

Del discurso a la acción: Clasificación de actos de habla en textos legislativos

From Discourse to Action: Classification of Speech Acts in Legislative Texts

Doaa Samy

Cairo University, Egypt

Universidad Complutense de Madrid, España

doasamy@cu.edu.eg

dkhalil@ucm.es

Resumen: Los actos de habla son unidades básicas de la comunicación lingüística que permiten realizar acciones a través del lenguaje. En el texto legislativo, los diferentes tipos de actos de habla cobran especial relevancia a nivel pragmático porque detrás del lenguaje, existe una intención (acto ilocutivo) que va más allá de las palabras para organizar y cambiar la realidad en una sociedad. La teoría lingüística propone diferentes tipos de actos de habla. Este trabajo tiene como objetivo clasificar automáticamente tres tipos por su relevancia en el texto legislativo: 1) Los actos asertivos que describen hechos y realidades; 2) los actos directivos que definen las normas o regulan las relaciones y las competencias de la materia en cuestión; y 3) los actos compromisorios que reconocen los derechos y se comprometen a velar por estos derechos. Para la clasificación, se ha anotado un conjunto de 1325 enunciados divididos en subconjuntos de entrenamiento, validación (80%-20%) y un conjunto de prueba (250 enunciados). Se han entrenado y se han evaluado varios clasificadores automáticos multi-etiqueta y multiclase basándose en tres tipos de modelos: modelos clásicos de aprendizaje automático, modelos fundacionales del lenguaje (*LLMs*) de tipo “encoder” y un modelo fundacional generativo de tipo “decoder” mediante instrucciones *prompting* de 5 niveles (GPT 3.5). Los clasificadores basados en modelos “encoder” (*BERT* y *RoBERTaLex*) han obtenido los mejores resultados. *BERT* ha alcanzado un *f1-macro* de 0,85 y un *f1-micro* de 0,87. *RoBERTaLex* ha obtenido 0,86 en *f1-macro* y *f1-micro*.

Palabras clave: Modelos fundacionales, actos de habla, pragmática computacional, procesamiento de textos legales.

Abstract: Speech acts are basic units of linguistic communication which perform actions through words. Certain types of speech acts are especially significant in legislative texts as they go beyond words revealing intentions aiming at shaping the reality of a society. The linguistic theory proposes different types of speech acts. However, this study focuses on the automatic classification of three types for their relevance in the legislative context including: 1) Assertive acts describing events and reality; 2) directive acts setting regulations and, finally, 3) commissive acts indicating commitment to basic rights and principles. For the training and evaluation, a dataset of 1325 statements was manually labeled and further splitted into train and validation sets (80%-20%). Then, the resulting trained classifiers were further evaluated against a test dataset of 250 statements. Different classifier were trained over three types of models: Classical machine learning models, foundational Large Language Models (*LLMs*) based on “encoders”; namely *RoBERTaLex* and *BERT* and finally, generative models based on “decoders”, namely *GPT3.5* through a 5-shot prompt tuning. The classifier based on encoder *LLMs* (*BERT* and *RoBERTa*) outperformed the rest of models. *BERT* achieved *f1-macro* score of 0.85 for all classes and a *f1-micro* score of 0.87 (*BERT*) and 0.86 (*RoBERTa*).

Keywords: Language models, Speech Acts, Computational Pragmatics, Legal Text Processing.

1 Introducción y marco teórico

Los actos de habla son unidades básicas de la comunicación lingüística que permiten realizar acciones a través del lenguaje (Instituto Cervantes, 2023). El primero en introducir esta teoría fue el filósofo británico Austin (1962). Su obra póstuma *Cómo hacer cosas con las palabras* asienta las bases de esta teoría. Siete años más tarde, su discípulo Searle (1969) consolidó la teoría acuñando el término “actos de habla”.

Según esta teoría, en un acto de habla se activan tres dimensiones simultáneamente: la dimensión locutiva, ilocutiva y perlocutiva. La dimensión locutiva se centra en el mensaje refiriéndose al acto físico de emitir el mensaje y su formulación fónica, gramatical y semántica. La ilocutiva se centra en la intención del emisor y la acción que quiere realizar más allá de sus palabras. Por último, la dimensión perlocutiva se centra en el efecto provocado en el receptor por el enunciado.

En la legislación, nos encontramos ante una situación comunicativa concreta donde el emisor es la autoridad competente que expide la ley, i.e. el texto legislativo (entendido como mensaje) y un receptor genérico representando a múltiples posibles destinatarios como ciudadanos, profesionales del derecho, juristas, órganos judiciales, autoridades competentes de aplicar la ley, instituciones públicas o privadas, administraciones del Estado, etc.

En esta situación comunicativa, los actos de habla ilocutivos cobran especial relevancia porque mediante estos actos, el órgano legislador pretende llevar a la realidad acciones que implican organizar las relaciones y establecer las normas.

El marco teórico establece diferentes tipologías de los actos de habla (Austin, 1962; Searle, 1969; Vendler, 1980; Bach y Harnish, 1979). No obstante, el presente estudio se centra en un subconjunto de tres tipos de actos por su relevancia en el contexto legislativo. Los actos asertivos que describen los hechos. Los actos directivos que establecen las normas. Este tipo de actos es el más común en los textos legislativos. Finalmente, los actos compromisorios que reconocen los derechos o muestran un compromiso por velar y proteger estos derechos. Para el resto de los actos fuera del alcance de este estudio se ha contemplado una cuarta categoría “Otros”.

El estudio se estructura en siete secciones incluyendo esta introducción. En la Sección 2 se define el objetivo general, los objetivos específicos y el alcance. La Sección 3 ofrece un análisis del estado de la cuestión. La metodología y la descripción de los datos se describen en la Sección 4. El proceso de anotación se detalla en la Sección 5. Los experimentos para entrenar y adaptar los modelos con los resultados obtenidos se explican en la Sección 6. Finalmente, las conclusiones resumen los principales hallazgos.

