

4-1-2024

Deontic Modality in a comparable legislative corpus (Arabic-Spanish): A Tool for Automatic Detection and Contrastive Analysis

Follow this and additional works at: <https://jfa.cu.edu.eg/journal>



Part of the [Spanish and Portuguese Language and Literature Commons](#)

Recommended Citation

(2024) "Deontic Modality in a comparable legislative corpus (Arabic-Spanish): A Tool for Automatic Detection and Contrastive Analysis," *Journal of the Faculty of Arts (JFA)*: Vol. 84: Iss. 2, Article 9.

DOI: 10.21608/jarts.2023.217946.1410

Available at: <https://jfa.cu.edu.eg/journal/vol84/iss2/9>

This Original Study is brought to you for free and open access by Journal of the Faculty of Arts (JFA). It has been accepted for inclusion in Journal of the Faculty of Arts (JFA) by an authorized editor of Journal of the Faculty of Arts (JFA).

La modalidad deóntica en un corpus legislativo comparable (árabe-español): Una herramienta de detección automática y análisis contrastivo^(*)

Dr. Doaa Ahmed Samy
Profesor Asistente, Departamento de Lengua Española
Facultad de Arts - Universidad de El Cairo

Resumen: El presente estudio ofrece una herramienta para la detección automática de la modalidad deóntica (obligación, permiso y prohibición) en los textos legislativos en español y árabe. La herramienta está desarrollada en el lenguaje de programación de Python y utiliza la librería de Procesamiento del Lenguaje Natural (PLN) de Stanza. A través de las fuentes oficiales de legislación, se ha recopilado un corpus comparable de textos legislativos en español y árabe. Se trata de 5 textos legislativos españoles y sus correspondientes egipcios incluyendo las leyes de tráfico, propiedad intelectual, protección de datos personales, el código civil y el estatuto de trabajadores. A través de esta herramienta, se realiza un análisis contrastivo y cuantitativo de la modalidad deóntica en ambas lenguas. Los resultados revelan que la obligación es el tipo de modalidad deóntica más frecuente en los textos legislativos en ambas lenguas. En el corpus árabe se observa una distribución más equilibrada entre los tres tipos deónticos que en el corpus español donde la obligación es el tipo de modalidad deóntica por excelencia. Por otro lado, se observa que el verbo es la categoría gramatical más empleada para expresar la modalidad deóntica tanto en español como en árabe.

Palabras Clave: Modalidad deóntica, Lenguaje legislativo, Lingüística contrastiva, Lingüística computacional, Lingüística de corpus.

^(*) Bulletin of the Faculty of Arts Volume 84 Issue 4 April 2024

Abstract: This study presents a computational tool for the automatic detection and analysis of deontic modality (obligation, permission and prohibition) in legislative texts in Arabic and Spanish. The tool is developed in Python and it uses Stanza for text processing in both languages. A comparable corpus of legislative texts has been collected from official data sources. It consists of 5 Spanish legislative texts and their corresponding Egyptian legislative texts including the laws for Traffic, Intellectual Property, Personal Data Protection, Civil Code and the Statute of Labourers. The analysis uncovers the dynamics of how each language expresses the deontic modality within a contrastive framework. It also provides insights into the main linguistic features showing how legislative texts express obligation, prohibition or permission from a qualitative and quantitative perspective. Results show that obligation is the most frequent type of deontic modality in legislative texts in both languages. However, in the Arabic corpus, the distribution among the three types of deontic modality is more balanced than in Spanish, where the obligation is the dominant type of modality. On the other hand, the 'verb' is the Part-of-Speech mostly used to express modality in both languages.

Key Words: Deontic Modality, Legislative language, Contrastive Linguistics, Computational Linguistics, Corpus Linguistics.

صيغة الوجوب في النصوص التشريعية باللغتين العربية والإسبانية: أداة
حاسوبية للكشف الآلي والتحليل التقابلي

الملخص

يقدم هذا البحث أداة حاسوبية للكشف الآلي وتحليل صيغة الوجوب والتي تشمل الإلزام والسماح والمنع في النصوص التشريعية باللغتين العربية والإسبانية. وقد تم تطوير الأداة الحاسوبية باستخدام لغة Python ومجموعة Stanza لمعالجة النصوص في كلتا اللغتين. ولتطوير هذه الأداة الحاسوبية تم تجميع ذخيرة لغوية مقارنة (corpus) من النصوص التشريعية المأخوذة عن المصادر الرسمية. تتكون الذخيرة اللغوية من 5 نصوص تشريعية إسبانية والنصوص المقابلة لها من التشريعات المصرية بما في ذلك قوانين المرور وحقوق الملكية الفكرية وحماية البيانات الشخصية والقانون المدني ونظام

العمال.

يكشف البحث عن ديناميكية كل لغة فيما يختص بكيفية التعبير عن صيغة الوجوب والسمات اللغوية المميزة من خلال إطار تقابلي يعتمد على التحليل الكيفي والكمي .

وتظهر النتائج أن صيغة الوجوب الإلزامية هي النوع الأكثر شيوعاً في النصوص التشريعية في اللغتين، إلا أن توزيع أنواع صيغ الوجوب في اللغة العربية تتميز بتوازن أكثر من توزيع هذه الصيغ في النصوص التشريعية الإسبانية. من ناحية أخرى، تعتمد اللغتان على التكوينات التي تستخدم الفعل للتعبير عن صيغة الوجوب.

1. Introducción

Lengua y ley: instrumentos de comunicación

Lengua y Ley coinciden en su carácter instrumental y función comunicativa. La Lengua es un medio de comunicación general entre los seres humanos, mientras que la Ley es el medio institucional para comunicar las normas en una sociedad. De ahí, los textos legislativos representan una función específica en el marco de la comunicación lingüística general.

Por esta razón, el lenguaje siempre ha desempeñado un papel fundamental en la formalización, la comunicación, la interpretación y la aplicación de las leyes. Desde las reglas de Maat en el Antiguo Egipto y el código de Hammurabi en Mesopotamia hasta nuestros tiempos, las leyes se han basado en el lenguaje para expresar las normas que regulan la sociedad.

En el Antiguo Egipto, aunque no se conoce un código legislativo explícito, los historiadores apuntan a las reglas de Maat (la diosa egipcia de la justicia) como la fuente de normas que regulaban las relaciones en la sociedad. El egipcio antiguo ha elegido expresar estas reglas empleando la negación y los verbos conjugados en la primera persona del singular en un pasado hipotético, donde el alma después de la muerte y en el Juicio de la Otra Vida confiesa lo que “no ha hecho” en su vida terrenal: ‘No he robado’, ‘No he blasfemado’, etc. (Loro-Andrés, 2017). Desde un punto de vista lingüístico-

discursivo, las reglas de Maat reflejan que la fuente de la legislación en el Egipto Antiguo reside en el individuo, quien actúa como emisor ante los dioses acerca de sus hechos en la vida terrenal. Este acto comunicativo se realiza mediante oraciones declarativas negativas con una modalidad epistémica que refleja la certeza sobre los hechos (como explicaremos más adelante).

Por otro lado, el código de Hammurabi emplea otro marco discursivo que se acerca a los códigos actuales. Probablemente por este motivo, se considera como el primer código legislativo en la historia humana. En el código de Hammurabi, se emplea el uso de estructuras condicionales reales donde la oración subordinada de la condición (prótasis) se expresa con verbos conjugados en el pasado o el presente y la oración principal (apódosis) expresa la consecuencia en el futuro [Ej. “Si un hombre ha dado a otro dinero en sociedad, partirán delante de los dioses a partes iguales los beneficios y las pérdidas en que incurrieran”.] (Franco, 1962:342). Desde el punto de vista discursivo, se observa una diferencia respecto a las reglas de Maat. Se trata de un acto comunicativo diferente donde el emisor es el Rey Hammurabi, el receptor es el ciudadano y la enunciación se realiza mediante las oraciones condicionales reales con carácter informativo y un grado de certeza que apunta a una modalidad epistémica.

Hoy en día, la legislación como acto comunicativo incluye a los órganos del poder judicial como emisores, el texto legislativo como enunciado y, finalmente, los órganos del poder ejecutivo y el poder judicial junto al ciudadano representan los posibles receptores.

En este proceso, los órganos legislativos competentes se basan en una consolidada tradición científica y metodológica que regula la redacción de los textos legislativos. La dimensión lingüística desempeña un papel importante en esta metodología donde se explicitan pautas lingüísticas y buenas prácticas para garantizar la calidad y la claridad del texto legislativo. Estas pautas se manifiestan de forma directa e indirecta en distintos mecanismos lingüísticos y metalingüísticos. Es en estos mecanismos metalingüísticos donde reside la modalidad de enunciado en su vertiente deóntica, objeto de

este estudio.

