** (B)?⁹
18. vamos por qué no hacemos documentales de el apareamiento humanos y nos dejamos ya de tonterías **no** (B)?
19. (...) pero **no** (B) es **más** (I) que el Ying y el YANG, unidos y dispares.
20. Una amiga mia está flipada con este grupo y hasta que **no** (B) me dejó el disco para que lo escuchase no paró.

El modelo tampoco etiqueta correctamente casos de oraciones comparativas. Las considera negativas y, por tanto, etiqueta los marcadores de negación contenidos en estas oraciones, como se puede ver en (21) a (23):

21. La comodidad de el vehículo para los ocupantes; la comodidad de la conducción, la cual **no** (B) es **tan** (I) diferente de la de un turismo (...)
22. Tokio_Hotel son los nuevos idolos de quinceañeras, con un estilo **no** (B) **tan** (I) pasteloso como los antiguos idolos de adolescentes de los años 80 y 90 Take_That , Backstreet boys, New_Kids_On_The_Block (...)
23. **No** (B) se profundiza **tanto** (O) en sentimientos como sí veremos en otras partes .

Una vez más, el etiquetado correcto en este caso hubiese sido clasificar todas las palabras de oraciones como (21) a (23) como “O”. Lo mismo sucede con oraciones como (24) a (26), que expresan contraste:

24. Recordaré **no_sólo** (B) cuando jugaba en ella , **sino** (I) cuando me iba haciendo mayor
25. que la corrupción **no** (B) está solo dentro_de mí , **sino** (I) en el desorden que se ha apoderado de todos nosotros
26. La mitad de el disco **no** (B) es propiamente de ellos , **sino_que** (O) son arreglos orquestales para la película , creados por su productor George_Martin.

Nótese que el modelo CRF no siempre es totalmente coherente. Así, en (26), etiqueta ‘no’ como inicio de marcador de negación pero no clasifica ‘sino_que’ como continuación, sino como “O”. De todas formas, puesto que se trata de una oración que expresa contraste (añade una corrección o un matiz a la información anterior), debería etiquetar todos los elementos como “O”.

2. Errores contenidos en expresiones multipalabra: identificamos algunas expresiones multipalabra que contienen marcadores de negación que no fueron correctamente detectados por el sistema como expresiones negativas. Esta es la lista completa de elementos no detectados:

9 En los ejemplos de (17) a (20) hemos transcrito el etiquetado, erróneo, generado por nuestro sistema y no el etiquetado que sería correcto que, en este caso, clasificaría los marcadores de negación como “O”.

3. Metodología para la detección automática de marcadores de negación

en_mi_vida	en_ningún_momento
a_no_ser	sin_necesidad_de
ni_tan_siquiera	ni_mucho_menos
nunca_más	ni_por_asomo
en_absoluto	mucho_menos

Lista de marcadores de negación multipalabra no detectados por el modelo CRF

La dificultad de identificar estos marcadores radicaba en que, si se añadían reglas específicas para detectarlos o una lista de todos ellos para que fuesen etiquetados se corría el riesgo de que el sistema detectase cada una de estas palabras como marcadores de negación en otros contextos en los que aparecían aisladas y no lo eran. En consecuencia, para no generar nuevos falsos positivos, hubo que aplicar un método de ensayo-error añadiendo los marcadores uno a uno y analizando si el cambio mejoraba o empeoraba la precisión del modelo.

3. Errores en el etiquetado de los elementos ‘tan’, ‘tanto’ y ‘muy’, ‘mucho’: el etiquetado de los elementos ‘tan’ y ‘tanto’ complicó su detección como marcador de negación. Como hemos visto en el punto 1, cuando este elemento aparece en oraciones comparativas, no debería aparecer etiquetado como marcador porque pertenece a una oración que realmente no expresa negación. Sin embargo, cuando aparece con un valor intensificador, formando parte del marcador discontinuo ‘no’... ‘tan’/ ‘tanto’, en algunos casos debería ser etiquetado como “I”, es decir, como un elemento dentro de un marcador de negación y no como otro marcador. En otros, no debería ser considerado marcador y por tanto la etiqueta BIO correcta es “O”. En el ejemplo (27) podemos ver que cuando el adjetivo indefinido ‘tanto’ debería ser etiquetado como “I” nuestro modelo CRF lo etiqueta como “O” :

27. etiquetado correcto → Aquí **ya** (B) **no** (I) hay **tanta** (I) acción y la narración se centra más en reflexiones, pensamientos y sentimientos de el protagonista .
etiquetado modelo CRF → Aquí **ya** (O) **no** (B) hay **tanta** (O) acción y la narración se centra más en reflexiones, pensamientos y sentimientos de el protagonista .

En cambio, cuando aparece el adverbio ‘tan’, el etiquetado correcto sería “O”, esto es, sería no considerarlo marcador de negación. Sin embargo, nuestro modelo lo detecta como marcador discontinuo y lo etiqueta como “I”:

28. etiquetado correcto → bueno **no** (B) lo dejaré **tan** (O) mal
etiquetado modelo CRF → bueno **no** (B) lo dejaré **tan** (I) mal

Probablemente a causa de esta diversidad de etiquetas, en función de si ‘tan’/ ‘tanto’ aparece o no en una oración comparativa o si es un adverbio o un adjetivo, nuestro modelo extrapola la conclusión de que ‘tan’/ ‘tanto’ es un marcador de negación y lo etiqueta como “I”

cuando aparece detrás de un ‘no’, incluso en contextos en los que aparece varias palabras después y ya no forma parte de ningún marcador de negación, como en (29) y (30):

29. En todo el libro apenas hay diálogo , pero **no** (B) tenéis que asustaros por sus párrafos eternos porque muy a menudo son prácticamente **tan** (I) ligeros como un best seller actual (aunque ya quisieran muchos autores dominar la escritura como lo hacía Terenci_Moix) .
30. Bueno espero **no** (B) haber os aburrido con el resumen de este libro, y que se os haya hecho esta opinión **tan** (I) amena como a mí el libro.

