Resumen de tesis **Alessandra Teresa Cignarella**

Dependency Syntax in the Automatic Detection of Irony and Stance

La presente tesis se enmarca dentro del amplio panorama de estudios relacionados con el Procesamiento del Lenguaje Natural (NLP). En concreto, se trata de un trabajo de Lingüística Computacional (CL) cuyo objetivo principal es estudiar en profundidad la contribución de la sintaxis en el campo del análisis de sentimientos y, en concreto, aplicado a estudiar textos extraídos de las redes sociales o, más en general, de contenidos online.

Además, dado el reciente interés de la comunidad científica por el proyecto Universal Dependencies (UD), en el que se propone un formato de anotación morfosintáctica destinado a crear una representación ``universal'' de la morfología y sintaxis aplicable a diferentes idiomas, en este trabajo se utiliza este formato con el propósito de realizar un estudio desde una perspectiva multilingüe (italiano, inglés, francés y español). Aunque el formato UD se concibió originalmente para ser aplicado a textos escritos en la lengua ``estándar'' desde el punto de vista de las normas morfosintácticas y la puntuación, recientemente se ha comenzado a aplicar el mismo esquema también a contenidos generados por usuarios en línea (User-Generated Content, UGC), es decir, a textos extraídos de redes sociales, blogs, foros y plataformas de microblogging, como las páginas de Reddit, Twitter o Wikipedia, en el que se utiliza un registro más informal. Inevitablemente, la aplicación de este formato de anotación a un registro textual tan peculiar, en el que los textos van acompañados de elementos multimedia como enlaces, fotos y videos, emojis y puntuación no estandarizada, ha abierto varios problemas en la comunidad de Universal Dependencies, muchos de los cuales siguen siendo objeto de un debate abierto y acalorado en la actualidad.

En este trabajo, por lo tanto, se presenta una descripción exhaustiva del formato de anotación morfosintáctica de UD, en particular, subrayando las cuestiones más relevantes en cuanto a su aplicación a los UGC generados en las redes sociales. El objetivo final es analizar y comprobar si estas anotaciones morfosintácticas sirven para obtener información útil para los sistemas/modelos de detección de la ironía y del stance o posicionamiento. Se presentarán dos subáreas de análisis de sentimientos y se utilizarán como ejemplos de estudio para probar las hipótesis de la investigación: el primer caso se centra en el área de la detección automática de la ironía y el segundo en el área de la detección del stance o posicionamiento. En ambos casos, se proporcionan los antecendentes y trabajos relacionados notas históricas que pueden servir de contexto para el lector, se introducen/plantean los problemas encontrados y se describen las distintas actividades propuestas para resolver estos problemas en la comunidad de la lingüística computacional. Se presta especial atención a los recursos actualmente disponibles, así como a los desarrollados específicamente para el estudio de los fenómenos antes mencionados. Finalmente, a través de la descripción de una serie de experimentos, llevados a cabo tanto en campañas de evaluación como en estudios independientes, se describe la contribución que la sintaxis puede brindar a la resolución de esas tareas, como subcampos del análisis de sentimientos.

Esta tesis es el resultado de toda la investigación que he llevado a cabo durante mi doctorado en una colección revisada de mi carrera de doctorado de los últimos tres años y medio, y se ubica dentro de la tendencia creciente de estudios dedicados a hacer que los resultados de la Inteligencia Artificial sean más explicables, yendo más allá del logro de puntajes más altos en la realización de tareas, sino más bien haciendo comprensibles sus motivaciones y qué los procesos sean más comprensibles para los expertos en el dominio.

La contribución principal y más novedosa de este trabajo consiste en la explotación de características (o rasgos) basadas en la morfología y la sintaxis de dependencias, que se utilizaron para crear las representaciones vectoriales de textos procedentes de redes sociales en varios idiomas y para dos tareas diferentes. A continuación, estas características se han emparejado/combinado con

una variedad de clasificadores de aprendizaje automático, con algunas redes neuronales y también con el modelo de lenguaje BERT.

Los resultados sugieren que la información sintáctica basada en dependencias utilizada es muy informativa para la detección de la ironía y menos informativa en lo que respecta a la detección del posicionamiento. No obstante, la sintaxis basada en dependencias podría resultar útil en la tarea de detección del posicionamiento si, en primer lugar, la detección de ironía se considera un paso previo al procesamiento en la detección del posicionamiento. También creo que el enfoque basado casi completamente en sintaxis de dependencias que propongo en esta tesis podría ayudar a explicar mejor un fenómeno pragmático tan difícil de detectar e interpretar como la ironía. De hecho, los diversos estudios que se presentan permitieron analizar si las estructuras sintácticas, independientemente del idioma , pueden aportar información útil para comprender y clasificar si un mensaje es irónico o no.