El presente estudio adopta un enfoque descriptivo y contrastivo para el análisis de las marcas de la deonticidad por su relevancia en los textos legislativos. Se aplican técnicas de Lingüística de Corpus y Lingüística Computacional mediante el desarrollo de una herramienta automática cuyo objetivo es identificar y contrastar las marcas de deonticidad (verbos, perífrasis verbales y otras estructuras sintácticas) tanto en español como en árabe en un corpus comparable de legislaciones españolas y egipcias.

La estructura del trabajo consiste en 5 secciones. La Sección 1 ofrece un análisis del estado de la cuestión para situar el trabajo dentro del panorama actual abordando diferentes aspectos como las pautas para redacción legislativa y la interdisciplinariedad. La Sección 2 presenta el objetivo y el alcance del trabajo. La metodología se describe en la Sección 3 incluyendo la descripción del proceso de recopilación del corpus y su pre-procesamiento. En la Sección 4 se detallan las fases del análisis en ambas lenguas. Los resultados se analizan en la Sección 5. Por último, se destacan las conclusiones y las futuras líneas de trabajo.

2. Marco teórico y estado de la cuestión

2.1. El texto legislativo y la modalidad deóntica

El texto legislativo tiene como finalidad comunicativa comunicar las normas que rigen las relaciones entre los órganos y los individuos en una sociedad. Para ello, la redacción legislativa adopta mecanismos lingüísticos y metalingüísticos que definen los rasgos de este lenguaje considerado como un lenguaje de especialidad. Esta aproximación coincide con la definición de Parodi (2005) del lenguaje de especialidad como “un conjunto de rasgos lingüísticos caracterizadores que co-ocurren sistemáticamente en torno a temáticas concretas y que, a su vez, encierra una coincidencia funcional de los mismos”.

Los mecanismos lingüísticos se manifiestan en los diferentes niveles de análisis como el nivel léxico semántico con el empleo de términos especializados; en el nivel morfológico mediante tiempos

verbales y formas impersonales; en el nivel sintáctico mediante estructuras sintácticas o perífrasis verbales características del ámbito legal, etc.

En el nivel metalingüístico, nos encontramos ante la modalidad como otro mecanismo relevante en el lenguaje. Basada en la Lógica y la Filosofía, la modalidad se asocia con la subjetividad y la expresividad en el lenguaje. Se distingue entre la modalidad de la enunciación como actitud del emisor ante su interlocutor (el *modus* o la forma de transmitir el enunciado) y la modalidad del enunciado como actitud del emisor ante lo enunciado (el *dictum* o el contenido) (Matte Bon, 1995; Bybee y Fleischman, 1995; Centro Virtual Cervantes, 2023).

La modalidad de la enunciación refleja la forma de comunicación entre el emisor y su interlocutor y se manifiesta en la modalidad de la oración: interrogativa, declarativa, imperativa, etc. (Centro Virtual Cervantes, 2023).

La modalidad del enunciado refleja las modalidades lógicas: la epistémica, la deóntica y la axiológica/apreciativa. La modalidad epistémica indica la actitud del emisor en cuanto a la certeza/verosimilitud del enunciado [Ej. Seguramente la convocatoria se publicará este mes.]. La modalidad deóntica indica la actitud del emisor acerca de la obligatoriedad del enunciado [Ej. La convocatoria se debería publicar este mes.]. Por último, la modalidad axiológica representa cómo el emisor aprecia el enunciado desde su escala de valores, por eso también recibe el nombre de modalidad apreciativa [Ej. Me parece bien que se publique la convocatoria este mes.] (Centro Virtual Cervantes, 2023).

Los estudios consultados han abordado la modalidad deóntica desde distintos enfoques, i.e. comparativo, pragmático-discursivo, traductológico, etc.

Desde el enfoque contrastivo, se destaca el estudio de la modalidad deóntica entre la legislación nacional española y la legislación europea en español (Polo, 2013). Desde la perspectiva pragmático-discursiva, algunos estudios se centran en la modalidad deóntica en una ley específica como la ley fiscal (Burgoa-Toledo,

2010), en un tipo de documento legal como la Constitución (Morillo-Velarde Pérez, 2011) o en el destinatario (Wiltos, 2020).

Otro enfoque relevante es el traductológico como los estudios de El-Farhaty y Elewa (2020) o el estudio de Abdel-Fattah (2005) sobre la traducción de la modalidad deóntica entre el árabe y el inglés.

2.2. La redacción legislativa

En Europa, en general, y en España, en particular, existen normativas que regulan la redacción legislativa como la resolución del 28 de junio de 2005 que define los criterios lingüísticos para la redacción de textos legislativos (BOE, 2005). Asimismo, existen iniciativas para la modernización del discurso jurídico en español de cara a la claridad, la transparencia y la accesibilidad de los textos jurídicos (Montolío Durán, 2012; Da Cunha y Escobar, 2021).

En el mundo árabe, en general, y en Egipto, en particular, existen directrices lingüísticas y guías de buenas prácticas para la redacción legislativa en árabe, pero estas directrices carecen de carácter normativo (Al-Leheiby, 2019). Por este motivo, la Organización Mundial de Cooperación y Desarrollo Económico (OCDE) ha elaborado un manual sobre las buenas prácticas en la redacción legislativa para la zona del Medio Oriente y, en particular para Egipto (OCDE, 2019).

En estas normas o recomendaciones tanto en árabe como en español se destaca una serie de aspectos lingüísticos a tener en cuenta en la redacción legislativa como emplear un léxico unívoco, evitar el uso de las formas pasivas y negativas para garantizar una mayor claridad, optar por oraciones breves para evitar ambigüedad, etc.

2.3. El texto legislativo: Un área de estudio interdisciplinaria

Según la afirmación de Jiménez-Salcedo y Moreno-Rivero (2017), la relación entre Lengua y Ley se trata de una relación intrínseca que ha creado un espacio interdisciplinario abarcando diferentes perspectivas lingüísticas como la enseñanza, el estudio del

lenguaje con fines específicos, la traducción, la terminología, la estilística y la pragmática (Ballesteros & Samaniego, 2014; Chierichetti & Garofalo, 2011).

Este espacio interdisciplinario ha recibido diferentes denominaciones como ‘Lingüística legal’ (*Legal Linguistics*) o ‘Lingüística jurídica’. Ejemplo de ello, el libro editado por Vogel (2019) donde se describe el estado de la cuestión en diferentes áreas geográficas incluyendo África y España. Por otro lado, y desde un enfoque traductológico, se ha acuñado el término ‘Jurilingüística’ donde también se observa el interés por el uso de corpus en los estudios del lenguaje jurídico (Jiménez-Salcedo & Moreno-Rivero, 2017; Moreno-Rivero & Jiménez-Salcedo, 2020; Pontrandolfo, 2012).

El español jurídico

Centrándonos en el marco del español jurídico, destacamos algunos estudios de referencia que han abordado los textos jurídicos y sus rasgos lingüísticos desde varios enfoques: el léxico, el sintáctico, el terminológico, el discursivo y el pragmático. El libro ‘Lenguas, lenguaje y derecho’ de Jesús José Prieto de Pedro (1991) ofrece un análisis exhaustivo de las características del lenguaje jurídico en español subrayando algunos rasgos como el uso de latinismos y neologismos, el empleo de tecnicismo, la tendencia a la reiteración y el uso de los verbos en forma impersonal. Calvo Ramos (1980) destaca los latinismos y cultismos, la afijación, las formas nominales y la estructura de la oración. Por otra parte, Alcaraz Varó y Hughes (2002) analizan el lenguaje jurídico desde la lingüística oracional incluyendo fenómenos morfológicos, sintácticos, léxicos, terminológicos. Mientras que, otros como Greenberg (2011) o Medina-Montero (2013), se centran en el nivel de la lingüística del texto destacando rasgos relativos al discurso, la función comunicativa, la estilística, la pragmática y la macroestructura del texto.

El árabe jurídico

En árabe, son menos los estudios que abordan las características del lenguaje jurídico. En su mayoría coinciden con las observaciones sobre el lenguaje jurídico en otras lenguas como los tecnicismos, la ambigüedad, algunos usos arcaicos, el uso de los

verbos de forma impersonal. Sin embargo, señalan como un rasgo distintivo el uso de términos de origen religioso, sobre todo en algunos casos como las leyes de familia (Al-Leheiby, 2019).