En el caso de ‘muy’/ ‘mucho’, nuestro sistema tiene dificultades para etiquetarlo como segundo elemento (“I”) de un marcador de negación. La causa podría ser el etiquetado diferente de estos términos según su categoría gramatical en la anotación del corpus. Así, en el caso de ‘mucho’, cuando este es un adverbio, se considera que no forma parte del marcador de negación:

31. Antes_de hacer la reserva de la habitación intenté buscar opiniones de gente que lo hubiera visitado y **no** (B) encontré **mucho** (O).

Sin embargo, nuestro modelo lo etiquetó como “I”. En cambio, si ‘mucho’ es un adjetivo, debe ser etiquetado como “I”, como se puede ver en (32):

32. sí, el aumento de rendimiento era gratis, pero para cuatro mini-tornillos creo que **no** (B) pasaremos **muchos** (I) problemas económicos

En cambio, nuestro modelo no detecta ‘mucho’ como marcador de negación y lo etiqueta como “O”. Suponemos que las dificultades del CRF para etiquetar correctamente este elemento vienen dadas por las diferentes funciones que puede presentar esta palabra.

En el caso de ‘muy’, puesto que solo puede tener una función gramatical, la de adverbio, aparece siempre etiquetado como parte de un marcador de negación (“I”), como se puede observar en los ejemplos (33) a (36):

33. **No** (B) es **muy** (I) normal que te cambien un cd 3 veces.
34. Pues la historia **no** (B) es que sea **muy** (I) original
35. hombre **no** (B) soy **muy** (I) alta (1.6) pero tampoco un pitufo (...)
36. De niña, ya todos notaban algo, sí, algo indefinible que hacía que la gente la siguiera, que quisiera tener la cerca, **sin** (B) saber **muy_bien** (I) por qué.

Como se puede ver en (36), la etiqueta “I” es la correcta también cuando ‘muy’ se coordina con la preposición ‘sin’. No obstante, nuestro modelo tiene dificultades en reconocer este elemento y lo etiqueta siempre como “O”.

4. Errores en la detección de marcadores discontinuos: en español, una oración negativa puede contener marcadores de negación complejos, a menudo discontinuos. Una combinación

3. Metodología para la detección automática de marcadores de negación

frecuente es la formada por ‘no’ y un segundo elemento que corresponde a uno de los pronombres o adjetivos ‘ninguno’/ ‘ninguna’ o de los pronombres ‘nada’ o ‘nadie’. El valor de estos marcadores es el de reforzar la negación:

37. Desde mi punto de vista el final **no** (B) aporta **nada** (I), es como si la trama no hubiera avanzado.

En este caso, nuestro modelo etiqueta ‘nada’ como “B”, es decir, como marcador de negación, pero interpreta que forma parte de una expresión negativa diferente. En (38), ‘ninguno’ forma parte de un marcador discontinuo (de ahí la etiqueta “I”), pero nuestro modelo no lo detecta como marcador de negación y lo etiqueta como “O”:

38. Si en algunos modelos de gama media Nokia **no** (B) pone esmero **ninguno** (I) en este aspecto.

El modelo basado en CRF también presenta dificultades para reconocer los marcadores discontinuos formados por ‘sin’ como elemento inicial (“B”) y los pronombres ‘ninguno’, ‘nada’ o ‘nadie’ como continuación (“I”). En el ejemplo (39), además, aparece más de un marcador de negación, lo que dificulta todavía más el etiquetado. Incluimos ambas versiones (la correcta y la que predice nuestro modelo) para facilitar la comprensión de este ejemplo:

39. etiquetado correcto → he de decir que funciona perfectamente permitiendo te grabar hasta un máximo de 48x y **sin** (B) tener **nunca** (I) **ningun** (B) fallo en la grabación.
etiquetado modelo CRF → he de decir que funciona perfectamente permitiendo te grabar hasta un máximo de 48x y **sin** (B) tener **nunca** (B) **ningun** (I) fallo en la grabación.

Nuestro modelo CRF interpreta que ‘ningún’ forma parte del marcador iniciado por ‘nunca’ y considera que ‘sin’ es independiente. Cabe mencionar que este es un ejemplo particularmente complejo.

La negación en estructuras coordinadas plantea también problemas de difícil solución. En los ejemplos que se muestran a continuación, los marcadores ‘no’ o ‘sin’ (en caso de que introduzca un sintagma preposicional) se coordinan con la conjunción ‘ni’:

40. Y es que así es, **no** (B) me lo he leído entero porque **no** (B) me apetecían más nauseas **ni** (I) obviedades absurdas.
41. En primer lugar llama la atención la sencillez de su diseño exterior, **sin** (B) grandes adornos **ni** (I) formas extrañas que requiera cualquier examen aerodinámico.

En (40) ‘ni’ se coordina con el segundo ‘no’, de ahí la etiqueta “I”, y en (41) ‘ni’ se coordina con ‘sin’. Sin embargo, el modelo CRF etiqueta ambos marcadores como “B”, esto es, los considera nuevos marcadores de negación.

