



UNIVERSIDAD
COMPLUTENSE
MADRID

MÁSTER EN LINGÜÍSTICA Y TECNOLOGÍA
ESPECIALIDAD LINGÜÍSTICA COMPUTACIONAL

TRABAJO DE FIN DE MÁSTER
Curso académico 2024-2025

**Desarrollo de una herramienta
computacional para textos más
inclusivos respecto al género y la
diversidad funcional**

APELLIDOS Y NOMBRE: Katsani María

DNI: AT0539699

TUTOR: Doaa Samy Khalil Shower

CONVOCATORIA: Extraordinaria

Índice

Índice de figuras.....	3
Índice de tablas.....	4
Siglas.....	5
1. Resumen.....	6
2. Abstract.....	6
3. Introducción	7
4. Estado de la cuestión.....	9
5. Marco teórico lingüístico.....	12
5.1. Debate lingüístico sobre la inclusión de género.....	12
5.2. Evolución de términos sobre la diversidad funcional.....	17
6. Metodología.....	19
6.1. Estrategias para un lenguaje inclusivo con relación al género.....	19
6.2. Selección y justificación de expresiones lingüísticas para el uso del lenguaje inclusivo en relación con el grupo social de personas con discapacidad.....	22
6.3. Desarrollo.....	23
6.3.1. Datos y herramientas disponibles.....	23
6.3.2. Preparación de los datos y pre-procesamiento.....	26
6.3.3. Entrenamiento.....	29
6.3.4. Retos, dificultades y excepciones.....	32
6.4. Análisis de los resultados.....	33
6.4.1. Análisis de resultados.....	33
6.4.2. Análisis de resultados de pruebas prácticas	34
6.4.3. Análisis y comparación de resultados con otras herramientas.....	38
7. Conclusiones.....	43
8. Bibliografía.....	44
9. ANEXO.....	49

Índice de figuras

Figura 1. Ejemplo de la herramienta Themis	10
Figura 2. Ejemplo de la herramienta CaDi.....	11
Figura 3. Ejemplo de la herramienta spaCy NER Annotator.....	23
Figura 4. Esquema de preprocesamiento.....	29
Figura 5. <i>Input Frame</i>	30
Figura 6. Esquema de entrenamiento.....	30
Figura 7. Ejemplo de reconocimiento de expresiones relacionadas a la inclusión de género spaCy.....	34
Figura 8. Ejemplo I de sugerencias spaCy.....	34
Figura 9. Ejemplo de reconocimiento de expresiones relacionadas a la inclusión de género BERT.....	35
Figura 10. Ejemplo I de sugerencias BERT.....	35
Figura 11. Ejemplo de reconocimiento de expresiones relacionadas a la diversidad funcional spaCy.....	36
Figura 12. Ejemplo II de sugerencias spaCy	36
Figura 13. Ejemplo de reconocimiento de expresiones relacionadas a la diversidad funcional BERT.....	37
Figura 14. Ejemplo II de sugerencias BERT.....	37

Índice de tablas

Tabla 1. Palabras elegidas relacionadas al género y sus alternativas.....	22
Tabla 2. Palabras elegidas relacionadas a la diversidad funcional, sus alternativas y justificación.....	22
Tabla 3. Datos del corpus.....	24
Tabla 4. Datos del entrenamiento.....	24
Tabla 5. Datos de validación.....	26
Tabla 6. Datos de prueba.....	26
Tabla 7. Ejemplo del etiquetario.....	27
Tabla 8. Parámetros de entrenamiento.....	33
Tabla 9. Métricas de los resultados spaCy.....	32
Tabla 10. Métricas de los resultados BERT.....	33

Siglas

BERT	<i>Bidirectional Encoder Representations from Transformers</i>
COCEMFE	Confederación Española de Personas con Discapacidad Física y Orgánica
CORPES XXI	Corpus del Español del Siglo XXI
GPT	<i>Generative Pre-Trained Transformer</i> (Transformer generativo preentrenado)
LC	Lingüística Computacional
NER	<i>Named Entity Recognition</i> (Reconocimiento de entidades nombradas)
OMS	Organización Mundial de la Salud
ONCE	Organización Nacional de Ciegos Españoles
ONG	Organización No Gubernamental
PLN	Procesamiento del Lenguaje Natural
RAE	Real Academia de la lengua Española
RTVE	Radiotelevisión Española
UCM	Universidad Complutense de Madrid
UN	Naciones Unidas

1. Resumen

El lenguaje desempeña un papel crucial en la configuración de la realidad social. Aunque se ha avanzado en la concienciación sobre el uso inclusivo del lenguaje para combatir estereotipos sigue existiendo mucho debate, y todavía persisten prácticas lingüísticas que se pueden considerar discriminatorias. Motivado por la necesidad de una comunicación más inclusiva, este trabajo de fin de máster tiene como objetivo desarrollar una herramienta basada en técnicas de Procesamiento del Lenguaje Natural (PLN) que identifique sesgos lingüísticos en textos escritos, especialmente en cuestiones de género y diversidad funcional y sugiera alternativas. Para ello, se han abordado diferentes técnicas del estado de la cuestión en PLN que demostraron ser efectivos en proporcionar soluciones para un lenguaje más inclusivo. En conclusión, esta investigación contribuye al avance de tecnologías que impulsan la equidad y el respeto por la diversidad, reafirmando el papel del lenguaje como un medio para promover una sociedad más justa.