Estudios lingüísticos de textos legales en español y árabe. Siendo el español y el árabe el objeto del presente estudio, el análisis del estado del arte revela que la traducción es la línea de investigación por excelencia que ha abordado los textos legales en ambas lenguas (El Krirh, 2017; Soto-Aranda & Kabbour, 2020). Algunos estudios ponen mayor énfasis en la terminología (Gogazeh, 2014) o en la enseñanza, pero siempre desde la perspectiva de la traducción (Saad, 2015; Bouteffah Aghbalou, 2017).

La Lingüística legal y Tecnología

En el espacio interdisciplinario generado del encuentro entre la lengua y el derecho caben más disciplinas y enfoques como la Informática, el Procesamiento del Lenguaje Natural (PLN) y la Inteligencia Artificial (*Legal Informatics*) (Hinson, 2005; Katz et al., 2021). Esta última línea de desarrollo ofrece soluciones prácticas a los desafíos encontrados al abordar los grandes volúmenes de textos legales y administrativos. Estas soluciones permiten búsquedas inteligentes o extracción de información relevante en millones de textos y van destinados a un amplio abanico de usuario con diferentes necesidades y finalidades desde investigadores, abogados, jueces, legisladores, órganos responsables de aplicar la ley, documentalistas, administraciones públicas y ciudadanos. Los ciudadanos como grupo destinatario está siendo objeto de interés cada vez más en los últimos años teniendo en cuenta la necesidad de una comunicación clara entre las administraciones públicas y la ciudadanía (Da Cunha, 2020).

El presente estudio se enmarca en este último espacio interdisciplinar. De ahí, el presente trabajo contribuye a este panorama aportando un enfoque contrastivo y computacional mediante una herramienta de análisis automático que permite una aproximación tanto cualitativa como cuantitativa.

3. Objetivo y alcance

3.1. Objetivo general y objetivos específicos

El presente estudio tiene como objetivo ofrecer una herramienta práctica que permite la detección y el análisis de la modalidad deóntica en los textos legislativos en español y árabe. Esta detección se presenta dentro de un marco previo de análisis contrastivo para comprender los rasgos de los textos legislativos en español y en árabe.

Para conseguir este objetivo general, se detallan los objetivos específicos a continuación:

- Recopilar un corpus comparable de textos legislativos españoles y egipcios
- Analizar las características del corpus en ambas lenguas para su procesamiento
- Desarrollar la herramienta automática para la detección de marcas de deonticidad
- Analizar los resultados de detección de deonticidad

3.2. Planteamiento

Para el estudio, se ha optado por un corpus comparable formado de textos originales del mismo género en ambas lenguas y con misma funcionalidad comunicativa, i.e. describir las normas legislativas en los ámbito de

La solución propuesta aplica técnicas de Procesamiento del Lenguaje Natural (PLN) y análisis basado en un corpus de muestra permitiendo, de esta manera, tratar grandes cantidades de textos y extraer la información relevante de forma rápida y eficiente. Además, permite formular conclusiones basadas en un estudio empírico que parte de los datos.

Se trata de una herramienta desarrollada específicamente para esta finalidad, teniendo en cuenta la naturaleza lingüística de las dos lenguas en cuestión y la modalidad lingüística objeto de estudio.

El desarrollo se ha llevado a cabo en el lenguaje de Python y se ha empleado la librería abierta de *Stanza* (Qi et al., 2020) para el procesamiento morfosintáctico, dado que ofrece recursos y módulos de análisis lingüístico para las dos lenguas, objeto del estudio.

La herramienta recibe como entrada una colección de textos en español o en árabe y tras un análisis morfosintáctico automático, se extrae la información relevante a la modalidad deóntica de permiso, obligación y prohibición en las dos lenguas (español y árabe). Los textos de salida consisten en las expresiones deónticas extraídas en su contexto junto a un análisis cuantitativo que permite valorar la distribución de los diferentes tipos de modalidad deóntica y sus formas gramaticales.

Optar por un desarrollo específico en lugar de utilizar herramientas genéricas de análisis textual viene motivado por una serie de razones. Primero, estas herramientas ofrecen soluciones genéricas, difíciles de adaptar a un problema concreto o un fenómeno lingüístico concreto. Segundo, son pocas las herramientas que cuentan con módulos para el tratamiento del árabe. Tercero, la mayoría de estas herramientas no están abiertas de uso gratis y suelen requerir una suscripción de pago. El código de la herramienta desarrollada para este estudio se hará disponible con acceso abierto a la comunidad científica a través del repositorio de *github*¹.

3.3. Preguntas de la investigación

El presente trabajo plantea una serie de cuestiones de carácter analítico y metodológico. A continuación se detallan estas preguntas:

- ¿Cómo se define la modalidad deóntica? ¿Cómo se manifiesta esta modalidad en los textos legislativos árabes y españoles?

- ¿Cómo se puede emplear las técnicas de análisis textual y Procesamiento del Lenguaje Natural para detectar la modalidad deóntica en los textos legislativos en ambas lenguas?

3.4. Alcance

El estudio se basa en un corpus comparable recopilado desde fuentes oficiales en España y Egipto. Se trata de una muestra de textos legislativos a diferencia de otros tipos de textos legales (Ej. textos jurisprudenciales, textos académicos o textos judiciales).

Al tratarse de un corpus comparable, los textos que lo forman son textos originales en ambas lenguas. Representan el mismo género textual y la misma funcionalidad comunicativa, i.e. describir las normas legislativas en España y en Egipto. El corpus consiste en 5 textos legislativos españoles y 5 legislaciones egipcias equivalentes incluyendo las leyes de tráfico, propiedad intelectual, protección de datos personales, la ley civil y el estatuto de trabajadores.

El fenómeno lingüístico, objeto de análisis es la modalidad deóntica (permiso, obligación y prohibición) en la variante del español peninsular y la variante egipcia del árabe.

4. Metodología

El estudio reúne unos rasgos principales que se detallan a continuación:

Es un estudio aplicado que desarrolla una herramienta práctica que detecta la modalidad deóntica en los textos legislativos en español y árabe. Por tanto, es una aproximación empírica basada en corpus con evidencias.

Es un estudio computacional que emplea técnicas computacionales del Procesamiento del Lenguaje Natural (PLN) haciendo uso de librerías de acceso abierto como *Stanza* (Qi et al., 2020). Estas técnicas se emplean por primera vez en un estudio de esta índole.

La herramienta desarrollada permite un análisis descriptivo tanto cuantitativo como cualitativo sobre el uso actual de la modalidad deóntica en la legislación española y egipcia. Eso permite, a su vez, un análisis contrastivo de esta modalidad en las dos lenguas.

Es un trabajo con un fin práctico que puede ser un punto de partida para otras soluciones informáticas destinadas tanto a investigadores de Lingüística o del Derecho como a profesionales del dominio legal para una detección rápida y eficiente de las

obligaciones, los permisos o las prohibiciones en los textos legislativos.

La metodología consiste en las siguientes fases:

- a) Identificación de las fuentes y recopilación de los datos.
- b) Depuración y pre-procesamiento de los datos
- c) Cálculo de indicadores léxico-estadísticos del corpus comparable
- d) Identificación de la modalidad deóntica
- e) Análisis de la modalidad deóntica detectada.

5. Análisis

5.1. Identificación de las fuentes y recopilación de los datos

Se ha recurrido a las fuentes oficiales que ofrecen la legislación nacional española y egipcia en sus portales electrónicos. Para la legislación española, se han recopilado los datos del portal del Boletín Oficial del Estado (BOE)² donde se publica toda la legislación como datos abiertos en diferentes formatos (pdf, txt y xml). Para este estudio, se ha optado por el formato de XML, a partir del cual se han extraído los datos y el contenido textual sin los metadatos.

Es importante señalar que los textos legislativos en España se encuentran disponibles como ‘Datos abiertos’ gracias a la Ley de Reutilización de la Información del Sector Público (RISP) basada en la Directiva 2003/98/CE, de 17 de noviembre de 2003, del Parlamento Europeo y del Consejo Europeo. La apertura de los datos del sector público tiene como objetivo fomentar la transparencia administrativa y el potencial económico del dato en la era digital.

No obstante, este no es el caso para los textos legislativos en Egipto, donde no existe una ley que regule la apertura de los datos del sector público pese a los grandes esfuerzos realizados en los últimos años en el ámbito de la transparencia democrática, la transformación digital y los servicios del gobierno electrónico. Por eso, la recopilación de los datos de la legislación egipcia se ha enfrentado con dos retos:

1. La disponibilidad y el acceso a los datos: En principio, se ha recurrido al portal Boletín Oficial egipcio, pero los datos

completos no están abiertos y es necesaria una suscripción con pago para acceder al texto completo de las leyes. Ante este reto, se ha optado por fuentes alternativas en otros portales oficiales que permitan el acceso libre a subconjuntos de la legislación como la Corte de Casación³ o el portal oficial del Colegio de Abogados Egipcios⁴. Sin embargo, estas fuentes no ofrecen toda la legislación ni permiten realizar búsquedas avanzadas.