5. Casos especiales: detectamos otros ejemplos mal etiquetados que eran menos frecuentes. Por ejemplo, el modelo CRF presenta dificultades para identificar correctamente el marcador de negación ‘ningún’/ ‘ninguno’. Por un lado, no lo etiqueta como elemento “I” en marcadores discontinuos, como ya hemos visto en ejemplos como (39). Así, en (42) nuestro sistema etiqueta ‘ningún’ como “B” cuando debería aplicar la etiqueta “I”:

42. Por_favor directores españoles hagan cine español y no intenten hacer réplicas a los americanos que por historias de fantasmas **no** nos darán **ningún** oscar.

Por otro lado, cuando aparece en posición inicial, como en (43), el sistema no lo detecta y lo etiqueta como “O”:

43. etiquetado correcto → **Ninguno** (“B”) tiene empatía a el prójimo, solo convicción de su poder.
etiquetado modelo CRF → **Ninguno** (“O”) tiene empatía a el prójimo, solo convicción de su poder.

Esto puede deberse a que el marcador ‘ningún’/ ‘ninguno’ suele aparecer en segunda posición, como marcador discontinuo que aparece tras ‘no’ o ‘sin’. No es tan frecuente que aparezca en posición inicial y, por tanto, el modelo CRF no ha podido ver muchos ejemplos de este tipo en la fase de entrenamiento y de ahí que no lo detecte correctamente.

El marcador de negación ‘aún no’ y ‘todavía no’ tampoco aparece correctamente etiquetado por el modelo CRF. La predicción correcta sería etiquetar ‘aún’ y ‘todavía’ como inicio del marcador, esto es, “B”, y ‘no’ como segundo elemento con la etiqueta “I”. Sin embargo, nuestro modelo etiqueta ambos marcadores como “B”:

44. Por problemas de visión he tenido que estar un tiempo alejada de el ordenador , y **aun** (B) **no** (B) es que bien al_100_% pero si que estoy muchísimo mejor.
45. Para mi este ordenador para tener casi tres añitos esta bastante bien y **aun** (B) **no** (B) se ha quedado anticuado.
46. A continuación os voy a contar un truco para mejorar el hardware de el ordenador que me mandó un amigo mío que es informático por internet, yo **todavía** (B) **no** (B) lo he hecho asi que allá con las consecuencias yo no se como puede salir, mi amigo me ha prometido que va bien....

Finalmente, el modelo CRF no etiqueta correctamente casos de negación léxica, como es el caso de ‘apenas’. Este adverbio puede tener un valor de negación o no, según el contexto. Así, en el enunciado (47) ‘apenas’ expresa negación pero nuestro modelo lo etiqueta como “O”, mientras que en (48) no expresa negación pero nuestro modelo lo etiqueta como “B”:

47. etiquetado correcto → que risa M! Luisa, si ese kilometraje esta hecho en un par de años para conductores que **apenas** (B) tocan el coche.

3. Metodología para la detección automática de marcadores de negación

etiquetado modelo CRF → que risa M! Luisa, si ese kilometraje esta hecho en un par de años para conductores que **apenas** (B) tocan el coche.

48. etiquetado correcto → Lo he ido retomando a ratos y en distintos momentos, pero **apenas** (O) leo unas páginas y me aburro.

etiquetado modelo CRF → Lo he ido retomando a ratos y en distintos momentos, pero **apenas** (B) leo unas páginas y me aburro.

Se trata de otro caso motivado por la escasez de ejemplos, ya que ‘apenas’ aparece tan solo 3 veces en todo el corpus de entrenamiento.

3.2.4. El sistema CLiC-Neg

Una vez analizados los errores generados por nuestro modelo basado en CRF, en nuestra primera versión del sistema CLiC-Neg (reglas) introducimos algunas reglas que afectaban la interpretación de los ítems léxicos como marcadores de negación. En una segunda versión del sistema CLiC-Neg (lista), introducimos una lista de marcadores de negación extraída de otro corpus.

Las reglas que introducimos en el modelo CRF para mejorar la detección de los marcadores de negación son las siguientes:

REGLA 1: Se introdujo esta regla para que el sistema no interprete como una estructura de negación las expresiones retóricas del tipo ‘, no?’. Se trata de una regla que detecta este tipo de negaciones retóricas y que tiene como objetivo evitar falsos positivos.

REGLA 2: Si aparece una secuencia que incluya ‘no’ y ‘nada más’ en un intervalo de [0,5] palabras, estos elementos no deben ser etiquetados como marcadores de negación, ya que este tipo de estructuras no expresan una negación.

REGLA 3: Si ‘ningún’ aparece en posición inicial, se trata de un marcador de negación simple y debería ser etiquetado como elemento inicial (“B”). De esta manera, pretendemos facilitar el etiquetado de un elemento que puede aparecer o bien solo en inicio de oración, y entonces funcionar como marcador simple, o bien detrás de ‘no’ y formar parte de un marcador complejo. Además, esta regla ayudaría a detectar los contextos en los que ‘ningún’ aparece en inicio de frase que, como vimos en el ejemplo (43), no son detectados por el modelo CRF.

REGLA 4: Si ‘no’/ ‘tampoco’/ ‘sin’ (palabra 1) aparece seguido, en un intervalo de [0,10] palabras, por ‘nada’/ ‘ningún’/ ‘nadie’ (palabra 2), estos dos elementos forman un marcador complejo donde la palabra 1 es el inicio (“B”) y la palabra 2 es la continuación (“I”).