2 Objetivo y alcance

2.1 Objetivo

El presente trabajo tiene como objetivo clasificar automáticamente los actos de habla en los textos legislativos distinguiendo entre los actos asertivos, directivos, compromisorios y otros actos que no se incluyen bajo estas tres categorías. Es una clasificación multiclase y multi-etiqueta porque en un solo enunciado se puede expresar más de una intención. Por ejemplo, un enunciado puede indicar una norma (acto directivo) a la vez que reconocer un derecho (un acto compromisorio).

Para alcanzar el objetivo general se plantean los siguientes objetivos específicos:

- Recopilar una muestra representativa de actos de habla a partir de un corpus de textos legislativos.
- Establecer unos criterios para la anotación y clasificación manual de esta muestra.
- Anotar un conjunto de datos de referencia “gold-standard” que sirva tanto para el entrenamiento como para la validación de los diferentes modelos.
- Entrenar clasificadores automáticos basados en diferentes familias de modelos: 1) modelos clásicos de aprendizaje automático (Ej. *RandomForest*, *OneVsRestClassifier*), 2) modelos fundacionales de lenguaje de tipo “encoder” (*BERT* y *RoBERTaLex*; ambos son modelos pre-entrenados con datos en español) y 3) modelos fundacionales generativos de tipo “decoder” (*GPT 3.5*).
- Evaluar los resultados obtenidos de los diferentes modelos.

Explorar este abanico de modelos permite comparar y evaluar la capacidad de las técnicas actuales en abordar esta tarea y, por tanto, valorar su viabilidad y su posible impacto en

soluciones y servicios finales destinados a diferentes tipos de usuarios en el dominio legal. Por ejemplo, la clasificación de actos de habla puede asistir a los profesionales del derecho en extraer la información relevante de los grandes volúmenes de textos legislativos distinguiendo de forma más rápida y eficiente entre los enunciados directivos que establecen una norma de los enunciados compromisorios que reconocen un derecho. Por otro lado, para un ciudadano este tipo de clasificación le puede ayudar a identificar los derechos y las obligaciones en una ley.

Asimismo, esta clasificación puede ser un módulo a integrar en sistemas conversacionales interactivos, sistemas de pregunta-respuesta o en soluciones para la simplificación del texto legislativo, la generación de resúmenes automáticos, etc.

2.2 Alcance

El estudio se centra en la dimensión ilocutiva referente a la intención o la acción que se pretende realizar con las palabras del enunciado. Dentro de esta dimensión ilocutiva, se han seleccionado tres tipos de actos: asertivos, directivos y compromisorios por su relevancia en el texto legislativo siguiendo la clasificación de López-Hernández (2005).

Otros estudios de carácter puramente lingüístico sin un componente computacional han contemplado otros tipos de actos de habla como los cualificatorios o los realizativos/performativos. Los primeros sirven para definir conceptos y los segundos convierten en realidad lo que se dice en el enunciado (López-Hernández, 2005).

Este último tipo es más significativo en los textos jurisprudenciales o sentencias. No se han contemplado en este estudio porque su uso en los textos legislativos es poco significativo y se limita a fórmulas tradicionales para declarar la entrada en vigor de una ley. Por ejemplo, al principio de una ley:

Rey de España: A todos los que la presente vieren y entendieren. Sabed: Que las Cortes Generales han aprobado y Yo vengo en sancionar la siguiente ley orgánica.

Además, los enunciados al final de una ley. Por ejemplo “Por tanto, Mando a todos los españoles, particulares y autoridades, que guarden y hagan guardar esta ley orgánica”.

Partiendo de los objetivos y el alcance, el presente estudio plantea tres cuestiones que se detallan a continuación:

- ¿Es viable abordar los actos de habla en el texto legislativo desde una perspectiva computacional mediante el entrenamiento de clasificadores automáticos?
- ¿Qué clasificadores automáticos realizarían mejor la tarea?
- ¿Cuáles son los retos que presenta esta tarea?

3 Estado de la cuestión

El análisis del estado de la cuestión llevado a cabo revela dos enfoques principales: 1) Los estudios de Lingüística y la Filosofía del Derecho que han tratado los actos de habla en el dominio legal desde una perspectiva pragmática y 2) los estudios computacionales que han abordado los actos del habla como un problema de clasificación automática.

Según este análisis, ningún estudio ha abordado los actos del habla en los textos legales como un problema de clasificación automática.

De ahí la aportación del presente estudio donde se retoma la teoría de los actos de habla desde un enfoque pragmático computacional. A la vez, es un enfoque aplicado a los textos legislativos que ofrece una solución automática empleando las técnicas actuales en el ámbito del Procesamiento del Lenguaje Natural (PLN) y la Inteligencia Artificial (IA).

3.1 Actos de habla en textos legales

Numerosos estudios han abordado los actos de habla en los textos legales. No obstante, todos estos estudios han tratado el tema desde una perspectiva puramente lingüística y pragmática o a veces desde la perspectiva de la filosofía del derecho. Ningún estudio de los señalados ha planteado una solución computacional automatizada. Además, varios se centran en el análisis de estos actos en inglés.

En cuanto a tipologías y clasificaciones de los actos de habla, se destaca el estudio pionero de Kurzon (1986) sobre los actos de habla en textos legales. Maccormick y Bankowski (1991) compararon los actos de habla con los actos jurídicos en su trabajo titulado “La teoría de los actos de habla y la teoría de los actos jurídicos”.

Siguiendo esta línea, Visconti (2009) agrupó los actos de habla en actos *a praxis* y actos *a poiésis*. Los primeros representan acciones sin pretender cambiar la realidad (Ej. definir

conceptos, constatar realidades). El segundo tipo son actos que pretenden cambiar la realidad del mundo a nuestro alrededor (Ej. ordenar, prometer, acusar, etc.). Por último, Durant y Leung (2016) prestaron especial interés a los actos realizativos/performativos.

Por otro lado, se destacan estudios aplicados que analizaron los actos de habla en diferentes tipos de textos legales. En este sentido, Blom y Trosborg (1992) y Trosborg (1995) analizaron cuantitativamente los actos de habla en estatutos y contratos en inglés basándose en un corpus limitado. En un estudio más reciente, Janicki (2018) ha retomado el análisis de los actos de habla en contratos.