2. El formato de los datos y las dificultades del procesamiento: Superado el primer reto, se ha observado que en algunos casos los datos accesibles no están en formato texto o XML. En cambio, están en formato PDF o imagen. Para poder analizar estos documentos, es necesario realizar un proceso previo de conversión a formato de texto utilizando herramientas de reconocimiento de caracteres. Esta solución supone unas limitaciones porque el proceso de conversión tiene un margen de error, que suele introducir ‘ruido’ de cara al procesamiento posterior debido a los caracteres no reconocidos y los desajustes de formato y, por tanto, requiere de mayor esfuerzo de depuración.

Dadas estas limitaciones, hemos seguido un número de criterios para la construcción del corpus comparable objeto del presente estudio como se detalla a continuación:

- Optar por las versiones actualizadas y vigentes de las leyes recopiladas.
- Seleccionar las leyes que cuentan con equivalentes en ambas lenguas para poder realizar un análisis contrastivo.
- Priorizar las leyes que están disponibles en formato de html o texto en el caso de la legislación egipcia para garantizar la calidad de los datos.
- Priorizar la calidad del dato para garantizar resultados fiables que se pueden validar.

5.2. Depuración y procesamiento del corpus

Después de recopilar los datos, se realiza un proceso de depuración básico para eliminar posibles causas de ‘ruido’ en los

textos a tratar como, por ejemplo, los espacios seguidos, tabulares, saltos de línea innecesarios o caracteres malcodificados.

Siguiendo el criterio de priorizar textos en formato XML o HTML, este proceso no ha supuesto mucho esfuerzo. No obstante, siempre se recomienda realizar una depuración básica para garantizar el buen funcionamiento de los procesos posteriores como la tokenización.

Finalizada la fase de depuración, se procede al procesamiento que consiste en estos análisis básicos realizados con la librería *Stanza*:

- a) El proceso de *tokenización* para identificar las diferentes palabras y separar las secuencias de caracteres alfabéticas consideradas palabras y los signos de puntuación o los caracteres numéricos. Este análisis nos permite analizar posteriormente la distribución de *tokens* y tipos.
- b) Los *tokens* se refieren a las ocurrencias repetidas de las palabras, mientras que los tipos se refieren a las clases únicas de estas ocurrencias. Por ejemplo, la palabra ‘obligación’ puede ocurrir 20 veces en un texto considerados como 20 *tokens* de un solo tipo ‘obligación’.
- c) La segmentación de las oraciones.
- d) La anotación morfosintáctica (*POS: Part of Speech*) en este proceso a cada palabra se le asigna la categoría gramatical correspondiente para identificar nombres, verbos, adjetivos, preposiciones, etc.
- e) La lematización donde a cada palabra se le proporciona el lema correspondiente.

El análisis morfosintáctico nos permite distinguir entre las unidades léxicas funcionales conocidas como ‘*stop-words*’ (determinantes, preposiciones, etc.) y las unidades léxicas nocionales, dotadas de contenido semántico como los sustantivos, los verbos, los adjetivos, etc. Esta distinción nos sirve para calcular las estadísticas léxicas, además de analizar cómo se manifiesta la modalidad deóntica en cada lengua y qué categorías gramaticales emplea.

La lematización, por otra parte, nos permite una identificación con mayor cobertura los patrones de la modalidad deóntica,

especialmente en lenguas altamente flexivas (flexión nominal y verbal) como el español o el árabe. A través de los lemas, se definen reglas genéricas capaces de detectar las diferentes formas en las que ocurre un verbo, un sustantivo o un adjetivo. Por ejemplo, con definir una regla que localiza el lema de los verbos ‘deber’ o ‘permitir’, se pueden identificar todas sus formas de flexión que aparecen en el texto. De la misma forma, se identifican adjetivos como ‘prohibido’, ‘permitido’ en plural o en femenino o los nombres en plural de ‘prohibición’ o ‘obligación’.

El resultado de esta fase de procesamiento consiste en una relación de los *tokens* que forma el texto y a cada *token* se le asigna la categoría gramatical y el lema correspondiente. Como dato adicional, se añade la información contextual sobre las unidades que ocurren antes y después, es decir de una ventana de análisis de +1/-1. Las Tablas 1 y 2 demuestran el resultado de la fase de procesamiento.

Tabla 1. Muestra de los resultados de procesamiento de textos en español.

	Oración	Token	Token_Antes	Token_Después	Token_POS	Token_Lema
0	la disposición final segunda de la ley 27/1...			.	el	ADP
1	la disposición final segunda de la ley 27/1...	la		disposición	DET	el
2	la disposición final segunda de la ley 27/1...	disposición	el	final	NOUN	disposición
3	la disposición final segunda de la ley 27/1...	final	disposición	segundo	ADJ	final
4	la disposición final segunda de la ley 27/1...	segunda	final	de	ADJ	segundo
...	...	...	...	...	...	...
95	el alcance temporal de esta habilitación legis...	legales	disposición	que	ADJ	legal
96	el alcance temporal de esta habilitación legis...	que	legal	él	PRON	que
97	el alcance temporal de esta habilitación legis...	se	que	encontrar	PRON	él

Tabla 2. Muestra de los resultados de procesamiento de textos en árabe.

	Oración	Token	Token_Antes	Token_POS	Token_Lema
0	٢٠٠٢ لسنة ٨٢ لقانون النشر	قانون	٢٠٢٠	NOUN	قانون
1	٢٠٠٢ لسنة ٨٢ لقانون رقم ٨٢ لسنة ٢٠٠٢	رقم	قانون	NOUN	رقم
2	٢٠٠٢ لسنة ٨٢ لقانون النشر	لسنة	٨٢	NOUN	لِسنة
3	٢٠٠٢ لسنة ٨٢ لقانون رقم ٨٢ لسنة ٢٠٠٢	تاريخ	٢٠٠٢	NOUN	تاريخ
4	٢٠٠٢ لسنة ٨٢ لقانون رقم ٨٢ لسنة ٢٠٠٢	نوع	٢٠٠٢	NOUN	نوع
...	...	...	...	...	...
66579	صدر برئاسة الجمهورية في ٢٧ ذي الحجة سنة ١٤١٤ هـ	الجريدة	ب	NOUN	جَرِيْدَة
66580	صدر برئاسة الجمهورية في ٢٧ ذي الحجة سنة ١٤١٤ هـ	الرسمية	جَرِيْدَة	ADJ	رَسْمِيّ
66581	صدر برئاسة الجمهورية في ٢٧ ذي الحجة سنة ١٤١٤ هـ	العدد	رَسْمِيّ	NOUN	عَدَد
66582	صدر برئاسة الجمهورية في ٢٧ ذي الحجة سنة ١٤١٤ هـ	في	)	ADP	في

5.3. Indicadores léxico-estadísticos

Al adoptar un enfoque contrastivo y partiendo de en un corpus comparable, se ha estimado imprescindible un análisis previo para contrastar los rasgos de los textos legislativos en español y árabe a través del cálculo de algunos indicadores léxico-estadísticos. Este análisis ha sido motivado por las diferencias significativas detectadas en los resultados de los recuentos básicos de los textos que forman el corpus.

Según López-Morales (2010), es importante distinguir entre la extensión de un texto y la extensión del vocabulario. El número de las palabras que componen el texto, representado por N , representa la extensión del texto. Por otro lado, el número total de los vocablos o tipos léxicos distintos empleados en el texto, representado por V , refleja la extensión del vocabulario.

En términos de Procesamiento del Lenguaje Natural, el valor de V es el equivalente al recuento de los ‘tipos’ de estos *tokens*, mientras que el valor de N es el equivalente al recuento de los *tokens* léxicos, entendidos como *tokens* de palabras descartando los *tokens* de signos de puntuación u otros símbolos.

A su vez, las unidades léxicas se dividen en unidades nocionales y unidades funcionales. Las unidades nocionales son aquellas con contenido semántico como los sustantivos, los verbos, los adjetivos, etc. En cambio, las unidades funcionales se refieren a los determinantes, las preposiciones, etc. que se conocen en la literatura del Procesamiento del Lenguaje Natural como *stop-words*.

La Tabla 3 muestra los recuentos del corpus comparable distinguiendo entre los ‘tokens’, los ‘tipos’ y entre *tokens* y tipos nocionales y ‘tokens’ y tipos funcionales.

Tabla 3. Datos cuantitativos del corpus comparable.