REGLA 5. Si ‘aún’ o ‘todavía’ aparecen inmediatamente seguidos de ‘no’, estas dos palabras forman parte de un marcador de negación complejo donde la palabra 1 (‘aún’ o ‘todavía’) es el inicio (“B”) y ‘no’ es la continuación (“I”). Con esta regla pretendemos evitar errores en el

etiquetado de estos términos, que a veces aparecen como inicio de un nuevo marcador, como en los ejemplos (44) a (46), y en otras ocasiones no es detectado como marcador, como en el siguiente ejemplo, donde el modelo CRF etiqueta ambos marcadores como “O”:

49. que aunque la gente tiene dinero para comprar se coches hoy más que nunca, **todavía** (O) **no** (O) tienen el suficiente para comprar se un buen coche (...)

REGLA 6. ‘Tan’ no debe ser nunca etiquetado como marcador de negación. Esta regla pretende eliminar un falso positivo frecuente.

Previa a la aplicación del modelo CRF, aplicamos estas seis reglas con la intención de mejorar no solo el etiquetado de los marcadores de negación, sino también el de los elementos que deberían ser considerados como externos a la negación. A continuación, todos los elementos que no fueron predichos por las reglas fueron procesados por el modelo CRF. Esto es, las decisiones de las reglas prevalecen sobre el modelo CRF si estas son aplicables.

Incluir estas reglas no solo no mejoró los resultados obtenidos con el modelo CRF sino que incluso disminuyó el valor-f promedio (de un 84.18% a un 81,89%). No se trata de un descenso notable pero, en cualquier caso, es suficiente para contradecir nuestra intuición de que estas reglas ayudarían a identificar mejor los marcadores de negación que el modelo CRF no ha podido detectar correctamente.

Para nuestra segunda versión del sistema CLiC-Neg, decidimos incluir solamente una lista de marcadores de negación multipalabra, extraídos del corpus NewsCom (Taulé et al., en prensa). La lista contenía marcadores de negación complejos tales como ‘aún_ni’, ‘cuanto_menos’, ‘ni_tan_siquiera’, ‘ni_de_coña’, ‘ni_de_lejos’, entre otros¹⁰. Consideramos que de esta manera evitaríamos falsos negativos ocasionados por la dificultad del modelo CRF para detectar negación en expresiones multipalabra.

Los resultados que obtuvimos fueron inferiores a los obtenidos con el modelo CRF. En consecuencia, procedimos a probar diferentes combinaciones con distintas listas que incluyeran diferentes expresiones multipalabra. Tras varias pruebas, comprobando con cada modificación los resultados de precisión, cobertura y valor-f del sistema, llegamos a la conclusión de que la mejor combinación posible consistía en eliminar de nuestra lista los términos ‘tan’ y ‘tanto’, que habíamos eliminado con la regla 6, y añadir ‘a_no_ser’, ‘en_absoluto’, ‘en_ningún_momento’ y ‘sin_necesidad_de’.

En la tabla 8 se muestra una comparación entre el valor-f obtenido por el modelo CRF, el de la primera versión del sistema CLiC-Neg, que combina un modelo CRF con las reglas que hemos

¹⁰ Véase el apéndice para comprobar la lista completa de marcadores de negación que se incluyeron en esta versión del sistema CLiC-Neg.

3. Metodología para la detección automática de marcadores de negación

descrito, y la segunda versión del sistema CLiC-Neg, que combina el modelo CRF con la lista final de expresiones multipalabra:

Dominio	CRF	CLiC-Neg (reglas)	CLiC-Neg (lista)
Coches	75,86%	76%	75%
Hoteles	85,18%	81%	89,29%
Lavadoras	86,42%	86,42%	86,36%
Libros	85,2%	84,13%	83,57%
Teléfono móviles	90,29%	84,69%	89,42%
Música	81,48%	77%	80%
Ordenadores	81,25%	80,85%	81,25%
Películas	87,81%	85%	88,46%
Promedio	84,18%	81,89%	84,17%

Tabla 8. Valores-f de los diferentes sistemas evaluados con el corpus de desarrollo

4. Resultados del NEGES 2019

A la espera de participar en el taller NEGES 2019 (septiembre del 2019), en el que se podrán analizar los resultados de los sistemas participantes y se podrán comentar los algoritmos elaborados por los distintos equipos, así como los errores de cada sistema, de momento podemos mostrar los resultados obtenidos por los participantes. Estos resultados se han calculado aplicando los *scripts* de evaluación del *SEM2012 Shared Task -Resolving the Scope and Focus of Negation (Morante & Blanco, 2012), como vimos en el apartado 3. Los sistemas han sido evaluados, por tanto, con las medidas de precisión, cobertura y valor-f.

La última versión del sistema CLiC-Neg, que combina una lista de marcadores de negación con el CRF, fue enviado para participar en la subtarea A del NEGES 2019 y quedó clasificado en primer lugar. La tabla 9 muestra los resultados obtenidos por el sistema CLiC-Neg en el corpus de evaluación utilizado por los organizadores del NEGES 2019:

Dominio	Precisión	Cobertura	valor-f
Coches	94,92%	82,35%	88,19%
Hoteles	87,5%	71,19%	78,51%
Lavadoras	92,98%	76,81%	84,13%
Libros	80,59%	75,79%	78,12%
Teléfono móviles	87,76%	75,44%	81,13%
Música	94,44%	78,16%	85,53%
Ordenadores	90,48%	93,83%	92,12%
Películas	88,67%	81,60%	84,99%
Promedio	89,67%	79,40%	84,09%

Tabla 9. Precisión, cobertura y valor-f del sistema CLiC-Neg en el corpus de evaluación

El sistema CLiC-Neg obtiene una precisión del 89,67%, una cobertura del 79,40% y, finalmente, un valor-f del 84,09%, el más elevado entre los participantes del NEGES 2019. No obstante, no mejora los resultados obtenidos por Loharja et al. (2018), quienes obtuvieron un valor-f un 2% superior al nuestro en 2018. La principal diferencia entre su enfoque y el nuestro es el conjunto de rasgos aplicado al modelo CRF. Por ejemplo, ellos aplican algunos rasgos en un nivel inferior que el de la palabra, ya que tienen en cuenta elementos como prefijos y sufijos, y el sistema CLiC-Neg, no.