Palabras clave: procesamiento de lenguaje natural, lenguaje inclusivo, género, diversidad funcional

2. Abstract

Language plays a crucial role in shaping reality. Although progress has been made in raising awareness about the use of inclusive language to combat gender stereotypes, much debate still exists, and discriminatory linguistic practices continue to persist. Motivated by the need for more inclusive communication, this master's thesis aims to develop a computational tool applying state-of-the-art Natural Language Processing (NLP) techniques to identify linguistic biases in written texts, particularly concerning gender and functional diversity, and, thus, suggest alternatives. Using natural language processing methodology, a tool was developed and evaluated, demonstrating its effectiveness in promoting more inclusive language. In conclusion, this research contributes to the advancement of technologies that enhance equity and respect for diversity, reaffirming the role of language as a means to promote a more equal society.

Key words: natural language processing, inclusive language; gender, diversity of functional abilities

3. Introducción

El lenguaje humano es un instrumento que desempeña un papel imprescindible en reflejar la realidad y los cambios sociales. En una sociedad contemporánea donde los valores de inclusión e igualdad son pilares fundamentales, el lenguaje también debería reflejar estos principios. Por eso, surge el debate en la sociedad contemporánea y en la comunidad de los expertos lingüistas sobre hasta qué punto la manera en la que se expresan y escriben los hablantes efectivamente refleja estos avances sociales.

El uso inclusivo del lenguaje es un medio poderoso para combatir estereotipos hacia ciertos grupos sociales y fomentar una cultura de inclusión y equidad. Sin embargo, a pesar de los avances en la concienciación sobre el lenguaje políticamente correcto, persisten prácticas lingüísticas que invisibilizan o discriminan a algunos grupos. Este fenómeno es evidente en diversos ámbitos, desde los medios de comunicación hasta los textos administrativos.

Por el lado contrario, la promoción de un lenguaje inclusivo se ha convertido en una prioridad para numerosos organismos oficiales como ministerios, universidades, organizaciones europeas e internacionales y organizaciones no gubernamentales. (ONGs), etc. han creado guías para un lenguaje inclusivo dentro de las cuales se incluyen instrucciones y sugerencias de uso del lenguaje inclusivo. Por ejemplo, el Ministerio español de Sanidad, Servicios Sociales e Igualdad ha publicado una recopilación de todas las guías generadas a nivel oficial (*Guías para el uso no sexista del lenguaje*) (Ministerio español de Sanidad, Servicios Sociales e Igualdad, 2015).

A nivel europeo, en 2008 el Parlamento Europeo fue la primera organización mundial en publicar la guía de uso de un lenguaje neutro en el Parlamento y lo actualizó en 2018 ([Parlamento Europeo, 2018](#)). En febrero 2024, la Comisión Europea publicó la guía para el uso del lenguaje como un medio de inclusión (*Guidelines for the use of language as a driver of inclusivity*) ([Comisión Europea, 2024](#))

A nivel internacional, organizaciones como las Naciones Unidas (UN) defienden la adopción de “una serie de estrategias para que el personal de las Naciones Unidas emplee un lenguaje inclusivo en cuanto al género. Las estrategias pueden aplicarse a todo tipo de situación comunicativa, oral o escrita, formal o informal, con un público tanto interno como externo.” (Naciones Unidas, 2024).

Es importante mencionar que algunas guías como la de la Universidad Complutense de Madrid no se limitan solamente al uso de lenguaje sino incluyen “[...] ejemplos muy útiles, así como la puesta en práctica de habilidades comunicativas efectivas, destinadas a evitar

la discriminación y la exclusión de cualquier persona que tenga relación con nuestra institución educativa.” (Universidad Complutense de Madrid, 2021)

En el marco de este contexto sociolingüístico, el presente estudio plantea una cuestión: ¿Se pueden emplear técnicas de Procesamiento del Lenguaje Natural (PLN) y Lingüística Computacional (LC) para desarrollar herramientas que puedan asistir en evitar sesgos y mejorar la inclusividad? ¿Existen herramientas de esta naturaleza que procesen los textos en español?

La hipótesis central del estudio parte de que una herramienta automatizada pueda asistir y sensibilizar a los usuarios sobre la importancia del lenguaje inclusivo y políticamente correcto para así ayudarles a corregir prácticas lingüísticas excluyentes de manera eficiente.

El objetivo general de este trabajo de fin de máster se centra en el desarrollo de herramientas informáticas que analizan los textos en español basándose en técnicas de PLN y LC para detectar expresiones lingüísticas objeto de una posible mejora, sugiriendo propuestas para un lenguaje más inclusivo.

El diseño de la herramienta y la selección de las expresiones a identificar junto a las propuestas sugeridas, se fundamentan en una base teórica lingüística donde se ha evaluado las diferentes perspectivas y se ha optado por propuestas donde se prioriza la eficacia comunicativa y la economía lingüística sobre las tendencias sociopolíticas. De esta manera, los experimentos llevados a cabo pretenden proporcionar a los usuarios una herramienta que les asista en corregir sesgos, promoviendo un uso del lenguaje más inclusivo y a la vez respetando la naturaleza lingüística y la eficacia comunicativa

Para conseguir este objetivo general, se plantean los siguientes objetivos específicos:

- Analizar el estado de la cuestión contrastando posturas acerca del debate lingüístico.
- Recopilar una muestra de textos escritos para el desarrollo de la herramienta.
- Explorar las diferentes técnicas que se pueden emplear.
- Diseñar y desarrollar herramientas basándose en las técnicas de PLN.
- Entrenar los modelos para la detección de expresiones como objeto de mejora.
- Evaluar los modelos y analizar los retos y las limitaciones.

La justificación de este estudio radica en la necesidad de contar con recursos efectivos que apoyen la implementación de un lenguaje inclusivo en diversos contextos. Se pueden considerar herramientas aplicadas las que incluyen las recomendaciones de la guías a la práctica, contribuyendo así al campo de la inclusión, ya que dichas guías no están accesibles siempre a todos los usuarios y hablantes, y a veces resulta bastante complicado memorizar todas las diferentes maneras de aplicarlas.