Bernal (2007) analizó los actos de habla en las decisiones judiciales. Por otro lado, Mey (2013) abordó los actos de habla en contratos, acuerdos matrimonios, sentencias, etc. desde un enfoque diacrónico. Recientemente, Kone (2020) ha analizado los actos de habla en los tratados de las Naciones Unidas. Finalmente, en su estudio titulado “When does Speech Perform Regulable Action?”, Weston (2022) ha criticado la teoría de los actos de habla aplicándola a la regulación de la libertad de expresión.

En cuanto a los estudios de actos de habla en textos legales en español, se destaca el estudio de López-Hernández (2005) en el que propuso una clasificación de las normas jurídicas como enunciados de actos ilocutivos.

Otros estudios se han centrado en actos de habla concretos o en documentos legales específicos. En esta línea, destacamos los trabajos de Cifuentes-Honrubia (2005; 2006; 2009) sobre la “autorización” como acto de habla. También, los trabajos que enfocaron los actos realizativos/performativos en el discurso legal (Fiorito, 2006; Amorebieta y Vera, 2020). Moreu-Carbonell (2020) estudió los actos comunicativos en el lenguaje administrativo destacando la particularidad del lenguaje jurídico español. Por último, Ibáñez-Macías (2021) analizó los actos declaratorios/cualificatorios en el Derecho constitucional.

3.2 Clasificadores automáticos de actos de habla

Desde el enfoque computacional, varios trabajos han desarrollado clasificadores automáticos de actos de habla, pero no en el dominio legal, sino en otros dominios como las redes sociales, los textos de mensajería instantánea, los correos

electrónicos, los foros educacionales, etc. Estos estudios se han centrado en la dimensión ilocutiva para una mejor detección de los intentos.

En cuanto a la clasificación de actos en diálogos y mensajería instantánea, se destaca el trabajo pionero de Mast et al. (1996) basado en la clasificación semántica. Twitchell et al. (2004) partieron de la teoría de actos de habla para modelar las conversaciones en mensajería instantánea. Carvalho y Cohen (2005) abordaron los actos de habla en los correos electrónicos para detectar si se tratan de solicitudes o de compromisos. Moldovan, Rus y Graesser (2013) entrenaron un modelo supervisado con árboles de decisiones para las conversaciones en línea. Samei et al. (2014) entrenaron un clasificador de actos de habla para analizar los diálogos de sistemas inteligentes de tutoría (*Intelligent Tutoring Systems*). Arguello y Shaffer (2015) también entrenaron un clasificador para los actos de habla en los foros educativos de los cursos abiertos en línea (*MOOC*).

Recientemente, los actos de habla en *tweets* han sido objeto de varios trabajos. Zhang, Gao y Li (2011) desarrollan una metodología para reconocer “qué hacen los tweets”. Vosoughi y Roy (2016) también analizan los actos de habla en *tweets*. Por último, Saha, Saha y Bhattacharyya (2019) desarrollan un clasificador de actos de habla en *tweets* basado en redes neuronales.

4 Metodología y datos

Para el presente estudio se ha adoptado una metodología empírica y aplicada siguiendo cuatro fases principales:

- Recopilación de los datos.
- Extracción de enunciados.
- Anotación de los enunciados según los actos de habla objeto de este estudio.
- Entrenamiento y evaluación de los clasificadores automáticos.

Los desarrollos se han realizado en el lenguaje de *Python* utilizando las librerías abiertas de:

- *Stanza* (Qi et al., 2020) para el preprocesamiento del texto del corpus y *pandas* (McKinney, 2010) para la estructuración de los conjuntos de datos.
- *Scikit-learn* (Pedregosa et al., 2011) para el entrenamiento y la evaluación de los

modelos clásicos de aprendizaje automático.

- *Transformers* de *HuggingFace* (Wolf et al., 2020) para el entrenamiento y la evaluación de los modelos fundacionales de tipo “encoder”.
- Modelos fundacionales “encoder” pre-entrenados con textos en español: *BERT* [distil-bert] (Cañete et al., 2023) y *RoBERTaLex* (Gutiérrez-Fandiño et al., 2021). Están disponibles con acceso abierto en el repositorio de *HuggingFace*.
- El interfaz conversacional *ChatGPT* (OpenAI, 2023) para el experimento con el modelo fundacional generativo *GPT 3.5*.

4.1 El corpus de textos legislativos

Existen corpus legales como Legal-ES (Samy, Arenas-García y Pérez-Fernández, 2020), el corpus “Spanish Legalese Language Model and Corpora” (Gutiérrez-Fandiño et al., 2021) o el corpus titulado “Spanish monolingual corpus from contents of Spanish State Official Gazette” disponible en *European Language Grid* (2022).

Para este estudio, se ha partido de un subcorpus limitado de textos legislativos, dado que se trata de una tarea concreta y no es necesario contar con un corpus grande. El subcorpus utilizado incluye tres fuentes principales: El Código Civil, la Ley de Protección de Datos Personales y la Ley de Régimen Jurídico de las Administraciones Públicas.

Conjunto de datos	Nº <i>tokens</i>	Nº <i>types</i>
Código Civil	129251	9732
Ley Orgánica de Protección de Datos Personales y garantía de los derechos digitales	40681	3930
Ley de Régimen Jurídico de las Administraciones Públicas y del Procedimiento Administrativo Común	31613	3849
Total	201545	17511

Tabla 1: Textos legislativos para extraer la muestra de los actos de habla.

A continuación, incluimos ejemplos de los tres tipos de actos de habla: asertivos, directivos, compromisorios junto a la categoría “Otros”.

Acto de habla	Ejemplo
Asertivo	Es preciso ahora que el marco que regula el régimen jurídico de las Administraciones Públicas sea objeto de una adaptación normativa expresa que lo configure de forma armónica y concordante con los principios constitucionales.
Directivo	Si el recurso se hubiera interpuesto ante el órgano que dictó el acto impugnado, éste deberá remitirlo al competente en el plazo de diez días, con su informe y con una copia completa y ordenada del expediente.
Compromisorio	Los poderes públicos velarán por los derechos y las necesidades de las personas que hayan padecido daños causados por catástrofes.
Otros	A todos los que la presente vieren y entendieren. Principios de la potestad sancionadora Disposición adicional decimotercera

Tabla 2: Ejemplos de enunciados con los tipos de actos de habla.