La tabla 10 muestra la comparación entre los resultados del sistema CLiC-Neg y los del resto de participantes en el NEGES 2019, en todos los dominios:

Dominio	Aspie96	CLiC	IBI	NLP_UNED
Coches	23,39%	88,19%	82,93%	87,3%
Hoteles	12,5%	78,51%	82,35%	83,02%
Lavadoras	28,57%	84,13%	81,96%	81,96%
Libros	20,51%	78,12%	76,57%	82,66%
Móviles	21,53%	81,13%	76,77%	82,66%
Música	28,2%	85,53%	83,33%	81,57%
Ordenadores	20,8%	92,12%	78,62%	85,91%
Películas	25,35%	84,99%	81,48%	85,53%
Promedio	22,58%	84,09%	80,50%	82,99%

Tabla 10. Valor-f obtenido por cada uno de los equipos participantes en el NEGES 2019 en el corpus de evaluación

Una vez analizados los resultados obtenidos por nuestro modelo y vistos los resultados del resto de modelos participantes, se aprecia lo siguiente:

- La mayor parte de los sistemas obtienen resultados parecidos. Esto es, la diferencia entre el sistema propuesto por la UNED y el nuestro es mínima y la diferencia entre el sistema CLiC-Neg y el propuesto por IBI no alcanza los 4 puntos.
- Para poder aprovechar al máximo los modelos de aprendizaje automático se requieren corpus más grandes. Dado el escaso número de oraciones y, en concreto, de estructuras negativas, contenidas en el corpus de desarrollo hay casos de marcadores de negación que se encuentran infrarrepresentados.
- En contra de nuestras expectativas, enriquecer el modelo CRF con reglas y listas de marcadores de negación no mejora el sistema. Consideramos que sería interesante profundizar en esta área y ver qué tipo de conocimiento lingüístico se tendría que incorporar en un modelo de aprendizaje automático. Por ejemplo, quizás una manera de mejorar resultados sería tratar el contexto de una manera diferente, o bien cambiando los bigramas que hemos usado o bien incorporar conocimiento sintáctico.

5. Conclusiones

En esta tesis de máster hemos desarrollado un sistema automático de detección de marcadores de negación en español basado en aprendizaje automático supervisado. Los resultados no han sido plenamente satisfactorios, ya que al intentar mejorar un modelo basado en CRF mediante conocimiento lingüístico (reglas y listas de marcadores de negación multipalabra) no hemos conseguido mejorar los resultados obtenidos por el sistema CRF básico. Como se trata de un trabajo preliminar, consideramos que hay que seguir trabajando en esta línea para determinar si realmente la incorporación de rasgos lingüísticos a modelos estadísticos puede mejorar el resultado.

Los sistemas que aplican técnicas de aprendizaje automático han obtenido resultados prometedores (Morante et al., 2008; Morante & Daelemans, 2009; Agarwal & Yu, 2010), tanto en la detección de marcadores de negación como del alcance de la negación. Todos superan los resultados obtenidos por los algoritmos basados en reglas, como NegEx. Por tanto, esto nos ha hecho decidimos a optar por esta técnica para elaborar nuestro sistema CLiC-Neg. Sin embargo, los resultados que hemos obtenido plantean el reto de superar el valor-f obtenido con un modelo puramente estadístico y determinar qué conocimiento lingüístico deberíamos añadir para alcanzar resultados que se aproximen al 95%-100%.

Para mejorar la detección de la negación sin duda se requiere mejorar los métodos aplicados y desarrollar nuevos corpus anotados con el fin de garantizar la representación de la compleja casuística que plantea la negación. Es decir, que la comunidad científica necesita más corpus, que dispongan de más ejemplos, anotados con negación y otros aspectos relacionados, como puede ser el alcance de la negación, el evento, la polaridad o el foco, especialmente para el desarrollo de sistemas orientados al análisis de sentimientos. Además, es imprescindible que estos corpus sean de libre acceso y que sean anotados con rigor, llevando a cabo pruebas de acuerdo entre anotadores.

En nuestra experimentación nos ha sorprendido el hecho de que las reglas tuvieran un efecto negativo en los resultados del modelo CLiC-Neg. Nuestra primera hipótesis consistía en que añadir reglas de tipo lingüístico resultaría útil para detectar ejemplos que no eran correctamente etiquetados usando solo el CRF, especialmente para detectar los marcadores de negación discontinuos, difíciles de etiquetar correctamente ya que lo más habitual es que el sistema detecte cada marcador como inicio de una nueva estructura negativa. Sin embargo, teniendo en cuenta los resultados observados, es evidente que la incorporación de reglas no ha mejorado el rendimiento del modelo, puesto que empeoraban más aspectos de los que mejoraban. Por tanto, podemos concluir que la mayor parte del trabajo de detección automática de la negación la llevan a cabo los modelos de aprendizaje automático.

A nivel personal, el ejercicio que he llevado a cabo me ha permitido experimentar con la implementación de un modelo estadístico, familiarizarme con esta tecnología y, si bien el resultado no ha sido todo lo satisfactorio que se deseaba, para mí ha significado un paso importante en el aprendizaje de las tecnologías del lenguaje.