En el marco teórico y lingüístico, se ha realizado un estudio breve sobre el debate lingüístico actual del lenguaje inclusivo respecto al género, y la evolución de los términos que se consideran políticamente correctos respecto a diversidad funcional.

En relación con la metodología implementada se aplican tres técnicas combinando la investigación cualitativa y cuantitativa para desarrollar y evaluar la herramienta. Para ello, se implementan técnicas de PLN como el reconocimiento de entidades (expresiones lingüísticas en este caso) para conseguir identificar estas expresiones y sugerir los puntos de mejora. Las tres técnicas aplicadas son: la librería spaCy, ajustando un modelo BERT, y por último una breve comparativa mediante instrucciones (*prompt*) a ChatGPT de OpenAI.

4. Estado de la cuestión

Este estado de la cuestión tiene como objetivo revisar la literatura y las herramientas existentes sobre el uso inclusivo del lenguaje enfocado en los aspectos relacionados con el género y la diversidad funcional. A través de esta revisión, se pretende identificar las principales contribuciones, herramientas existentes y detectar las posibles limitaciones o aspectos que no han sido tratados hasta la actualidad para plantear nuevas hipótesis.

Con relación al lenguaje inclusivo desde la perspectiva del género, cada vez hay más estudios de literatura que muestran el interés por este tema. El debate sobre el lenguaje inclusivo ha sido ampliamente abordado en varias obras. En la *Guía de comunicación no sexista* de Antonio Briz et al. (2021), se subraya que el objetivo es "lograr un trato lingüístico más igualitario, que visualice por igual a hombres y mujeres, así como evitar y erradicar el sexismo del lenguaje". De forma similar, *Género Neutro y Lenguaje Inclusivo: Una Pequeña Guía Práctica* de Lauredal y Gelen Jeleton (2019) promueve la adopción del género neutro en el castellano, argumentando que "el uso de términos como 'elle' o 'une' puede ayudar a que el lenguaje sea más inclusivo y representativo de todas las personas". Y María Martín, en su obra *Ni Por Favor Ni Por Favors*, que critica la postura conservadora de la Real Academia Española y sugiere que "hablar con lenguaje inclusivo no tiene que ser forzado ni notorio, sino una adaptación natural que refleje la realidad social" (Martín, 2020).

Desde mi punto de vista, el camino hacia un lenguaje más inclusivo debe ser colaborativo, combinando enfoques prácticos y adaptándose a las necesidades cambiantes de los hablantes. Por esta razón, no comparto el enfoque de género neutro propuesto por Lauredal y Gelen Jeleton (2019), ya que introduce reglas no normativas que podrían generar rechazo, confusión y, en última instancia, resultar contraproducentes. Aunque la postura de Antonio Briz et al. (2021) apunta a un objetivo ideal, considero que erradicar el sexismo del lenguaje sólo será posible si también se elimina de la sociedad. Por lo tanto, la postura de Martín (2020) está más alineada con la evolución natural del lenguaje, en el cual hay que reflejar los cambios sociales siempre y cuando sean accesibles y aceptados por la mayoría de los hablantes para tener un impacto real.

Por otro lado, y desafortunadamente, en el estudio realizado no se ha encontrado mucha literatura sobre lenguaje inclusivo relacionado con la diversidad funcional. Solo se ha podido localizar algunos artículos divulgativos y los manuales previamente mencionados.

Entre estos manuales, se destaca el manual de la Confederación Española de Personas con Discapacidad Física y Orgánica (COCEMFE). “COCEMFE es una ONG con propósito de conseguir una sociedad inclusiva que garantice el ejercicio pleno de los derechos de las personas con discapacidad física y orgánica.”(COCEMFE, 2024).

En la introducción se aclaran las diferencias entre discapacidad física y discapacidad orgánica subrayando que las personas con discapacidad, ante todo son personas y no quieren ser etiquetadas, considerando que la discapacidad es una característica más de todas las que tienen las personas, y no ha de ser lo único por la que una persona se reconoce. En general, se defiende evitar la calificación de las personas como “discapacitadas”, sino que son personas que tienen una discapacidad y se proponen alternativas para reafirmar esta postura.

Desde la perspectiva computacional se ha llevado a cabo un análisis del estado de la cuestión en cuanto a los recursos y las herramientas disponibles. En este análisis se han identificado algunas herramientas de pago y otras gratuitas que procesan los textos en español y ofrecen alternativas para un lenguaje más inclusivo. Entre las herramientas localizadas, dos son de libre uso y una de pago con fines comerciales. A continuación, se describen estas herramientas:

- [Themis](#) es un software con fines comerciales, desarrollado por Natural Language Analysis, que mediante el uso de la inteligencia artificial transforma los textos en lenguaje inclusivo solamente de género. La herramienta tiene la capacidad de analizar documentos de texto, carpetas, bandejas de correo electrónico y páginas web. Detecta términos no inclusivos o sexistas y ofrece una amplia gama de alternativas.