4.2 Preprocesamiento

Se ha preprocesado el subcorpus para garantizar su calidad en cuanto a formato, codificación, etc.

En general, los textos procedentes de fuentes oficiales como el Boletín Oficial del Estado (BOE) garantizan una buena calidad en cuanto a formato con poco ruido textual. Por lo tanto, el esfuerzo de la depuración ha sido limitado a ajustar algunos saltos de líneas y casos puntuales de caracteres mal codificados. Se ha utilizado *Stanza* (Qi et al., 2020) para la segmentación de oraciones que han constituido la unidad de análisis para este trabajo.

Cabe destacar que la oración es una unidad de análisis sintáctico-semántico, mientras que los actos de habla es un fenómeno pragmático que pueda coincidir o no con la oración. Esto depende del contexto lingüístico y extralingüístico. Por esto, el fenómeno pragmático puede reflejarse en una o más oraciones. No obstante, se ha decidido basarse en la oración como unidad mínima de análisis en este estudio por razones prácticas de cara al procesamiento computacional.

5 Anotación de los actos de habla

La anotación es una piedra angular para el entrenamiento de modelos de aprendizaje supervisado. La calidad de los datos de entrenamiento y la coherencia de los criterios son factores principales para que el modelo aprenda y generalice mejor. Además de servir como base para el entrenamiento y la evaluación de los modelos de clasificación automática, un conjunto de referencia estándar “gold standard” es en sí un recurso lingüístico de interés.

Para garantizar la calidad del conjunto anotado, el proceso se ha llevado a cabo en tres pasos: 1) extracción de enunciados que representan los tipos de actos de habla en cuestión, 2) establecimiento de los criterios y la anotación manual y 3) validación de la anotación. A continuación, describimos brevemente cada paso.

5.1 La extracción de los actos de habla

Con el corpus preprocesado y segmentado en oraciones, se procede a la extracción de enunciados para anotar los actos de habla.

Existen algunos retos a la hora de afrontar la tarea de extracción de enunciados que representan los tres tipos de actos de habla objeto de este estudio.

Primero, no todas las oraciones contienen los tipos de actos de interés (asertivos, directivos y compromisorios). Segundo, los actos de habla son fenómenos pragmáticos a nivel del enunciado que no siempre coincide con el límite de la oración. Por lo tanto, la interpretación del acto de habla a nivel de oración requiere de un conocimiento del contexto extralingüístico por parte del anotador. Tercero, la ocurrencia de los tres tipos está desequilibrada.

Por ejemplo, identificar actos de habla directivos es más fácil porque son bastante frecuentes, mientras que los actos compromisorios son menos frecuentes y habría

que examinar muchos enunciados para localizar ejemplos de actos compromisorios.

Por estos motivos, se ha incluido una cuarta categoría “Otros” en la que se contemplan los enunciados que no se clasifican bajo ningún tipo de los tres tipos de actos señalados. Asimismo, se ha recurrido a algunas estrategias para aumentar los datos de los actos compromisorios para tener una cantidad suficiente a la hora de entrenar y evaluar los modelos.

Los datos anotados se han extraído en 3 conjuntos sumando un total de 1375 oraciones. Los dos primeros conjuntos consisten en muestras aleatorias.

Para el tercer conjunto, se han extraído muestras dirigidas de enunciados principalmente de tipo compromisorio. Se ha optado por esta estrategia para equilibrar esta categoría de cara al entrenamiento y la evaluación, ya que esta clase se ha quedado infrarrepresentada en las muestras aleatorias.

En la búsqueda dirigida solo aplicada al caso de los actos compromisorios, se han extraído como posibles candidatos enunciados que contengan lemas como “garantizar”, “velar”, “derecho”. Sin embargo, es importante señalar que no todas las menciones a derechos o garantías son necesariamente actos de habla compromisorios. Existen varios casos de ejemplos en los que no se aplican los criterios de actos compromisorios.

Por este motivo y para evitar generalizaciones imprecisas, es necesario tener en cuenta estas estrategias.

Primero, incluir ejemplos negativos donde aparecen formas de “velar”, “garantizar” o “derecho” sin que se trate de un acto compromisorio. Por ejemplo, en el enunciado “La Ley recoge esta concepción constitucional [...] y fija las garantías mínimas de los ciudadanos [...]”, no se trata de un acto compromisorio, sino de un acto asertivo que describe una realidad sobre la ley.

Segundo, resolver estos casos ambiguos mediante el proceso de anotación manual.

Tercero, incluir una clase “Otros” para aquellos casos en los que no se cumplen los criterios de ninguno de los tres tipos (asertivos, compromisorios o directivos).

De esta forma, se garantiza una mejor representatividad de la muestra.

La Figura 1 muestra la distribución en el conjunto completo de entrenamiento y validación (1075 enunciados) después de equilibrar la categoría de actos compromisorios.

Del total de 1375, se han descartado aproximadamente 50 oraciones porque presentaban problemas resultantes del proceso de segmentación. Por ejemplo, algunas eran solamente un dígito o un fragmento mal segmentado de una referencia a una ley por la presencia de barras o guiones, etc.

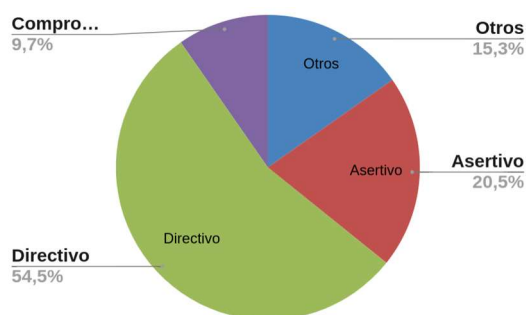


Figura 1: Distribución de actos de habla en el conjunto de entrenamiento y validación (1075 enunciados).

De ahí, el total anotado es 1325 enunciados con cuatro categorías: actos asertivos, directivos, compromisorios u otros. Para el proceso de entrenamiento y validación se ha utilizado un conjunto de 1075 enunciados. Para la prueba de los modelos ya entrenados, se ha utilizado un conjunto de 250 enunciados.