Apéndice. Lista completa de los marcadores de negación multipalabra incorporados al modelo CRF

- 0
- a_excepción_de
- a_falta_de
- a_no_ser_que
- aún_ni
- aún_no
- bajo_ningún_concepto
- brillan_por_su_ausencia
- brilló_por_su_ausencia
- casi_nada
- casi_nunca
- cero
- cualquier_cosa_menos
- cuanto_menos
- en_mi_vida
- ni_idea
- ni_por_asomo
- ni_tampoco
- no_precisamente
- nula
- nunca_mas
- sin_apenas
- sin_siquiera
- tanto
- ya_no
- en_ningún_sentido
- en_toda_su_vida
- excepto
- falta_de
- falto_de
- faltos_de
- hasta_que_no
- jamas
- mucho_menos
- nada_de_nada
- nada_más
- nada_mas
- ni_a_la_de_tres
- ni_de_coña
- ni_de_lejos
- ningún_tipo_de
- ni_siquiera
- ni_tan_siquiera
- no_siempre
- nunca_más
- para_absolutamente_nada
- si_no
- tan
- todavía_no

Bibliografía

- Agarwal, S. & H. Yu (2010): «Biomedical negation scope detection with Conditional Random fields». *Journal of the American Medical Informatics Association*, 17(6): 696-701.
- Alkorta, J., K. Gojenola & M. Iruskietia (2018): «Saying no but meaning yes: negation and sentiment analysis in Basque». *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. Brussels, Belgium, October 31. Association for Computational Linguistics: 85-90.
- Altuna, B., A. Minard & M. Speranza (2017): «The Scope and Focus of Negation: A Complete Annotation Framework for Italian». *Proceedings of the Workshop Computational Semantics Beyond Events and Roles (SemBEaR)*. Valencia, Spain, April 4. Association for Computational Linguistics: 34-42.
- Aronow, D. B., F. Fangfang, & W.B. Croft (1999): «Ad hoc classification of radiology reports». *Journal of the American Medical Informatics Association*, 6(5): 393-411.
- Banjade, R. & V. Rus (2016a): «DT-Neg: Tutorial dialogues annotated for negation scope and focus in context». *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*: 3768-3771.
- Banjade, R., N. Niraula & V. Rus (2016b): «Towards Detecting Intra-and Inter-Sentential Negation Scope and Focus in Dialogue». *The Twenty-Ninth International Flairs Conference*. Association for the Advancement of Artificial Intelligence: 198-203.
- Beltrán, J. & M. González (en prensa): «Detection of Negation Cues: The CLiC-Neg System».
- Blanco, E. & D. Moldovan (2011): «Semantic representation of negation using focus detection». *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. Association for Computational Linguistics: 581-589.
- Blanco, E. & D. Moldovan (2014): «Retrieving implicit positive meaning from negated statements». *Natural Language Engineering* 20(4): 501-535.
- Bolea, M. (2019): *Detecció automàtica d'oracions amb negació* (TFG). Universitat de Barcelona, Barcelona.
- Costumero, R., F. López, C. Gonzalo-Martín, M. Millán & E. Menasalva (2014): «An approach to detect negation on medical documents in Spanish». *International Conference on Brain Informatics and Health*. Springer, Cham: 366-375.
- Cotik, V., R. Roller, F. Xu, H. Uszkoreit, K. Budde, & D. Schmidt (2016): «Negation detection in clinical reports written in German». *Proceedings of the Fifth Workshop on Building and Evaluating Resources for Biomedical Text Mining (BioTxtM2016)*: 115-124.
- Cruz Díaz, N.P. (2014): *Negation and speculation detection in medical and review texts*. Tesis Doctoral, Universidad de Huelva.
- Cruz Díaz, N.P., M. Taboada, & R. Mitkov (2016): «A machine-learning approach to negation and speculation detection for sentiment analysis». *Journal of the Association for Information Science and Technology* 67(9): 2118-2136.