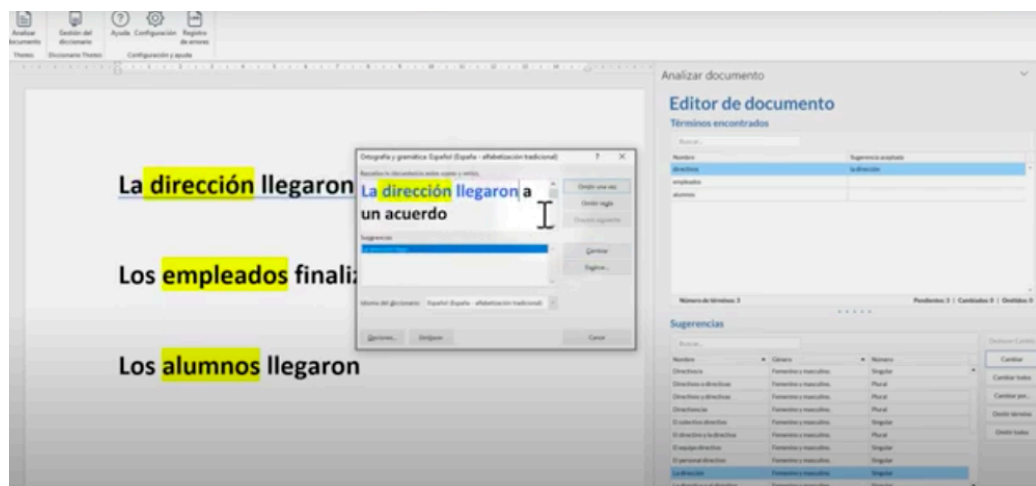


Figura 1. Ejemplo de la herramienta Themis. Extraído de themis.es

- [CaDi traductor de lenguaje incluyente](#), desarrollado por la Universidad Iberoamericana de México, es una herramienta de apoyo de uso libre para ayudar a emplear el lenguaje incluyente en la redacción de textos.

Incluye de base algo más de 500 palabras con dos o tres opciones en lenguaje inclusivo con la finalidad de evitar el uso del lenguaje desde una práctica androcéntrica.

Se ha consultado la herramienta en diferentes momentos durante la preparación del presente trabajo y se han observado mejoras en las sugerencias propuestas. Por ejemplo, en las primeras consultas que se realizaron, la herramienta sustituía la palabra “alumnos” por “alumnos y alumnas” proponiendo así una sugerencia ineficiente desde el punto de vista lingüístico y comunicativo creando así un efecto indeseado. En las consultas posteriores, la herramienta sugiere como alternativa la palabra “el alumnado” que es una sugerencia más adecuada y eficiente.

Pegar / agregar / capturar texto
*** En caso de palabras con errores ortográficos, la herramienta no emitirá texto sugerido para esas palabras**

los alumnos de la clase de matemáticas estaban emocionados por el nuevo proyecto que el profesor les había asignado. Durante la clase, los alumnos trabajaron en grupos para resolver los problemas y discutieron las posibles soluciones entre ellos. Al finalizar, el profesor elogió a los alumnos por su dedicación y les recordó que podían consultarlo si tenían dudas antes de la entrega final.

391/1000

Borrar

Comprobar Texto

Texto sugerido

el alumnado de la clase de matemáticas estaban emocionados por el nuevo proyecto que el profesor les había asignado. Durante la clase, el alumnado trabajaron en grupos para resolver los problemas y discutieron las posibles soluciones entre ellos. Al finalizar, el profesor elogió a el alumnado por su dedicación y les recordó que podían consultarlo si tenían dudas antes de la entrega final.

391/1000

Figura 2. Ejemplo de la herramienta CaDi. Fuente propia

- [Ventana Inclusiva y Xanela Inclusiva](#), desarrolladas por imaxin|software son aplicaciones correctoras de estilo que se instalan en el editor de texto, para un uso inclusivo y no discriminatorio de la lengua castellana y gallega. Estos correctores están financiados por la Fundación Organización Nacional de Ciegos Españoles (ONCE). Desafortunadamente, después de varios intentos no se ha podido descargar y probar dicha aplicación, por cuestiones de fallos en su página web que no permite la descarga.

Es importante mencionar que ninguna de esas herramientas propone mejoras respecto a la inclusión de personas con diversidad funcional o personas de otras etnias, etc. Por eso, eligiendo cuidadosamente las mejoras que se van a aplicar en los textos este trabajo viene a responder a esta necesidad.

5. Marco teórico lingüístico

Antes de comenzar a examinar el marco teórico del lenguaje inclusivo, es crucial definir qué es el lenguaje inclusivo, cómo se relaciona con el género y la diversidad funcional, y qué entendemos por lenguaje políticamente correcto.

Cuando se habla de lenguaje inclusivo, en varias ocasiones se asocia con incluir exclusivamente a las mujeres en la forma de hablar. A veces, incluso se confunde con el lenguaje políticamente correcto. Sin embargo, se trata de términos con distintos matices, y ambos abarcan más conceptos que cuyo alcance va más allá de la simple inclusión de las mujeres.

- Definición Lenguaje inclusivo: "El lenguaje inclusivo es un recurso comunicativo que busca visibilizar a todos los grupos sociales, especialmente aquellos que han sido históricamente discriminados o invisibilizados, como las mujeres, las personas con diversidad funcional, las personas de diferentes etnias y culturas, entre otros." (Instituto Cervantes, 2021)
- Definición Lenguaje políticamente correcto: "Lenguaje políticamente correcto es el que evita el uso de palabras o comportamientos que pueden ser ofensivos o perjudiciales para ciertos grupos de personas." (Cambridge Advanced Learner's Dictionary & Thesaurus, Cambridge University Press, 2013)

En otras palabras, mientras el lenguaje políticamente correcto busca no ofender a ningún individuo, el lenguaje inclusivo tiene como objetivo visibilizar a todos los grupos sociales. El lenguaje inclusivo no se trata simplemente de evitar ofender; en cambio, se trata de reconocer y dar visibilidad a grupos históricamente marginados. Ambos términos están estrechamente conectados, ya que para visibilizar a un grupo es necesario referirse a él sin utilizar términos que puedan tener connotaciones negativas y resultar ofensivos. De esta manera, tanto el lenguaje políticamente correcto como el inclusivo buscan que todas las personas se sientan incluidas y reconocidas como iguales dentro de la sociedad.