5.2 Criterios de anotación de los actos de habla

Para la anotación manual, se han establecido unos criterios para distinguir las tres categorías de actos de habla ilocutivos (asertivos, directivos y compromisorios) donde la intención es la base principal para la clasificación. Los enunciados que no se clasifican bajo ninguna de las tres categorías, se clasifican con la etiqueta “Otros”. Los criterios establecidos incluyen lo siguiente:

- **Asertivos.** Se anotan como actos asertivos las oraciones que definen una realidad o relatan unos hechos. Este tipo de actos es común en las disposiciones de una ley. En un texto de ley, las disposiciones suelen estar al principio del texto legislativo y suelen reflejar el contexto y los motivos por los cuales, se ha elaborado esta ley.
- **Directivos.** Se anotan como actos directivos los enunciados cuya interpretación indica establecer unas normas, regular ciertas relaciones o definir competencias. El acto directivo

puede ser directo empleando modalidades deónticas como “deber”, “haber de”, “estar prohibido”. También el acto directivo puede ser indirecto cuando se describen unos escenarios hipotéticos como “El responsable entregará en el plazo de un mes [...]”, “la autoridad competente convocará”, etc.

- **Compromisorios.** Se anotan como actos compromisorios aquellos enunciados que reconozcan los derechos y se comprometan a velar por ellos de forma directa o indirecta. En los casos directos, se indica de forma explícita como por ejemplo “velar por” o “respetando el derecho”. En los casos indirectos, el texto de la ley compromete a otros órganos o agentes indicando que es parte de su responsabilidad. Por ejemplo: “Es responsabilidad de [...] velar por el cumplimiento del derecho de igualdad”.
- **Otros.** En los casos donde no se cumple ninguno de los criterios establecidos, se anotan como “Otros”.
- En los casos ambiguos, se recurre al contexto original de dónde se ha extraído la muestra para aclarar dudas. Si persiste la ambigüedad, se da prioridad al acto directivo teniendo en cuenta su relevancia en el texto legislativo.
- **Casos multi-etiqueta.** Los enunciados que realizan dos acciones se anotan con los actos correspondientes (multi-etiqueta). Es decir, si en un enunciado, se describe una realidad y luego se establece una norma en el mismo enunciado, se anotan las dos categorías: el acto asertivo y el acto directivo.
- Del mismo modo, se anotan con las categorías directiva y compromisorias los enunciados donde a la vez se establece una norma y se reconoce el compromiso por un derecho. Ejemplo de ello es este enunciado:

Cada Administración pública establecerá los días y el horario en que deban permanecer abiertos sus registros, garantizando el derecho de los ciudadanos a la presentación de documentos.

En total, los enunciados con multi-etiquetas representan sólo el 3,4% del conjunto de entrenamiento y validación.

5.3 Validación de la anotación

Una vez concluido el proceso de anotación, se procede a la validación de las muestras anotadas, sobre todo, los casos ambiguos para tomar una decisión acerca de su clasificación.

En un escenario óptimo, la anotación debería llevarse a cabo por varios anotadores. Luego, en el proceso de validación se calculan los acuerdos entre anotadores para proceder a la armonización.

No obstante, en este proyecto, no se ha contado con recursos para tener un equipo de anotación. La mayoría del proceso se ha realizado por un solo anotador, un lingüista experto. Por eso, la validación consistía en revisar los casos ambiguos y asegurarse del cumplimiento de los criterios de una forma coherente.

6 Entrenamiento

El total de enunciados anotados y validados (1325) se han dividido en: 1075 enunciados para el entrenamiento y la validación, por un lado y 250 enunciados como un conjunto de prueba final para evaluar los modelos entrenados. Los datos anotados y los modelos entrenados se encuentran disponibles en el repositorio de *Github*.¹

El primer conjunto de 1075 enunciados se ha dividido de forma aleatoria en dos subconjuntos de entrenamiento (80%) y validación (20%) asegurándose de que: 1) estén representadas todas las categorías; 2) los conjuntos de entrenamiento y validación sean iguales en todos los experimentos para garantizar la comparabilidad de los resultados

6.1 Modelos clásicos de aprendizaje automático

Se han entrenado clasificadores basados en modelos clásicos disponibles en la librería de *scikit-learn* como *RandomForest* u *OneVsRestClassifier* en combinación con vectorizadores como *Tf-idf* o *HashingVectorizer*. Estos modelos tienen la ventaja de ser eficientes y requieren pocos recursos de cómputo para entrenarse. El inconveniente es que dependen de las palabras, i.e. el léxico empleado y, por lo tanto, tienen limitaciones a la hora de generalizar.

Los mejores resultados se han obtenido entrenando un clasificador *OneVsRestClassifier* con un vectorizador *Tf-idf* obteniendo un *f1-macro* de 0,74 y un *f1-micro* de 0,77 en el conjunto de prueba.

Los actos directivos y compromisorios obtuvieron mejores resultados que los actos asertivos. En nuestra opinión, esto se debe a: 1) los actos directivos son los más frecuentes; 2) los actos asertivos, en cambio, son más ambiguos y más variados sin rasgos distintivos lingüísticamente y 3) los actos compromisorios, sí, se caracterizan por el empleo de un léxico concreto o patrones más específicos.

La Tabla 3 muestra los resultados obtenidos por este tipo de modelos. No obstante, para una evaluación más rigurosa, se ha realizado una validación cruzada con 5 y 10 pliegues. Con la validación cruzada, los resultados bajan significativamente demostrando la limitación de estos modelos.

Modelos clásicos			
Conjunto: Validación			
	Precisión	Recall	f1-score
<i>Asertivo</i>	0,87	0,38	0,53
<i>Directivo</i>	0,81	0,87	0,84
<i>Compromisorio</i>	0,91	0,77	0,83
<i>Otros</i>	0,76	0,74	0,75
f1-score micro	0,77		
f1-score macro	0,74		
Validación cruzada			
	CV = 5	CV =10	
<i>f1-micro</i>	0,48	0,53	
<i>f1-macro</i>	0,23	0,22	
<i>Accuracy (mean)</i>	0,36	0,42	
Conjunto: Prueba (250 enunciados)			
<i>Asertivo</i>	0,61	0,19	0,29
<i>Directivo</i>	0,91	0,72	0,84
<i>Compromisorio</i>	0,70	0,47	0,56
<i>Otros</i>	0,69	0,62	0,65
f1-score micro	0,70		
f1-score macro	0,58		

Tabla 3: Resultados del modelo clásico *OneVsRestClassifier*.