Corpus árabe				Corpus español			
Total <i>tokens</i> (<i>vocablos+signos</i>)	74 632	Total tipos	6857	Total <i>tokens</i> (<i>vocablos+signos</i>)	317 298	Total tipos	15 154
Tokens (palabras)	66 584	Tipos (palabras)	6346	Tokens (palabras)	287 924	Tipos (palabras)	13065
<i>Tokens</i> funcionales (stop-words)	26 274	Tipos funcionales (stop-words)	239	<i>Tokens</i> funcionales (stop-words)	142 615	Tipos funcionales (stop-words)	228
<i>Tokens</i> nocionales	40 310	Tipos nocionales	6139	<i>Tokens</i> nocionales	145 309	Tipos nocionales	12893

Basándose en esta distinción, se han seleccionado algunos indicadores de referencia como la diversidad léxica, el intervalo de aparición de palabras y la densidad léxica (López-Morales (2010). Además de otros indicadores como la media de repetición de una palabra. A continuación, se detallan las métricas aplicadas:

- a) La diversidad léxica, ‘riqueza léxica’ (LV) o *Type-Token Ratio* (TTR): $LV = V / N$
- b) El porcentaje de vocablos (PV): $PV = V \times 100 / N$
- c) El intervalo de aparición de palabras de contenido nocional (IAT). Este considera la cantidad total de palabras del texto y la divide por el conjunto formado por las palabras nocionales

(PN) (López-Morales, 2010): $IAT = N / PN$.

d) La densidad léxica (DL). El índice de densidad léxica suele ser de una media de 40 %, es decir más del 40 % de un texto corresponde a unidades nocionales. La densidad léxica se calcula de la siguiente forma: $DL = PN / N$ (Fuentes-Riffo et al., 2019).

e) Adicionalmente, se ha añadido otro índice al que nos referimos como:

Media de la repetición léxica en un texto (RL): $RL = N / PN$.

Tabla 4. Comparativa de indicadores léxico-estadísticos.

	Corpus ar	Corpus es
Diversidad léxica (LV)	0.1	0.04
Porcentaje de vocablos (PV)	9.5	4.5
Intervalo aparición vocablos nocionales (IAT)	1.6	1.9
Densidad Léxica (DL)	0.6	0.5
Repetición del léxico nocional (RL)	6.5	11.2

Según este análisis, la extensión de los textos legislativos en español es significativamente mayor que la extensión de los textos en árabe. Es necesario recordar que no se trata de traducciones, sino textos comparables. Existen dos posibles razones detrás esta diferencia. La primera es de naturaleza legislativa, ya que es posible que la legislación española abarque más detalles con un mayor desglose de artículos y condiciones. La segunda razón es de naturaleza lingüística, dado que los textos en árabe, en general, suelen ser más cortos y sintéticos por el carácter altamente flexivo y por el empleo frecuente de los clíticos que permite, en algunos casos, que una oración se pueda sintetizar en solo *token*. Por ejemplo, una oración puede estar formada de un solo *token* en árabe como ‘وسيقرره’ (y se decidirá), mientras que, en un texto en español, este *token* equivaldrá a 3 *tokens*.

A pesar de que los textos en español son más extensos, se observa que los textos árabes presentan una mayor diversidad y densidad léxica.

La probabilidad de que se repita un léxico nocional en español (11.2) es significativamente mayor que en árabe (6.5). Por tanto, el intervalo de aparición de un vocablo nocional en árabe (1.6) es menor que en español (1.9).

Teniendo en cuenta las categorías gramaticales, la Tabla 5 demuestra la distribución de las diferentes categorías gramaticales en el corpus español y árabe según el análisis proporcionado por *Stanza*⁵.

Se observa que los sustantivos constituyen la categoría más frecuente en ambas lenguas. No obstante, en árabe la frecuencia de sustantivos es significativamente mayor representando el 40 % del total de palabras. Las frecuencias de otras categorías como los verbos, las preposiciones y los adjetivos son parecidas en ambos corpus. Por otro lado, cabe destacar que el uso de los adverbios es muy poco frecuente en el corpus árabe.

Tabla 5. Distribución de las categorías gramaticales en el corpus comparable.

Corpus árabe			Corpus español		
Categoría gramatical (POS)	Frec. ⁶	Frec. relativa	Categoría gramatical (POS)	Frec.	Frec. relativa
ADJ	7471	0.11	ADJ	29 054	0.10
ADP	11 947	0.18	ADP (Preposición)	57 944	0.20
ADV	63	0,001	ADV	6838	0.02
AUX	424	0.01	AUX	8033	0.03
CCONJ	6585	0.01	CCONJ (Conjunción)	11 973	0.04
DET	2607	0.04	DET	51 340	0.18
			INTJ	18	0.00
NOUN	27 601	0.41	NOUN	80 197	0.27
PART	859	0.01	PART	95	0.00
PRON	3707	0.06	PRON	16 297	0.06
PROPN	5	0.00	PROPN (Nombres propios)	630	0.00
SCONJ	569	0.01	SCONJ	4966	0.02
VERB	4746	0.07	VERB	20 539	0.07

5.4. Identificación de la modalidad deóntica

Una vez analizado el corpus comparable con sus rasgos léxico-estadístico, se ha procedido al reconocimiento de tres tipos de la modalidad deóntica: Permisi6n/permiso, obligaci6n y prohibici6n.

El reconocimiento autom6tico de cada tipo de la modalidad de6ntica se ha basado en un conjunto de reglas definidas para cada lengua. Estas reglas se basan en la identificaci6n de los lemas que representan los tres tipos de6nticos indicados como, por ejemplo, ‘permitir’, ‘deber’, ‘prohibir’, ‘obligaci6n’, ‘prohibici6n’, etc.

Adicionalmente, se han enriquecido estas reglas con informaci6n contextual sobre las palabras que ocurren antes o despu6s del lema. Esta informaci6n contextual permite: 1) Identificar per6frasis verbales como ‘haber de’ o ‘tener que’ que indican obligaci6n en espa6ol. 2) Excluir falsos positivos como la expresi6n ‘deber a’, dado que no se trata de una obligaci6n. 3) Identificar la negaci6n de los verbos, dado que la negaci6n afecta el tipo de modalidad de6ntica. Por ejemplo, ‘no deber’ o ‘no es obligatorio’ no expresan obligaci6n a pesar de incluir el lema ‘deber’ o ‘obligatorio’, sino permiso. De la misma manera, ‘no permitir’ ‘يجوز لا’ indica una prohibici6n (i.e. no se permite).

- (1) Si (lema=’جآز’ [permitir] y palabra_antes=’لا’ [no]):
Entonces modalidad de6ntica: Prohibici6n

Las tendencias actuales en Tecnolog6as del Lenguaje emplean los modelos del lenguaje y t6cnicas de aprendizaje autom6tico para tareas de extracci6n de informaci6n (detecci6n y clasificaci6n). No obstante, en este estudio, se ha optado por el uso de las reglas. Esta decisi6n se basa en la eficiencia de las reglas en tratar el fen6meno de la modalidad de6ntica dado su alcance definido que se puede formalizar en un n6mero acotado de reglas. Adem6s, el uso de las reglas ofrece una mayor precisi6n en la detecci6n de la modalidad de6ntica, lo cual resulta m6s conveniente para estudios con fines lingüísticos o legales.

Cabe destacar que existen casos de ambigüedad como el uso del verbo ‘poder’ en espa6ol que puede expresar un permiso o una

prohibición en su forma negativa. Además, puede expresa una modalidad epistémica sobre posibilidad. Estos verbos de carácter ambiguo, se han descartado para evitar posibles falsos positivos procedentes de la modalidad epistémica. Sin embargo, es una línea de investigación que se puede abordar en el futuro con un enfoque centrado en técnicas de desambiguación empleando modelos de clasificación que tienen en cuenta el contexto semántico.

Apoyándose en las reglas, el proceso de detección se aplica en tres módulos; uno por cada tipo de modalidad deóntica. El resultado de la detección se muestra en una tabla donde se indica el caso detectado junto a su categoría gramatical y el contexto donde aparece (la oración y las palabras que aparecen antes y después). A modo de ejemplo, se incluyen dos muestras de los resultados de la detección en español y en árabe.

Tabla 6. Muestra de la modalidad deóntica de obligación detectada en español.

	Oración_Obligación	Obligación_Token	Token_Antes	Token_Después	Token_POS	Token_Lema
0	la disposición final segunda de la ley 27/1...	hubieran	que	de	AUX	haber
1	1.º decidir si su obra ha de ser divulgada y e...	ha	obra	de	AUX	haber
2	2.º determinar si tal divulgación ha de hacers...	ha	divulgación	de	AUX	haber
3	si, posteriormente, el autor decide reemprende...	deberá	obra	ofrecer	AUX	deber
4	esta cantidad no podrá incluir total o parcial...	deba	que	satisfacer	AUX	deber
...	...	...	...	...	...	...
1896	en las obligaciones mancomunadas, cuando el ac...	obligaciones	el	mancomunado	NOUN	obligación

Tabla 7. Muestra de la modalidad deóntica de prohibición detectada en árabe.