Sin embargo, ambos términos han generado cierto debate en el ámbito social y en la comunidad científica sobre su uso y su razón de existir teniendo en cuenta que a pesar de tratarse de fenómenos lingüísticos, están estrechamente relacionados con los aspectos socio-políticos.

En esta sección, se van a examinar los argumentos más destacables tanto a favor como en contra del lenguaje inclusivo respecto al género, y se va a explorar la evolución de los términos relacionados a la diversidad funcional a lo largo de las últimas décadas.

5.1. Debate lingüístico sobre la inclusión de género

El debate acerca del lenguaje inclusivo es más complejo de lo que parece a primera vista, porque se trata de un problema ideológico que involucra a la lingüística, a la política y a la sociedad. Es indudable que el tema tiene importancia, prueba de ello es la ingente cantidad de documentos institucionales donde numerosos organismos han vertido recomendaciones orientadas a la eliminación de usos y expresiones consideradas sexistas y, por otra parte, la abundancia de reflexiones, artículos de opinión, etc. Como expresó Corral Herrera (2019) “el lenguaje es político, es decir, es una construcción humana. En todas nuestras construcciones subyace una ideología, una visión del mundo particular, y transmiten por tanto unos valores basados en los intereses de los grupos dominantes”. Eso significa, en cuanto a la política lingüística, el discurso científico muchas veces puede verse influido por ideologías sociales y políticas. La cuestión de si el lenguaje y el trabajo de los lingüistas deberían contribuir a cambiar la realidad es una cuestión de ética personal del investigador y del político. Además, frecuentemente refleja sus propias creencias ideológicas.

Es una realidad que el debate acerca del lenguaje inclusivo ha sido objeto de atención desde hace muchos años, empezando en los EEUU, para destacar el sexismo, el androcentrismo y el dominio de una cultura patriarcal, pero se intensificó en el mundo hispano en julio de 2018, cuando el gobierno español, solicitó formalmente en el congreso la adecuación de la Constitución de dicho país a un lenguaje “inclusivo, correcto y verdadero a la realidad de una democracia que transita entre hombres y mujeres”, y anunció que había solicitado un informe a la RAE. La primera reacción de la RAE no fue positiva, pero su director, se comprometió a dar un informe que sería “un trabajo serio y despolitizado” (El Universal, 2018)

Este informe consiste en tres secciones diferenciadas:

“La primera es el “Informe de la Real Academia Española sobre el uso del lenguaje inclusivo en la Constitución española”. Este texto ha sido elaborado a petición de la vicepresidenta del Gobierno. La segunda es un documento titulado “Sobre sexismo lingüístico, femeninos de profesión y masculino genérico. Posición de la RAE”. Y la tercera la constituyen dos anexos a este último documento que contienen, respectivamente, una muestra de las respuestas que el departamento @RAEinforma emite a diario a requerimiento de los hablantes sobre cuestiones de género y, por otro lado, una relación de enmiendas relacionadas con el lenguaje inclusivo incorporadas al Diccionario de la lengua española (DLE).”

(Real Academia Española de la Lengua, 2020)

El informe concluye que el texto constitucional utiliza un español correcto desde el punto de vista normativo, basándose en el uso predominante entre los hablantes de español, documentado en diversas obras y corpora lingüísticos. Además, se recomienda considerar una posible reforma enfocada en expresiones relacionadas con el matrimonio, las relaciones de pareja, las relaciones entre padres e hijos y la discapacidad.

Sin embargo, este informe ha sido criticado intensamente por otros lingüistas por no dar una respuesta satisfactoria que refleja la sociedad actual. De ahí, empieza el debate del lenguaje inclusivo en España en el que principalmente hay dos posturas: la de la Real Academia Española (RAE) con lo que esta autoridad defiende y la postura que promueve el lenguaje inclusivo como una acción para lograr la igualdad de género. Se toman posiciones ideológicas difíciles de conciliar en un diálogo que no siempre es cooperativo, sino más bien conflictivo. De este modo, estamos ante un problema lingüístico en el que las ideologías lingüísticas y políticas se cruzan. Por esto, es necesario diferenciar entre lenguaje inclusivo y exclusión social, conceptos que están en la base del problema. Asimismo, es necesario tener en cuenta conceptos fundamentales como la economía lingüística y la eficacia comunicativa.

En este contexto, surgen puntos de debate como el uso del masculino genérico y la interrelación lengua-sociedad-pensamiento y aceptación o rechazo de términos en el diccionario y otros puntos de controversia.

Uno de los puntos del debate, aunque no esté incluido dentro del informe de la RAE pero sí está presente dentro de la comunidad científica, es el argumento que el lenguaje está estrechamente conectado con la sociedad y por lo tanto existe el sexismo lingüístico, porque existe el sexismo ideológico. García Meseguer (1996) analizó el debate sobre el sexismo lingüístico en los usuarios de la lengua, afirmando que “la lengua española no es sexista, lo son los las personas que la utilizan”. El autor describe esta situación de esta manera. “Se incurre en sexismo lingüístico cuando se emplean vocablos (sexismo léxico) o se construyen oraciones (sexismo sintáctico) que debido a la forma de expresión escogida por el hablante, resultan discriminatorias por razón de sexo.” (Meseguer, 1996). Así es importante no confundir el plano social, de la realidad y el plano lingüístico. Por ejemplo: “*El gobierno español consta de tres mujeres y quince varones.*” describe una realidad sexista pero de forma no lingüísticamente sexista. Pues es sexismo social, no lingüístico. En cambio, en la frase “*El gobierno español consta de tres varones y quince hembras.*” encontramos sexismo lingüístico, ya que hay falta de simetría que se da al usar *hembra* en lugar de *mujer*, que es el par correlativo a *varón*. Otro ejemplo sería: “*En los pueblos la gente solo quiere estar tranquila, disfrutar con sus mujeres, vivir sin preocupaciones.*” En este ejemplo, se identifica *gente* con varones únicamente.