- Chapman, W. W., W. Bridwell, P. Hanbury, G.F. Cooper, & B.G. Buchanan (2001): «A simple algorithm for identifying negated findings and diseases in discharge summaries». *Journal of biomedical informatics*, 34(5), 301-310.
- Chapman, W. W., D. Hilert, S. Velupillai, M. Kvist, M. Skeppstedt, B.E. Chapman, & L. Deleger (2013): «Extending the NegEx lexicon for multiple languages». *Studies in Health Technology and Informatics*, 192: 677.
- Dadvar, M., C. Hauff, & F. de Jong (2011): «Scope of negation detection in sentiment analysis». *Proceedings of the Dutch-Belgian Information Retrieval Workshop (DIR 2011)*. University of Amsterdam: 16-20.
- Fabregat, H., L. Araujo Serna, & J. Martínez Romo (2018): «Deep learning approach for negation trigger and scope recognition». *NEGES 2018: Workshop on Negation in Spanish: Seville, Spain: September 19-21, 2018: proceedings book*: 43-48.
- Fancellu, F., A. López, & B. Webber (2016): «Neural networks for negation scope detection». *Proceedings of the 54th annual meeting of the Association for Computational Linguistics (volume 1: long papers)*: 494-504.
- Farkas, R., V. Vincze, G. Móra, J. Csirik & G. Szarvas (2010): «The CoNLL-2010 shared task: learning to detect hedges and their scope in natural language text». *Proceedings of the Fourteenth Conference on Computational Natural Language Learning---Shared Task*. Association for Computational Linguistics: 1-12.
- Garcelon, N., A. Neuraz, V. Benoit, R. Salomon & A. Burgun (2016): «Improving a full-text search engine: the importance of negation detection and family history context to identify cases in a biomedical data warehouse». *Journal of the American Medical Informatics Association*, 24(3): 607-613.
- Gindl, S. (2006): *Negation Detection in Automated Medical Applications. A Survey*.
- Gindl, S. (2008): *Negation detection in medical documents using syntactical methods*.
- Gkotsis, G., S. Velupillai, A. Oellrich, H. Dean, M. Liakata & R. Dutta (2016): «Don't let notes be misunderstood: a negation detection method for assessing risk of suicide in mental health records». *Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology*: 95-105.
- Goldin, I. & W. W. Chapman (2003): «Learning to detect negation with 'not' in medical texts». *Proc Workshop on Text Analysis and Search for Bioinformatics, ACM SIGIR*.
- Harkema, H., J.N. Dowling, T. Thornblade & W.W. Chapman (2009): «ConText: an algorithm for determining negation, experiencer, and temporal status from clinical reports». *Journal of biomedical informatics* 42(5): 839-851.
- Hajič, J., M. Ciaramita, R. Johansson, D. Kawahara, M.A. Martí, L. Màrquez, A. Meyers, J. Nivre, S. Padó, J. Stépánek, P. Straňák, M. Surdeanu, N. Xue, & Y. Zhang (2009): «The CoNLL-2009 shared task: Syntactic and semantic dependencies in multiple languages». *Proceedings of the Thirteenth Conference on Computational Natural Language Learning: Shared Task*. Association for Computational Linguistics:1-18.
- Jia, L., C. Yu & W. Meng (2009): «The effect of negation on sentiment analysis and retrieval effectiveness». *Proceedings of the 18th ACM conference on Information and knowledge management*, Association for Computational Linguistics:1827-1830.

- Jiménez-Zafra, S. M., N.P. Cruz Díaz, R. Morante & M.T. Martín-Valdivia (2019): «NEGES 2019 Task: Negation in Spanish». *Proceedings of the Iberian Evaluation Forum (IberLEF 2019), CEUR Workshop Proceedings, Bilbao, Spain, CEUR-WS (2019)*.
- Jiménez-Zafra, S. M., M.T. Martín-Valdivia, M.D. Molina-González & L.A. Ureña-López (2018a): «Relevance of the SFU Review_{SP}-NEG corpus annotated with the scope of negation for supervised polarity classification in Spanish». *Information Processing & Management*, 54(2): 240-251.
- Jiménez-Zafra, S. M., M. Taulé, M.T. Martín-Valdivia, L.A. Ureña-López & M.A. Martí (2018b): «SFU Review_{SP}-NEG: a Spanish corpus annotated with negation for sentiment analysis. a typology of negation patterns». *Language Resources and Evaluation* 52(2): 533-569.
- Jiménez-Zafra, S. M., N.P. Cruz Díaz, R. Morante, M.T. Martín-Valdivia (2018c): «Tarea 2 del Taller NEGES 2018: Detección de Claves de Negación». *Proceedings of NEGES 2018: Workshop on Negation in Spanish*: 35-41.
- Jiménez-Zafra, S. M., M. Martín, L.A. López, M.A. Martí & M. Taulé (2016): «Problematic cases in the annotation of negation in Spanish». *Proceedings of the Workshop on Extra-Propositional Aspects of Meaning in Computational Linguistics (ExProM)*:42-48.
- Jiménez-Zafra, S. M., E. Martínez Cámara, M.T. Martín-Valdivia & M.D. Molina González (2015): «Tratamiento de la Negación en el Análisis de Opiniones en Español». *Procesamiento del Lenguaje Natural* 54:37-44.
- Kang, T., S. Zhang, N. Xu, D. Wen, X. Zhang & J. Lei (2017): «Detecting negation and scope in Chinese clinical notes using character and word embedding». *Computer Methods and Programs in Biomedicine* 140: 53-59.
- Konstantinova, N., S.C.M. De Sousa, N.P. Cruz Díaz, M.J. Maña, M. Taboada & R. Mitkov (2012): «A review corpus annotated for negation, speculation and their scope». *Lrec*:3190-3195.
- Lafferty, J., A. McCallum & F.C. Pereira (2001): «Conditional random fields: Probabilistic models for segmenting and labeling sequence data».
- Loharja, H., L. Padró & J. Turmo Borrás (2018): «Negation cues detection using crf on spanish product review texts». *NEGES 2018: Workshop on Negation in Spanish: Seville, Spain: September 19-21, 2018: proceedings book*: 49-54.
- Manimaran, J. & T. Velmurugan (2017): «Evaluation of lexicon -and syntax-based negation detection algorithms using clinical text data». *Bio-Algorithms and Med- Systems*, 13(4): 201-213.
- Martí, M. A., M. Taulé, M. Nofre, L. Marsó, M.T. Martín-Valdivia & S.M. Jiménez-Zafra (2007a): «La negación en español: análisis y tipología de patrones de negación». *Procesamiento del Lenguaje Natural* 57: 41-48.
- Taulé, M., M.A. Martí & M. Recasens (2007b): «Ancora: Multilingual and multilevel annotated corpora» (disponible en http://clic.ub.edu/corpus/webfm_send/13).
- Mehrabi, S., A. Krishnan, S. Sohn, A.M. Roch, H. Schmidt, J. Kesterson, C. Beesley, P. Dexter, C. Max Schmidt, L. Hongfang & M. Palakal (2015): «DEEPEN: A negation detection