	Oración_Prohibición	Prohibición_Token	Token_Antes	Token_POS	Token_Lema
0	٢... - قيام الغير في جمهورية مصر العربية، بصنع من	يجوز	لَا	VERB	جَاز
1	... يقدم طلب البراءة من المخترع أو ممن المادة ١٢	يجوز	لَا	VERB	جَاز
2	... لا يجوز التنازل عن الاستخدام المرخص به لإحدى ا	يجوز	لَا	VERB	جَاز
3	...وفيما عدا طلبات الإلغاء المقترنة بطلب وقف التث	يجوز	لَا	VERB	جَاز
4	... لا يجوز للعاملين بمكتب براءات الاخت المادة ٣٩	يجوز	لَا	VERB	جَاز
...	...	...	...	...	...
132	...مادة ٥٠ - تكون بطلقة تحقيق الشخصية حجة على صحة	يجوز	لَا	VERB	جَاز
133	...ويجب على كل من تنطبق عليه أحكام المادة (٤٨) من	يجوز	لَا	VERB	جَاز

En estas muestras se observa que la detección por lema permite reconocer las posibles formas en las que puede aparecer un verbo como en el caso del verbo ‘deber’ en español que aparece como ‘debe’, ‘deberán’, ‘deban’ y ‘debiendo’.

Por otro lado, se detectan las expresiones de modalidad deóntica de prohibición que emplean la negación ‘no + el verbo permitir (يجوز لا)’ en forma impersonal el presente siendo esta forma la más empleada en la prohibición por los textos en árabe.

Por último, la herramienta consigue detectar expresiones de obligación basadas en perífrasis verbales en español como ‘haber de’ en la forma de ‘han de’.

6. Resultados y discusión

Procesando los textos del corpus comparable por los diferentes módulos de la herramienta, obtenemos los resultados de la detección y clasificación de la modalidad deóntica en ambas lenguas según demuestra a continuación la Tabla 8:

Tabla 8. Frecuencias de tipos de modalidad deóntica en el corpus comparable.

	Corpus árabe		Corpus español	
	frecuencia	%	frecuencia	%
Permiso	210	33.9 %	235	10.4 %
Obligación	271	43.8 %	1901	84.4 %
Prohibición	137	22.3 %	116	5.2 %
Total casos	618	100 %	2252	100 %

Frecuencia

Desde una aproximación cuantitativa, se observa que la frecuencia absoluta de casos de modalidad deóntica en español es mayor que en árabe (2269 casos frente a 618 casos en árabe). Sin embargo, una frecuencia normalizada que tiene en cuenta el volumen total de los textos legislativos en cada lengua (66 584 palabras en árabe y 287 924 palabras en español), demuestra que la probabilidad de que una palabra forme parte de una expresión deóntica es mayor en árabe (0.9 %) que en español (0.7 %). Es decir, la modalidad deóntica es más frecuente en los textos legislativos árabes que los textos legislativos españoles.

En cuanto al orden de frecuencia de los tipos de modalidad deóntica, se observa que es similar en ambas lenguas. La obligación es el tipo más frecuente, seguida de la modalidad deóntica de permiso y, por último, la prohibición. Esta distribución es una prueba de cómo el lenguaje es un fiel reflejo de la naturaleza y la propia función comunicativa de la legislación como instrumento de establecer las normativas mediante una serie de obligaciones.

Distribución

Los resultados de la detección y clasificación revelan una diferencia significativa en cuanto a la distribución de los tres tipos en cada lengua. En español, la obligación es dominante (84.4 %), seguida de una presencia limitada del permiso (10.4 %) y la prohibición (5 %). En cambio, en árabe, los tres tipos se distribuyen de forma más equilibrada donde la obligación representa el 43.8 %, los permisos; el

33.9 % y las prohibiciones; el 22.3 %. Una posible razón detrás de esta diferencia se puede encontrar en la filosofía del derecho donde hay sistemas que se apoyan más en la obligación que en la prohibición o puede tratarse de una cuestión puramente estilística, fruto de la tradición legislativa.

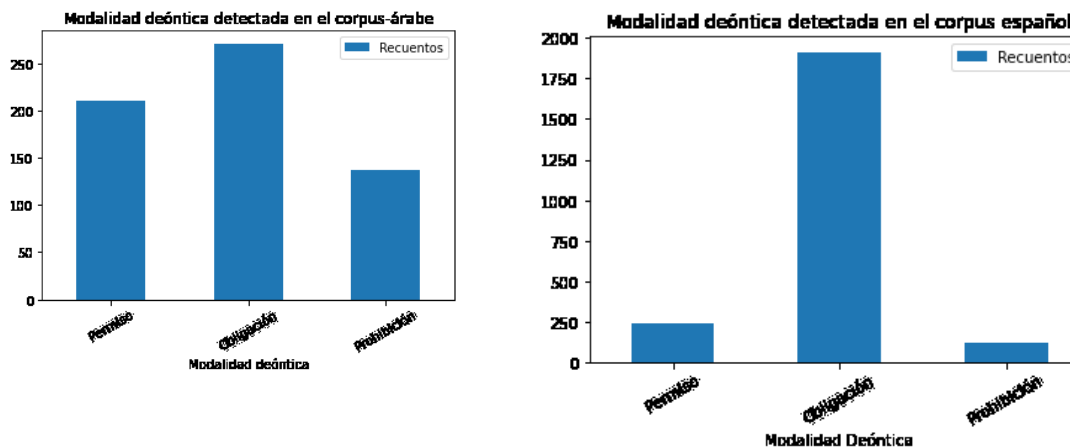


Gráfico 1. Distribución de los tipos de modalidad deóntica en el corpus comparable.

Categorías gramaticales

A través del análisis morfosintáctico, se puede analizar las categorías gramaticales que emplea cada lengua para expresar los diferentes tipos de modalidad deóntica. La distribución por categoría gramatical se refleja en la Tabla 9.

Tabla 9. Categorías gramaticales entre los tipos de modalidad deóntica.

	Sustantivo		Verbo		Adjetivo		Auxiliar		Adverbio	
	ar	es	ar	es	ar	es	ar	es	ar	es
Permiso	-	119	171	89	39	19	-	8	-	-
Obligación	41	584	190	104	40	243	-	928	-	42
Prohibición	1	28	132	51	4	37	-	-	-	-
Total por categoría	42	731	493	244	83	299	-	936	-	42

Las categorías gramaticales utilizadas varían entre las dos lenguas en cuanto a su frecuencia y su distribución entre los tipos de modalidad.

El español emplea cinco categorías gramaticales: Sustantivos, verbos, adjetivos, adverbios y verbos auxiliares, siendo la última la categoría más frecuente en español (41.5 %) para expresar la modalidad deóntica debido al empleo de perífrasis verbales como ‘deber+infinitivo’ o ‘haber+de+infinitivo’, sobre todo en la obligación (el tipo más frecuente en español). El sustantivo se sitúa en segundo lugar (32.6 %), seguido por el adjetivo en tercer lugar (13.3 %). En cuarto lugar, está el verbo (10.8 %) y, por último, está el adverbio (1.8 %).

En cambio, la modalidad deóntica en árabe se expresa mediante tres tipos de categorías gramaticales: Sustantivo, verbo y adjetivo, siendo el verbo la categoría empleada por excelencia para expresar todos los tipos de la modalidad deóntica representando el 79.8 % de los casos. En segundo lugar, está el adjetivo (13.4 %) y, por último, el sustantivo (6.8 %). Así, en árabe, no se ha detectado ningún caso de modalidad deóntica en adverbios ni en verbos auxiliares. Primero, el sistema gramatical del árabe apenas emplea el uso de los auxiliares. Segundo, es difícil distinguir automáticamente entre un adjetivo y un adverbio [إجباري\إجباريا] [obligatorio/obligatoriamente] contando solo con el análisis de categorías gramaticales (*Part-of-Speech*). Para ellos, es necesario añadir un nivel de análisis sintáctico, ya que morfológicamente el adverbio puede compartir la misma forma que un adjetivo o un sustantivo. La distinción se realiza a nivel sintáctico añadiendo una letra [‘i’] al final de la palabra cuya detección automáticamente requiere añadir un nivel de análisis sintáctico que supone un nivel de complejidad, estimado innecesario en este caso porque la precisión del análisis sintáctico automático en árabe no suele ser lo suficientemente alta y, por tanto, no es tan fiable.