El sexismo social y el lingüístico están en interdependencia; diacrónicamente, el primero origina el segundo, pero el segundo contribuye a reforzar al primero, ya que el lenguaje conforma, en parte, la mentalidad de los hablantes. El problema que ocurre está en los límites de cada uno porque desde una perspectiva de análisis pragmalingüístico, se considera que el sexismo ha de explicarse teniendo en cuenta tanto el sistema de la lengua como el contexto de uso, y por lo tanto, las circunstancias sociales e históricas actuales presentes en la mentalidad de los hablantes. Teniendo en cuenta eso, resulta difícil marcar conceptos de límites tan borrosos para el sexismo social y el sexismo lingüístico, y por eso es coherente hablar de discurso sexista que se define como: “El conjunto de rasgos y usos lingüísticos discriminatorios derivados de la situación y la acción social.” (Vigara, 2009)

Otro punto de debate y el que más debate ha generado está a nivel morfológico es el uso de masculino genérico. Este fenómeno es el que se comenta en el informe de la RAE sobre lenguaje inclusivo.

El masculino es el término no marcado para todos los hispanohablantes en un gran número de contextos, especialmente cuando se asocia con el plural. La expresión nominal “*los pasajeros del avión*” no invisibiliza a las pasajeras ni es irrespetuosa con ellas, sino que las abarca o las incluye, de acuerdo con el sentimiento lingüístico de los hispanohablantes de todo el mundo.

(RAE, 2019)

En la contraposición se encuentran argumentos que destacan efectivamente las mujeres no se ven incluidas en la palabras masculinas (como pasajeros, del ejemplo anterior) por el simple hecho de que existen diferentes palabras para referirse a lo femenino y a lo masculino, lo cual afirma que ambos sexos no son lo mismo ya que hace falta tener palabras diferentes. Por lo tanto, si no son lo mismo, no se pueden sentir identificadas las mujeres dentro de los términos masculinos.

Aunque la Nueva gramática de la lengua española indica que el masculino es el género no marcado, y por lo tanto sirve para nombrar lo masculino y lo femenino y que eso no hace el español sexista, no es verdad. Aunque se menciona que no tiene que ver con hombres y mujeres, sino con asignaciones gramaticales independientes del sexo, no es verdad. Aunque se diga que decir *el niño* y *la niña* es redundante, no lo es. Es redundante lo que se nombra dos veces. Un niño y una niña no son lo mismo. Porque si lo fueran, la propia palabra sería redundante y atentaría contra la economía del lenguaje. ¿Para qué iba una sociedad a formar dos palabras distintas para una realidad idéntica sin que una de ellas pueda abarcar a la totalidad del conjunto, como pasa con los sinónimos? Si el niño es genérico y designa al niño y a la niña? Y una vez la palabra niña, ¿por qué los niños no se sintieron identificados con ella? Por qué no somos lo mismo.

(Martín, 2019)

De esta manera se refleja una profunda división en cuanto a la percepción y la representación de género en el lenguaje. Mientras que la RAE y otros defensores argumentan que el masculino genérico es inclusivo y no invisibiliza a las mujeres, varias voces críticas señalan que este enfoque perpetúa la exclusión. Básicamente, la idea de que el masculino puede abarcar a ambos géneros se cuestiona al observar que si fueran verdaderamente iguales, no habría necesidad de palabras diferentes para referirse a hombres y mujeres. Esto sugiere que el uso exclusivo del masculino como genérico se puede pensar que refuerza la idea de que los hombres son la norma y las mujeres la excepción.

En otras ocasiones, la RAE también ha sido objeto de críticas por su decisión de aceptar o rechazar nuevos términos de género opuesto en sus diccionarios.

Como indica Márquez, han surgido varias propuestas de diversas instituciones y organismos que abogan por la creación de feminismos específicos. Como respuesta a estos

intentos de intervenir activamente en el lenguaje para promover un uso no sexista, varios escritores, académicos e investigadores han tomado partido desde una perspectiva "puramente lingüística" (Márquez, 2013).

Por su parte, la Academia no solo se limita a registrar los usos del lenguaje, sino que también actúa como guía, ofreciendo recomendaciones sobre cuál debería ser el uso correcto. Estas recomendaciones no siempre son coherentes, ya que la Academia intenta alcanzar un difícil equilibrio entre la tradición y las innovaciones que se imponen debido a la adaptación del idioma a los cambios sociales.

Por ejemplo, la evolución del lenguaje ocurre de manera paralela al incremento de la presencia de mujeres en un número cada vez mayor de posiciones en el ámbito laboral. Existe una tendencia general en el español a feminizar los sustantivos que designan profesiones, especialmente aquellos que terminan en -o. Es decir, los hablantes tienden a crear formas femeninas específicas para los nombres de profesiones, incluso en casos de sustantivos cuya terminación no ha mostrado históricamente variación de género.

A veces, la Academia parece adelantarse a los hablantes en la tendencia a crear formas femeninas (por ejemplo, *árbita*, *fiscala* o *concejala*). En otras ocasiones, llama la atención su resistencia a incorporar términos que ya cuentan con una amplia aceptación en el uso cotidiano (como *abogada*). Podría argumentarse que esta demora se debe a que, hasta hace poco, las mujeres han tenido una representación limitada en la judicatura; sin embargo, esta misma ausencia se observa en otros casos, como el de *árbita*.