¹<https://github.com/dosamy/SpeechActs-Legislative-Spanish>

6.2 Modelos fundacionales basados en *encoder*

Los grandes modelos fundacionales han supuesto un salto cualitativo y cuantitativo en el panorama del PLN. Los modelos de tipo “encoder” han demostrado una mejora sustancial en las tareas de comprensión del lenguaje (*Natural Language Understanding - NLU*).

La hipótesis inicial considera que basarse en este tipo de modelos “encoder” puede suponer un avance significativo respecto a los modelos clásicos de aprendizaje automático abordados en la sección anterior.

Se han realizado varias rondas de entrenamiento para la adaptación *fine-tuning* de dos modelos fundacionales de tipo “encoder”; ambos pre-entrenados y adaptados a la lengua española: *RoBERTaLex* y *BERT*. El primero es pre-entrenado con textos del dominio legal, lo cual se supone que pueda obtener mejores resultados en nuestra tarea. El segundo es pre-entrenado con textos del dominio general.

Ambos modelos se han entrenado mediante la librería de Transformers en 10 épocas aplicando los mismos parámetros (*Learning rate=2e-5, batch_size=8, weight*) en un ordenador con CPU.

Las Tablas 4 y 5 muestran los resultados obtenidos por *RoBERTaLex* y *Distil-BERT*.

Modelos fundacionales <i>encoder</i> : RoBERTaLex			
Conjunto: Validación			
	<i>Precisión</i>	<i>Recall</i>	<i>f1-score</i>
<i>Asertivo</i>	0,75	0,73	0,74
<i>Directivo</i>	0,89	0,92	0,90
<i>Compromisorio</i>	0,92	0,92	0,92
<i>Otros</i>	0,86	0,88	0,87
f1-score micro	0,86		
f1-score macro	0,86		
Conjunto: Prueba (250 enunciados)			
<i>Asertivo</i>	0,59	0,47	0,52
<i>Directivo</i>	0,95	0,83	0,88
<i>Compromisorio</i>	0,82	0,60	0,69
<i>Otros</i>	0,65	0,96	0,78
f1-score micro	0,78		
f1-score macro	0,72		

Tabla 4: Resultados del modelo *RoBERTaLex*.

Los resultados de *RoBERTaLex* como un modelo fundacional pre-entrenado en el dominio

legal superan significativamente los modelos clásicos alcanzando un *f1-score micro* de 0,86 y un *f1-score macro* de 0,86 en el conjunto de validación, y 0,78 y 0,72 respectivamente en el conjunto de prueba.

Por otro lado, se ha replicado el mismo proceso de adaptación “fine-tuning” con el modelo *BERT* pre-entrenado para la lengua española.

Modelos fundacionales <i>encoder</i> : BERT			
Conjunto: Validación			
	<i>Precisión</i>	<i>Recall</i>	<i>f1-score</i>
<i>Asertivo</i>	0,71	0,69	0,70
<i>Directivo</i>	0,90	0,94	0,92
<i>Compromisorio</i>	1,00	0,88	0,94
<i>Otros</i>	0,85	0,85	0,85
f1-score micro	0,87		
f1-score macro	0,85		
Conjunto: Prueba (250 enunciados)			
<i>Asertivo</i>	0,55	0,47	0,50
<i>Directivo</i>	0,96	0,83	0,89
<i>Compromisorio</i>	0,92	0,80	0,86
<i>Otros</i>	0,64	0,89	0,74
f1-score micro	0,78		
f1-score macro	0,75		

Tabla 5: Resultados del modelo *BERT*.

Las diferencias entre los modelos *RoBERTaLex* y *BERT* son sutiles. Los resultados obtenidos por *BERT* son mejores, pero es una diferencia mínima. En este sentido, es un resultado inesperado, ya que la hipótesis inicial suponía que el modelo de *RoBERTaLex* obtendría los mejores resultados al estar adaptado al dominio legal en español. No obstante, esta ventaja no ha supuesto una diferencia significativa en la tarea en cuestión.

Por otro lado, es evidente que la tarea sigue suponiendo cierta dificultad, sobre todo en los actos asertivos por su gran variedad y por su difícil distinción de otros tipos de actos.

Pese a ello, cabe destacar que los modelos pre-entrenados han supuesto una mejora significativa respecto a los modelos clásicos, sobre todo en la cobertura.

No obstante, los resultados empeoran en el conjunto de prueba, lo cual podría indicar que los modelos no generalizan lo suficientemente bien y requieren más datos de entrenamiento.

6.3 Pruebas con modelos fundacionales generativos basados en *decoder*

Por último y dado el gran éxito que están teniendo los modelos fundacionales generativos, se ha realizado una prueba con el modelo *GPT3.5* siguiendo una instrucción con 5 ejemplos (*5-shot prompting*) a través del interfaz conversacional ChatGPT.

Estos modelos están entrenados para realizar varias tareas de forma genérica, por lo tanto, utilizarlos en dominios y tareas específicas requiere una adaptación o bien mediante un proceso de “fine-tuning” del modelo o bien a través de una serie de instrucciones “prompting”. La adaptación “fine-tuning” es costosa porque supone una modificación de los parámetros del pre-entrenamiento y requiere de procesadores gráficos (*GPU*) y grandes capacidades de cómputo. La adaptación mediante instrucciones es más viable, pero ofrece resultados poco precisos.

Para la prueba realizada, se ha creado una instrucción con 5 ejemplos y se ha introducido esta instrucción a través del interfaz de diálogo *ChatGPT*. En la instrucción, se le ha indicado al modelo el objetivo de la tarea y las clases a tener en cuenta².