Los verbos

Enfocándose en los verbos (incluyendo los verbos auxiliares), se observa que en el corpus español el 11.6 % de los verbos auxiliares

desempeñan una función deóntica (936 de 8033 verbos). De igual manera, el 10 % de los verbos en el corpus árabe desempeñan una función deóntica (493 de 4746 verbos).

Dada la relevancia del uso de los verbos (incluyendo los auxiliares) para expresar la modalidad tanto en árabe como en español, es importante analizar qué tipos de declinación se emplea en cada lengua. Las tablas V y VI recopilan una muestra de los verbos empleados en la modalidad deóntica tanto en español como en árabe en orden descendente de frecuencia.

En español, se observa que los verbos (incluyendo los auxiliares) en la forma impersonal conjugados en el futuro con la tercera persona del singular o el plural es relativamente frecuente. Por ejemplo, en la obligación, las formas ‘deberá’ y ‘deberán’ del verbo ‘deber’ representan el 27.8 % de las formas empleadas en español para expresar la obligación. Asimismo, el uso del presente del indicativo y del subjuntivo se aprecia en las modalidades de prohibición y permiso. En la modalidad de permiso, el verbo ‘permitir’ en la tercera persona del singular en el presente de subjuntivo ‘permita’, es la segunda forma más común después del sustantivo ‘permiso’. Mientras que, en la prohibición, el verbo ‘prohibir’ en la tercera persona del singular en el presente de indicativo ‘prohíbe’ es la segunda forma más frecuente para expresar este tipo de modalidad después del adjetivo ‘prohibido’.

En árabe, predomina el uso ‘impersonal’ del verbo en el presente con la tercera persona del singular. El predominio del presente está justificado por el sistema verbal en árabe que solo cuenta con el presente, el pasado y el imperativo. Además, para referirse al futuro, es normal emplear el presente con alguna partícula antes, por eso, en un contexto legislativo el presente impersonal es recurso empleado por excelencia. Por otro lado, el uso del imperativo no emplea en este contexto, ya que su uso es restringido a casos concretos de instrucciones u órdenes siempre y cuando se dirija explícitamente a un destinatario con el vocativo.

Tabla 10. Tipos y frecuencias de modalidad deóntica en español.

Permiso		Obligación ⁷		Prohibición	
Tipo	Frec.	Tipo	Frec.	Tipo	Frec.
permiso	91	deberá	349	prohibido	31
permítala	29	obligación	296	prohíbe	25
permisos	28	obligaciones	233	prohibición	24
permitan	20	deberán	184	prohíba	7
permite	13	obligado	170	prohibidas	5
permitir	11	debe	125	prohibir	5
permitido	7	obligados	56	prohibiciones	4
permitidos	7	deba	45	prohibida	4
permitirá	4	debidamente	38	prohibidos	4
No obliga	3	deben	36	No permita	3
permiten	3	deban	36	No permitirá	2
No hayan de	2	deber	34	No permitan	1
		ha de	29		

Tabla 11. Tipos y frecuencias de modalidad deóntica en árabe.

Permiso	Obligación	Prohibición
Tipo	Tipo	Tipo
يجوز [se permite]	يجب [se obliga]	لا يجوز [no se permite]
جاز [permite]	يلتزم [se debe]	يحظر [se prohíbe]
المرخص [permitido]	تلتزم [se debe (femenino)]	محظوراً [prohibido]
المصرح [permitió]	الإجباري [obligatorio]	حظر [prohibición]
رخصت [permite]	وجوب [debe]	محظور [prohibido]
	الالتزامات [obligaciones]	
	التزام [obligación]	

Negación

Es importante señalar que el uso de las formas negativas no es frecuente en español. Este dato refleja el cumplimiento de la normativa y las guías de buenas prácticas en la redacción legislativa. Esta normativa indica explícitamente la importancia de evitar el uso de la negación.

En árabe, en cambio, el empleo de la negación es frecuente, sobre todo, en la prohibición, siendo la forma más común ‘no se debe’ ‘لا يجوز’. Esta realidad contradice con las buenas prácticas que recomiendan el uso de expresiones afirmativas.

Léxico

Finalmente, ambas lenguas emplean un vocabulario reducido para expresar los diferentes tipos de la modalidad deóntica. En español, el permiso se expresa mayoritariamente con ‘permiso’ y ‘permitir’ junto a sus derivados. La obligación se expresa mayoritariamente con las diferentes formas del verbo ‘deber’ y sus derivados (debido, debidamente). Resulta relevante en este contexto señalar que la perífrasis ‘tener que’ no ocurre en los textos legislativos analizados. Por último, la prohibición en español se basa fundamentalmente en el verbo ‘prohibir’ y sus derivados.

En árabe, el verbo ‘جاز’[permitir] protagoniza la modalidad deóntica tanto en su forma afirmativa para expresar el permiso como en su forma negativa para expresa la prohibición.

7. Conclusiones y líneas futuras

En este estudio se ha presentado una herramienta que permite la detección automática de la modalidad deóntica (permiso, obligación y prohibición) en los textos legislativos en español y árabe. La metodología se ha aplicado sobre un corpus comparable que consiste en 5 legislaciones en español y sus correspondientes en árabe. El enfoque contrastivo permite analizar las características generales de los textos legislativos en cuanto a la diversidad y la densidad léxica. Además, refleja cómo cada lengua expresa la modalidad deóntica y qué categorías gramaticales emplea. Los resultados reflejan que la obligación es el tipo de modalidad deóntica predominante en los textos legislativos en ambas lenguas. También se observa que los verbos y los verbos auxiliares constituyen el recurso principal para expresar la modalidad deóntica tanto en español como en árabe.

Este trabajo es un punto de partida para otras líneas de investigación como el estudio comparativo de la modalidad deóntica entre variantes del español y el árabe entre corpus legislativos

latinoamericanos y sus equivalentes españoles y europeos o entre la legislación egipcia y la legislación de otros países árabes. Asimismo, los resultados pueden servir para seguir indagando en la ambigüedad de la modalidad deóntica mediante la aplicación de técnicas avanzadas como los modelos del lenguaje y el aprendizaje automático profundo con el fin de clasificar la modalidad deóntica teniendo en cuenta el contexto semántico.

Notas

1 Se proporcionará el enlace con la publicación del artículo.

2 www.boe.es

3 <https://www.cc.gov.eg/>

4 <https://egypls.com/category/%d8%aa%d8%b4%d8%b1%d9%8a%d8%b9%d8%a7%d8%aa-%d8%a7%d9%84%d9%8a%d9%88%d9%85/>

5 Es importante advertir que los análisis automáticos pueden incluir un margen de error. Esta realidad se suele asumir cuando se trata de técnicas de PLN, dado que el proceso automático no es siempre preciso. Este margen de error aumenta en lenguas como el árabe por su ambigüedad.

6 En árabe existen *tokens* que no se han reconocido por *Stanza* y por esto no se les ha asociado ninguna categoría.

7 Se ha incluido una parte de la lista dada su extensión.

REFERENCIAS BIBLIOGRÁFICAS

- Abdel-Fattah, M. (2005). On the Translation of Modals from English into Arabic and Vice Versa: The Case of Deontic Modality. *Babel*, 51, (31-48). DOI: <https://doi.org/10.1075/babel.51.1.03abd>
- Alcaraz Varó, E. y Hughes, B. (2002). *El español jurídico*. Barcelona, Ariel.
- Allehaibi, A. (2019). Rules of Drafting Legislative Text. *Journal of Legal Sciences*, University of Baghdad, num,1. DOI: <https://doi.org/10.35246/jols.v34i1.121>
- Ballesteros Panizo, M. López Samaniego, A. (2017). Lengua y derecho: perspectivas de estudio. *Anuari de filologia. Estudis de lingüística*, 7, (43-50). Disponible en: DOI: <https://doi.org/10.1344/AFEL2017.7.3>
- Bally, C. (1942). «Syntaxe de la modalité explicite». En Cahiers de Ferdinand de Saussure, 2, pp. 3-13.
- BOE (2005). Resolución de 28 de julio de 2005 del Acuerdo del Consejo de Ministros, de 22 de julio de 2005, por el que se aprueban las Directrices de técnica normativa. Boletín Oficial del Estado, 80, 29 de julio de 2005 (España). Disponible en: <https://www.boe.es/buscar/doc.php?id=BOE-A-2005-13020>
- Bouteffah Aghbalou, B. (2017). *La didáctica de las traducciones jurídica y jurada en la combinación árabe- español: propuesta de un manual didáctico*. Tesis doctoral. Universidad de Málaga. España. Disponible en: <https://hdl.handle.net/10630/16235>
- Burgoa-Toledo, CA. (2011). La deóntica jurídica como clave en la interpretación de las leyes fiscales. *Contaduría y administración*, 235, (57-76). Disponible en: http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0186-10422011000300004&lng=es&tlng=es
- Bybee, J. y Fleischman, S. (1995). *Modality in Grammar and Discourse*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