Esta tendencia a la creación de formas femeninas es más fuerte cuanto mayor es el prestigio de la profesión. Por eso, no ha habido problemas en aceptar un femenino como *obrero*, pero sí se ha generado controversia con términos como *médica* o *notaría*.

Por otro lado, la Academia rechazó la forma analógica *azafato*, que se había popularizado de manera espontánea en el uso cotidiano, argumentando que se trataba de una "*masculinización estéticamente aberrante, aunque posible desde el punto de vista morfológico*" (El dardo en la palabra, 1997). A pesar de que su uso es común en la lengua cotidiana, la palabra en cuestión todavía no ha sido incluida en el DRAE.

Otro caso que merece la pena mencionar es el de *modisto*. Siguiendo las normas del español, al tratarse de un sustantivo con el sufijo -ista que designa una profesión, debería ser invariable en cuanto al género. Sin embargo, *modisto* tiene una entrada independiente en el DRAE. Esto plantea la pregunta de por qué *modisto* no se considera estéticamente aberrante, mientras que *azafato* sí, mostrando una cierta incoherencia. Se tiende a pensar en factores ideológicos: el masculino *modisto* supone una elevación de la profesión de costurera, asociándolo con la alta costura y el diseño de moda. En contraste, para *azafato*, se sugiere el uso de "auxiliar de vuelo" para evitar las connotaciones peyorativas que se asocian con el femenino.

Se podría resumir que sobre todo hay dos posturas teóricas sobre el debate de lenguaje inclusivo y en consecuencia el sexismo lingüístico.

- A. Quienes defienden la independencia de la lengua con respecto a la realidad; apoyándose en el carácter arbitrario del género, afirmando que el sistema de la lengua no es sexista, aunque sí pueda ser su uso. Los autores que se sitúan en esta línea consideran que no es necesario crear femeninos específicos ni modificar en nada los usos lingüísticos.
- B. Quienes parten de la interrelación entre lengua, pensamiento y realidad, considerando al género como categoría motivada. Dada esta interrelación, se afirma que la lengua en general, y la categoría del género en particular reflejan la organización social patriarcal dominante en las sociedades actuales. Desde esta posición, se defiende la intervención activa en la lengua, herramienta que puede utilizarse para transformar la sociedad.

Ambas posturas permiten la existencia de posiciones intermedias. Los dos enfoques pueden ser útiles para justificar tanto hipótesis teóricas como acciones de política lingüística.

La defensa del sistema de la lengua como algo independiente de la realidad lleva a negar el vínculo del género gramatical con el “sexo”, y así desautotiza cualquier intervención que tienda a reflejar la presencia emergente de la mujer en la vida pública. Y por otro lado, una intervención activa en la lengua como herramienta para transformar la sociedad genera dudas sobre la manipulación del pensamiento de los hablantes y de la lengua en cuestión.

Desde mi perspectiva, no es el uso extensivo del masculino en sí mismo lo que resulta problemático, sino el sutil e implícito desplazamiento de lo específico a lo genérico, que constituye la clave de la invisibilización lingüística y social de la mujer y se deriva un uso inadecuado pragmáticamente, de términos específicos como si fueran genéricos. El lenguaje inclusivo implica ser consciente de que hablar configura la realidad. Como indica Martín (2022) “No se puede cambiar el lenguaje inclusivo con un simple cambio cosmético. No es de recibo usar una herramienta de inclusión una vez y pasar de forma inconsecuente a discriminar acto seguido. Porque se ponen en evidencia dos cosas: falta de interés o falta de formación.”

Esta posible carencia de interés o de formación es lo que motiva este trabajo. No se busca imponer reglas a los hablantes, ni limitar su forma de pensar, y mucho menos dictar cómo debería expresarse un hablante de español para no ser etiquetado como machista.

5.2. Evolución de los términos sobre la diversidad funcional

En su informe de lenguaje inclusivo, entre otras sugerencias, la RAE ha indicado que sería

conveniente que cambiara el término “disminuidos físicos” en la constitución española.

Se usa como sustantivo el término disminuido en el art. 49:

Los poderes públicos realizarán una política de previsión, tratamiento, rehabilitación e integración de los disminuidos físicos, sensoriales y psíquicos”.

Si llegaran a plantearse en algún momento otras actualizaciones de la Constitución en materia de léxico, cabría considerar la posible sustitución de esa voz por discapacitados, mucho más común en la actualidad.

(RAE, 2019)

Este cambio se llevó a cabo en 2021, cuando se implementó el término “personas con discapacidad” para referirse a este grupo de personas.

El Consejo de Ministros ha aprobado este martes el anteproyecto de reforma del artículo 49 de la Constitución Española (CE) con el fin de eliminar el término "disminuido" y permitir que las personas con discapacidad puedan participar en la sociedad en igualdad de condiciones...El nuevo artículo 49 sustituye el término "disminuidos" por el de "personas con discapacidad". Sin embargo, la iniciativa va más allá de un cambio de terminología en el lenguaje constitucional y supone "una reforma integral y en profundidad en la estructura y contenido" de la norma, ha destacado la portavoz del Gobierno, María Jesús Montero, en la rueda de prensa tras el Consejo de Ministros.”

(Radiotelevisión Española, 2021)

Se puede justificar que el cambio que se realizó finalmente en la Constitución es más adecuado que la propuesta inicial de la RAE, ya que se prefiere implementar el término “personas con discapacidad” en lugar de “discapacitado” debido a las connotaciones negativas del segundo. Esta preferencia se explorará en detalle en la sección sobre la elección de términos por parte de entidades.

(France 24, 2024)

Para la Confederación Española de Personas con Discapacidad (COCEMFE), la reforma *"incorpora el enfoque de derechos humanos, la defensa de la inclusión y vida independiente, la perspectiva de género, la interlocución con las organizaciones y el lenguaje inclusivo"*.