Se ha introducido el conjunto de prueba formado por 250 enunciados y se han calculado las métricas de los resultados obtenidos por *GPT 3.5*. Con 5 ejemplos y sin un proceso de adaptación es difícil que un modelo generativo realice una tarea tan específica en un dominio tan concreto como el legislativo. Pese a ello, en los actos directivos, la categoría más frecuente, el modelo *GPT 3.5* ha alcanzado un *f1-score* de 0.70.

<i>ChatGPT 3.5 - Conjunto: Prueba</i>	
f1-score micro	0,56
f1-score macro	0,48

Tabla 6: Resultados de *ChatGPT 3.5* (*5-shots*).

7 Conclusiones

El problema de clasificación al que nos enfrentamos en este estudio es un problema multi-etiqueta y multiclase de una dimensión de cuatro clases. Es un problema complejo a nivel lingüístico por la ambigüedad de algunos actos.

² La instrucción está disponible en el repositorio de *GitHub*.

Los resultados obtenidos con modelos que representan el estado del arte son bastante satisfactorios (0,87 de *f-score*) aunque dejan un margen de mejora.

A nuestro modo de ver, los retos que supone la tarea se resumen en lo siguiente:

- Es una tarea que se basa principalmente en un criterio pragmático, i.e. la intención del emisor, que va más allá del texto. A pesar de que el emisor en el texto legislativo es una autoridad competente que suele seguir un estilo de lenguaje normalizado, el hecho de que se trate de la interpretación de la intención puede ser un reto, incluso para la anotación humana.
- La segmentación y la variabilidad de la extensión de los enunciados es otro factor que puede suponer un reto porque a veces se trata de enunciados cortos y otras veces se trata de enunciados extensos y complejos.
- Los enunciados que se clasifican con más de una etiqueta es otro reto por su complejidad conceptual y porque solo representan un porcentaje limitado que no alcanza el 5%.
- A modo de conclusión, los resultados obtenidos son satisfactorios, pero para una mejora, es necesario ampliar el conjunto de datos de entrenamiento para incluir más casos de multi-etiqueta, de actos asertivos y de actos compromisorios.

Por otro lado, una línea de mejora es la adaptación de un modelo generativo como *Llama 2* o *GPT 4.0*. No obstante, es un proceso que requiere de mayores recursos de cómputo.

Por último, con este estudio se abre una línea de investigación para evaluar la utilidad de incluir este tipo de clasificadores en algunos servicios y productos. Sistemas de preguntas-respuestas, generación de resúmenes, simplificación de textos legislativos son ejemplos de soluciones que pueden beneficiarse de la clasificación como un paso relevante para mejorar el resultado final.

Bibliografía

Amorebieta y Vera, J. 2020. Las sentencias de la Corte Penal Internacional: análisis preliminar del uso de los verbos performativos. *Quintú*

- Quimün Revista De lingüística*. Número 4, páginas 1-13.
- Arguello, J. y K. Shaffer. 2015. Predicting speech acts in MOOC forum posts. En *Proceedings of the International AAAI Conference on Web and Social Media, ICWSM*, 9:1, páginas. 2-11, Oxford (Reino Unido).
- Austin, J. L. 1962. *Cómo hacer cosas con palabras*. Paidós, Barcelona.
- Bach, K. y R.M. Harnish. 1979. *Communication and Speech Acts*. MIT Press, Massachusetts.
- Bernal, C. L. 2007. A Speech Act Analysis of Judicial Decisions. *European Journal of Legal Studies*. 1:2, páginas 1-24.
- Blom, B., y A. Trosborg. 1992. An Analysis of Regulative Speech Acts in English Contracts - Qualitative and Quantitative Methods. *HERMES - Journal of Language and Communication in Business*. 5:9, páginas 83-111.
- Cañete, J., G. Chaperon., R. Fuentes, J. Ho, H. Kang y J. Pérez. 2020. Spanish Pre-Trained BERT Model and Evaluation Data. En *Proceedings of Practical ML for Developing Countries: learning under limited/low resource scenarios Workshop at the Eighth International Conference on Learning Representations, PML4DC@ICLR 2020*, páginas 1-9, Addis Ababa (Etiopía). Modelo disponible en: <https://huggingface.co/dccuchile/distilbert-base-spanish-uncased>
- Cao, D. 2007. Legal Speech Acts as Intersubjective Communicative Action. En Wagner, A., W. Werner y D. Cao (Eds.), *Interpretation, Law and the Construction of Meaning*, páginas 65-82. Springer, Dordrecht.
- Carvalho, V. R. y W. Cohen. 2005. On the collective classification of email speech acts. En *Proceedings of the 28th Annual ACM/SIGIR International Conference on Research and Development in Information Retrieval*, páginas. 345-352, Salvador (Brasil).
- Cifuentes-Honrubia, J.L. 2006. *Los actos de habla*. Liceus, Madrid.
- Cifuentes-Honrubia, J.L. 2009. Autorización: actos de habla e interpretación jurídica. En Jiménez Ruiz, J.L. y L. Timofeeva (Eds.), *Estudios de lingüística: Investigaciones lingüísticas en el siglo XXI, ELUA*, páginas 45-61, Universidad de Alicante.
- Durant, A. y H. Leung. 2016. *Language and Law*. Routledge, Oxon.
- European Language Grid. 2022. *Spanish monolingual corpus from contents of Spanish State Official Gazette*. Dataset (Text corpus). Disponible en: <https://live.european-language-grid.eu/catalogue/cpid/mxaKn3VZmpuKCfGhEdSPUX/>
- Fiorito, L. 2006. On Performatives in Legal Discourse. *Metalogicon*, 19:2, páginas 101-112.
- Gutiérrez-Fandiño, A., J. Armengol-Estapé, A. Gonzalez-Agirre y M. Villegas. 2021. *Spanish legalese language model and corpora*. arXiv. DOI: arXiv:2110.12201. Modelo disponible en: <https://huggingface.co/PlanTL-GOB-ES/RobERTale>
- Ibáñez-Macías, A. 2021. Las declaraciones como actos de habla en el Derecho constitucional. *Revista Telemática de Filosofía del Derecho*. (24), páginas 3-36.
- Janicki, E. A. 2018. Contracts as speech acts: Bringing Jakobson to the conversation. *GeorgeTown Law Journal*. 107:2018, páginas 201-233.
- Kone, N. 2020. Speech Acts in UN Treaties: A Pragmatic Perspective. *Open Journal of Modern Linguistics*. 10:6, páginas 813-827.
- Kurzon, D. 1986. *It is Hereby Performed: Explorations in Legal Speech Acts*. John Benjamins Publishing, Amsterdam.
- López-Hernández, J. 2005. Clasificación de las normas jurídicas como enunciados de actos ilocutivos. *Anuario de derechos humanos*. (6), páginas. 455-510.
- López-Hernández, J. 2005. Las normas jurídicas como actos ilocutivos: Conceptos y clases. *Cuadernos Electrónicos de Filosofía del Derecho*. (11), páginas 1-48.
- MacCormick, N. y Z. Bankowski. 1991. La teoría de los actos de habla y la teoría de los actos jurídicos. *Anuario de Filosofía del derecho*. (8), páginas 219-238.