-
- Chierichetti, L. y Garofalo, G. (2011). Lengua y Derecho: líneas de investigación interdisciplinaria. *Revista de Llengua i Dret. Journal of Language and Law*; (56). Disponible en: <http://revistes.eapc.gencat.cat/index.php/rld/article/view/2416>
- Da Cunha, I. y Escobar, MA. (2021). Recomendaciones sobre lenguaje claro en español en el ámbito jurídico-administrativo: análisis y clasificación. *Pragmalingüística*, 29, (129-148). <https://doi.org/10.25267/Pragmalinguistica.2021.i29.07>
- Da Cunha, I. (2020). [Una aplicación tecnológica que ayuda a la ciudadanía a escribir textos a la Administración pública](#). *Linguamática* 12(1), (97-115) Disponible en: DOI: <https://doi.org/10.21814/lm.12.1.316>
- El Krirh, K. (2017). *Sobre la traducción jurídica y judicial: Estudio de la variación aplicado a la Terminología del Derecho Penal y Procesal Penal en español, francés y árabe*. Tesis doctoral. Universidad de Málaga. España. Disponible en: <http://hdl.handle.net/10630/11503>
- El-Farahaty, H. y Elewa, A. (2020). A Corpus-Based Analysis of Deontic Modality of Obligation and Prohibition in Arabic/English Constitutions. *Estudios de Traducción*. 10, (107-136). DOI: <https://doi.org/10.5209/estr.68334>
- Franco, G. (1962). *Las leyes de Hammurabi*. *Revista De Ciencias Sociales*, 3, (331-356). Disponible en: <https://revistas.upr.edu/index.php/rcs/article/view/9466>
- Fuentes-Riffo, K., Hernández-Osuna, S. y Salcedo-Lagos, P. (2019). Descripción de la diversidad y densidad léxicas en noticias escritas por estudiantes de periodismo. *Revista Brasileira de Lingüística Aplicada*, 25(3), (499-528). Disponible en: <https://doi.org/10.1590/1984-6398201914113>
- Gogazeh, Z. (2014). La problemática de la traducción de la terminología jurídica en el libro Manual de traducción jurídica árabe-español (2014). *Philologia hispalensis* 28 (1), (25-34). DOI: <https://doi.org/10.12795/PH.2014.v28.i01.02>

-
- Greenberg, M. (2011). Legislation as Communication? Legal Interpretation and the Study of Linguistic Communication. En Marmor, A. y Soames, S. (Eds.) *Philosophical Foundations of Language in the Law*. Oxford University Press. Disponible en: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1726567
- Hinson, CL. (2005). Legal Informatics: Opportunities for Information Science. *Journal of Education for Library and Information Science*, 46(2), (134–153). DOI: <https://doi.org/10.2307/40323866>
- Jiménez-Salcedo, J. y Moreno-Rivero, J. (2017). On jurilinguistics: The principles and applications of research on language and law. *Journal of Language and Law*, 68, (1-4). DOI: <https://doi.org/10.2436/rld.i68.2017.3064>
- Katz, DM., Dolin, R. y Bommarito, MJ. (Eds.) (2021). *Legal informatics*. Cambridge University Press.
- López-Morales, H.(2011). Los índices de riqueza léxica y la enseñanza de lenguas. En *Del texto a la lengua: La aplicación de los textos a la enseñanza-aprendizaje del español L2-LE* (pp. 15-28). Asociación para la Enseñanza del Español como Lengua Extranjera. Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=5419218>
- Matte Bon, F. (1995). Gramática comunicativa de la lengua española. Madrid: Edelsa, nueva edición revisada.
- Medina Montero, JF. (2013). El español de los textos jurídicos comunitarios y españoles: el caso de los marcadores del discurso presentes en un corpus de reglamentos del DOUE y en otro de leyes del BOE.
- En Ondelli, S. (Coord.), *Realizzazioni testuali ibride in contesto europeo. Lingue dell'UE e lingue nazionali a confronto*, (127-144). EUT Edizioni Università di Trieste: Trieste. Disponible en: <http://hdl.handle.net/10077/9144>
- Montolío Durán, E. (2012). *Hacia la modernización del discurso jurídico*. Publicacions i Edicions de la Universitat de Barcelona. DOI: <http://dx.doi.org/10.2436/20.803>

-
- Moreno Rivero, J. y Jiménez-Salcedo, J. (2020). La jurilingüística como marco de la traducción jurídica: géneros, corpus y formación. *Estudios de Traducción*, 10, (9-16). DOI: <https://doi.org/10.5209/estr.72082>
- Morillo-Velarde Pérez, R. (2011). Tradiciones discursivas y derecho: la modalidad discursiva deóntica en la constitución de 1812. En *Sintaxis y análisis del discurso hablado en español: Homenaje a Antonio Narbona*, 683. Disponible en: <https://core.ac.uk/download/pdf/51401778.pdf#page=107>
- OECD, 2019: *Good Governance in Egypt: Legislative Drafting Manual for Better Policy*, OECD Publishing, Paris. DOI: <https://doi.org/10.1787/g2g9dd64-en>
- Parodi, G. (2005). Discurso especializado y Lingüística de Corpus: hacia el desarrollo de una competencia psicolingüística, *Boletín de Lingüística*, 17, (61-88). Disponible en: http://ve.scielo.org/scielo.php?script=sci_arttext&pid=S0798-97092005000100004
- Polo, A. (2013). Valores deónticos en un corpus comparable. En Stefano Ondelli (Ed.), *Realizzazioni testuali ibride in contesto europeo. Lingue dell'UE e lingue nazionali a confronto*, pp. 107-126. Trieste, EUT Edizioni Università di Trieste. Disponible en: https://www.openstarts.units.it/bitstream/10077/9143/1/Polo_RealizzazioniTestualiIbride.pdf
- Pontrandolfo, G. (2012). Legal Corpora: an overview. *Rivista Internazionale di Tecnica della Traduzione*, 14, (121-136). Disponible en: https://www.academia.edu/11163914/Legal_Corpora_An_Overview
- Pontrandolfo, G. (2017). Legal Linguistics in Spain. En Vogel, F. (Ed.), *Legal Linguistics Beyond Borders: Language and Law in a World of Media, Globalisation and Social Conflicts*. Dunckel&Humbolt. Berlin. DOI: <https://doi.org/10.3790/978-3-428-55423-2>

-
- Prieto de Pedro, JJ. (1991). *Lenguas, lenguaje y derecho*, Cívitas, Madrid.
- Qi, P., Zhang, Y., Zhang, Y., Bolton, J. y Manning, C. (2020). [Stanza: A Python Natural Language Processing Toolkit for Many Human Languages](https://doi.org/10.48550/arXiv.2003.07082). En *Association for Computational Linguistics (ACL) System Demonstrations*. DOI: <https://doi.org/10.48550/arXiv.2003.07082>
- Real Academia Española de la Lengua (1999). *Gramática descriptiva de la Lengua Española*, dirigida por Ignacio Bosque y Violeta de Monte, Madrid, Espasa.
- Saad, MS. (2015). *Curso práctico de traducción legal*, I Árabe/Español. Instituto Egipcio de Estudios Islámicos, Madrid.
- Soto Aranda, B. y Kabbour, R. (2020). Traducción comentada de un acta notarial de celebración de matrimonio civil del español al árabe. *Anaquel de Estudios Árabes*, 31, (179-205). DOI: <https://doi.org/10.5209/ange.66701>
- Vogel, F. (Ed.) (2019). *Legal Linguistics Beyond Borders: Language and Law in a World of Media, Globalisation and Social Conflicts.: Relaunching the International Language and Law Association (ILLA)*. Duncker & Humblot GmbH, 2019. Disponible en: <http://www.jstor.org/stable/j.ctv1q69j9f>
- Wiltos, A. (2020). El destinatario del enunciado modalizado deónticamente a partir del análisis de textos de derecho primario de la Unión Europea en español. *Studia Iberystyczne* 19 (diciembre) (299-310). DOI: <https://doi.org/10.12797/SI.19.2020.19.13>.
- علي أحمد حسن (٢٠١٩). قواعد صياغة النص. مجلة العلوم القانونية. المجلد ٤٣ العدد ١. جامعة بغداد
- . DOI: <https://doi.org/10.35246/jols.v34i1.121>