- system for clinical text incorporating dependency relation into NegEx». *Journal of Biomedical Informatics* 54:213-219.
- Mitchell, K.J., M.J. Becich, J.J. Berman, W.W. Chapman, J. Gilbertson, D. Gupta, J. Harrison, E. Legowski & R.S. Crowley (2004): «Implementation and Evaluation of a Negation Tagger in a Pipeline-based System for Information Extraction from Pathology Reports». *Medinfo*:663-667.
- Morante, R., A. Liekens & W. Daelemans (2008): «Learning the scope of negation in biomedical texts». *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics: 715-724.
- Morante, R. & W. Daelemans (2009): «A metalearning approach to processing the scope of negation». *Proceedings of the Thirteenth Conference on Computational Natural Language Learning*, Association for Computational Linguistics: 21-29.
- Morante, R., S. Schrauwen, & W. Daelemans (2011): «Corpus-based approaches to processing the scope of negation cues: an evaluation of the state of the art». *Proceedings of the Ninth International Conference on Computational Semantics*. Association for Computational Linguistics: 350-354.
- Morante, R. & E. Blanco (2012): «*SEM 2012 shared task: Resolving the scope and focus of negation». **SEM 2012: The First Joint Conference on Lexical and Computational Semantics–Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*: 265-274.
- Morante, R. & W. Daelemans (2012): «ConanDoyle-neg: Annotation of negation in Conan Doyle stories». *Proceedings of the Eighth International Conference on Language Resources and Evaluation, Istanbul*.
- Morante, R. & C. Sporleder (2012): «Modality and negation: An introduction to the special issue». *Computational Linguistics* 38(2): 223-260.
- Mutalik, P.G., A. Deshpande & P.M. Nadkarni (2001): «Use of general-purpose negation detection to augment concept indexing of medical documents: a quantitative study using the UMLS». *Journal of the American Medical Informatics Association* 8(6): 598-609.
- Packard, W., E.M. Bender, J. Read, S. Oepen & R. Drìdan (2014): «Simple negation scope resolution through deep parsing: A semantic solution to a semantic problem». *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*: 69-78.
- Packer, Abel Laerte (1999): «SciELO: An electronic publishing model for developing countries». *Electronic publishing'99. Conference*.
- Peng, Y., X. Wang, L. Lu, M. Bagheri, R. Summers & Z. Lu (2018): «NegBio: a high-performance tool for negation and uncertainty detection in radiology reports». *AMIA Summits on Translational Science Proceedings 2018*: 188.
- Pyysalo, S., F. Ginter, J. Heimonen, J. Björne, J. Boberg, J. Järvinen & T. Salakoski (2007): «BioInfer: a corpus for information extraction in the biomedical domain». *BMC Bioinformatics* 8(1): 50.

- Ramshaw, L. A. (1999): «Text chunking using transformation-based learning». *Natural language processing using very large corpora*. Springer, Dordrecht, 157-176.
- Shivade, C., M.C. de Marneffe, E. Fosler-Lussier & A.M. Lai (2015): «Extending NegEx with kernel methods for negation detection in clinical text». *Proceedings of the Second Workshop on Extra-Propositional Aspects of Meaning in Computational Semantics (ExProM 2015)*: 41-46.
- Sohn, S., S. Wu, & C.G. Chute (2012): «Dependency parser-based negation detection in clinical narratives». *AMIA Summits on Translational Science Proceedings 2012*: 1.
- South, B. R., S. Phansalkar, A.D. Swaminathan, S. Delisle, T. Perl, & M.H. Samore (2007): «Adaptation of the NegEx algorithm to Veterans Affairs electronic text notes for detection of influenza-like illness (ILI)». *AMIA Annual Symposium Proceedings*: 1118-1118.
- Taboada, M., J. Brooke, M. Tofiloski, K. Voll & M. Stede (2011): «Lexicon-based methods for sentiment analysis». *Computational linguistics* 37(2): 267-307.
- Tanushi, H., H. Dalianis, M. Duneld, M. Kvist, M. Skeppstedt & S. Velupillai (2013): «Negation scope delimitation in clinical text using three approaches: NegEx, PyConTextNLP and SynNeg». *19th Nordic Conference of Computational Linguistics (NODALIDA 2013), May 22-24, 2013, Oslo, Norway*. Linköping University Electronic Press: 387-474.
- Taulé, M., M. Nofre, M. González & M.A. Martí (en prensa): «Focus of negation: its identification in Spanish». *Natural Language Engineering* 1(1): 1-18.
- Vincze, V., G. Szarvas, R. Farkas, G. Móra & J. Csirik (2008): «The BioScope corpus: biomedical texts annotated for uncertainty, negation and their scopes». *BMC Bioinformatics* 9(11): S9.
- Vlachidis, A., & D. Tudhope (2015): «Negation detection and word sense disambiguation in digital archaeology reports for the purposes of semantic annotation». *Program* 49(2): 118-134.
- Wiegand, M., A. Balahur, B. Roth, D. Klakow & A. Montoyo (2010): «A survey on the role of negation in sentiment analysis». *Proceedings of the workshop on negation and speculation in natural language processing*: 60-68.
- Wu, S., T. Miller, J. Masanz, M. Coarr, S. Halgrim, D. Carrell & C. Clark (2014): «Negation's not solved: generalizability versus optimizability in clinical natural language processing». *PloS one* 9(11): e112774.
- Zhang, J. Ch. (2014): *Diseño e implementación de un algoritmo para la detección de la negación de textos clínicos en español*. Trabajo Fin de Grado. Universidad Politécnica de Madrid. Facultad de Informática.
- Zhu, Q., J. Li, H. Wang & G. Zhou (2010): «A unified framework for scope learning via simplified shallow semantic parsing». *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics: 714-724.