- Mast, M., H. Niemann, E. Nöth y E.G. Schukat-Talamazzini. 1996. Automatic classification of dialog acts with Semantic Classification Trees and Polygrams. En: Wermter, S., E. Riloff, y G. Scheler. (Eds.), *Connectionist, Statistical and Symbolic Approaches to Learning for Natural Language Processing. International Joint Conference on Artificial Intelligence, IJCAI 1995*. Lecture Notes in Computer Science. (1040), páginas 217-229, Springer, Berlin, Heidelberg.
- McKinney, 2010. Data structures for statistical computing in python. En *Proceedings of the 9th Python Science Conference*. (445), páginas 56-61.
- Mey, J. L. 2013. Atos de fala e a Lei. *Cadernos De Linguagem E Sociedade*. 4:2, páginas 11–27.
- Moldovan, C., V. Rus y A.C. Graesser. 2011. Automated Speech Act Classification For Online Chat. En *Proceedings of the 22nd Midwest AI and Cognitive Science Conference, MAICS*, páginas 23-29, Ohio (Estados Unidos).
- Moreu Carbonell, E. 2020. Nuestro lenguaje: el giro lingüístico del derecho. *Revista de Derecho Público: Teoría y Método*. (1), páginas 313-362.
- OpenAI 2023. ChatGPT [Large language model]. Disponible en: <https://chat.openai.com/chat>
- Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel y E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*. (12), páginas 2825-2830.
- Qi, P., Y. Zhang, Y. Zhang, J. Bolton. y C.D. Manning. 2020. Stanza: A Python natural language processing toolkit for many human languages. En *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020*, páginas 101–108.
- Robles-Morchón, G. 2009. *Comunicación, Lenguaje y Derecho: algunas ideas básicas de la teoría comunicacional del Derecho*. Real Academia de Ciencias Morales y Políticas, Madrid.
- Saha, T., S. Saha y P. Bhattacharyya. 2019. Tweet act classification: A deep learning based classifier for recognizing speech acts in Twitter. En *International Joint Conference on Neural Networks, IJCNN*, páginas 1-8, Budapest (Hungary).
- Samei, B., H. Li, F. Keshtkar, V. Rus y A.C. Graesser. 2014. Context-based speech act classification in intelligent tutoring systems. En *International conference on Intelligent Tutoring Systems, ITS*, páginas 236-241, Honolulu Hi (Estados Unidos).
- Samy, D., J. Arenas-García y D. Pérez-Fernández. 2020. Legal-ES: A Set of Large-Scale Resources for Spanish Legal Text Processing. En Samy, D. J. Arenas-García y D. Pérez-Fernández (Eds.) *Proceedings of Workshop on Language Technologies in Government and Public Administration, LT4Gov@LREC-2020*, páginas 32-36, Marseille (Francia).
- Searle, J. 1969. *Speech Acts*. Cambridge University Press, Cambridge.
- Trosborg, A. 1995. Statutes and contracts: An analysis of legal speech acts in the English language of the law. *Journal of Pragmatics*. 23:1, páginas 31-53.
- Twitchell, D.P, M. Adkins, J.F. Nunamaker y J.K. Burgoon. 2004. Using Speech Act Theory to Model Conversations for Automated Classification and Retrieval. En *Proceedings of the 9th International Working Conference on the Language-Action Perspective on Communication Modelling, LAP*, páginas 121-130, Tilburg (Holanda).
- Vendler, Z. 1980. Telling the facts. En Searle, J.R., F. Kiefer y M. Bierwisch (Eds.), *Speech Act Theory and Pragmatics. Texts and Studies in Linguistics and Philosophy*, páginas 273-290. Springer, Dordrecht.
- Visconti, J. 2009. Speech acts in legal language: introduction. *Journal of Pragmatics*. 41:3, páginas 393-400.
- Vosoughi, S., y D. Roy. 2016. Tweet Acts: A Speech Act Classifier for Twitter. En *Proceedings of the International Conference on Web and Social Media, AAAI*. 10:1, páginas 711-714, Cologne (Alemania).
- Weston, J. 2022. When Does Speech Perform Regulable Action? — A Critique of Speech Act Theory’s Application to Free Speech

- Regulation. *International Journal of Language and Law*. (11), páginas 78-97.
- Wolf, T., D. Lysandre, S. Víctor, J. Chaumon, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison, S. Shleifer, P. Von Platen, C. Ma, Y. Jernite, J. Plu, C. Xu, T. Le Scao, S. Gugger y M. Drame. 2020. Transformers: State-of-the-art natural language processing. En *Proceedings of the 2020 conference on Empirical Methods in Natural Language Processing: System Demonstrations, EMNLP*, páginas 38-45.
- Yin, H. y Y. Chen. 2020. Speech act in diplomacy: How China makes commitments in diplomatic press conference. *International Journal of Languages, Literature and Linguistics*. 6:4, páginas 170-176.
- Zhang, R., D. Gao y W. Li. 2011. What Are Tweeters Doing: Recognizing Speech Acts in Twitter. En *Analyzing Microtext: Papers from the 2011 AAAI Workshop*, páginas 86-91, San Francisco (Estados Unidos).