A diferencia de la controversia en torno al lenguaje inclusivo en términos de género, la adopción de términos más inclusivos respecto a la diversidad funcional ha suscitado menos resistencia. Aunque los cambios se logran con el tiempo y después de la insistencia de varios organismos que luchan por los derechos de las personas con discapacidad, la bibliografía científica muestra una mayor aceptación de nuevos términos relacionados con la diversidad funcional en comparación con la sustitución de palabras inclusivas respecto al género. Quizás esto se deba a una empatía generalizada hacia este grupo de personas, lo que reduce la resistencia en comparación con otros grupos como las mujeres y personas de otros géneros.

Por último, el término *diversidad funcional* es un término que también ha generado dudas sobre su utilidad. En unas guías aparece como un término aceptado, mientras en otras guías se considera un término inadecuado, dando preferencia al término *grupo colectivo de personas con discapacidad* basándose en que es la denominación oficial a nivel recogida por la Organización Mundial de la Salud (OMS). Desde un punto de vista lingüístico el término es totalmente aceptable y representa la nueva ideología para comprender la discapacidad. De esta forma, se pone el foco en la variedad de formas de funcionar que existen entre los seres humanos, sin implicar que una sea mejor o peor que otra. Al utilizar el término *diversidad funcional*, se reduce el estigma y la carga negativa. Sin embargo, sí que se desaconsejan utilizar términos como *personas con otras capacidades*, *personas con capacidades diferentes* que son muy imprecisos y no estandarizados.

En conclusión, se puede observar una evolución significativa en los términos utilizados para referirse al grupo social de personas con discapacidad; señal, que afirma que es posible interferir al lenguaje creando una evolución intencional por las autoridades y en consecuencia por los hablantes del mismo y como indica Martín: “*El lenguaje está vivo: todo lo que se construye, se puede analizar, deconstruir, y transformar*” (Martín, 2019).

6. Metodología

El lenguaje inclusivo y su evolución tienen numerosos aspectos que una herramienta computacional no puede abarcar todos. Por este motivo, se ha decidido enfocar este trabajo en los aspectos morfológicos del lenguaje escrito .

Esta sección explica los pasos seguidos para realizar los experimentos cuyo objeto es el reconocimiento de expresiones léxicas. Asimismo, se explican los criterios adoptados para la identificación de estas expresiones para la herramienta.

6.1. Estrategias para un lenguaje inclusivo con relación al género

En los medios de comunicación y en las redes sociales, frecuentemente se sugieren maneras de hablar o escribir de forma inclusiva. Sin embargo, estas sugerencias pueden ser contraproducentes, ya que aunque parezcan fáciles de aplicar, en varias ocasiones no son gramaticalmente correctas ni siempre aplicables. Es útil examinar algunas de estas prácticas comunes para explicar cuál es la mejor opción seleccionada para este proyecto.

Una práctica habitual es la implementación de "x" o "@" para sustituir la "o" masculina, transformando palabras como "todos" en "todxs" o "tod@s". El principal inconveniente de esta práctica es que hace las palabras impronunciables. Incluso cuando se usa "e", que es pronunciable, sigue siendo gramaticalmente incorrecto y puede generar generalizaciones confusas.

Otra opción por la que optan algunos hablantes es la de duplicar los sustantivos en ambos

géneros. Por ejemplo, en vez de decir “los niños”, decir “los niños y las niñas”. Esta es una idea que no es gramaticalmente incorrecta, pero después de unos sustantivos, el texto acaba siendo repetitivo y antinatural.

Para abordar este problema, hay varias estrategias que se pueden emplear:

- **Uso de términos colectivos neutros:** Una práctica común en este proyecto es encontrar alternativas que representen un conjunto sin indicar el género. Por ejemplo, en vez de decir “*los vecinos*”, se dice “*el vecindario*”.

Esta es la estrategia que se ha seleccionado para implementar en este trabajo. La eficiencia en la comunicación no se ve afectada por este cambio y se respeta la economía lingüística. Al mismo tiempo es viable computacionalmente y no presenta inconvenientes debidos a la dependencia morfológica que puede tener la frase.

- **Empleo de abstractos:** Es una práctica bastante parecida a la anterior en la cual se prefieren términos más abstractos, por ejemplo en vez de decir “*el legislador establece...*” se puede decir “*la legislación establece...*”.

La estrategia en cuestión afecta sutilmente el resultado a nivel lingüístico, haciéndolo un poco más formal, ya que el enfoque pasa de una persona a una entidad. Sin embargo, la eficiencia en la comunicación no se ve afectada por este cambio y se respeta la economía del lenguaje. Desde el punto de vista computacional, es viable implementarla, pero no se incluyó en este trabajo, ya que este tipo de cambio suele ocurrir en contextos específicos y no en situaciones cotidianas.

- **Cambio de pronombres de género por pronombres neutros:** Se pueden sustituir los pronombres que indican género por otros neutros. Por ejemplo, en lugar de “*aquellos que dispongan de pasaporte*”, se puede decir “*quienes dispongan de pasaporte*”, o cambiar “*cada uno*” por “*cada cual*”.

La estrategia en cuestión es fácil de aplicar y se utiliza con frecuencia en el día a día. Desde una perspectiva computacional, es sencilla de implementar, pero presenta un desafío relacionado con la búsqueda de corpus. Dado que términos como “*aquellos*” y “*cada uno*” no siempre se refieren a personas, encontrar un corpus adecuado se vuelve más complejo y requiere mayor atención y recursos.

- **Preferencia por sustantivos o adjetivos invariables:** Utilizar sustantivos o adjetivos que no cambian según el género. Por ejemplo, “*Pablo y Ana son muy distintos*” se puede transformar en “*Ana y Pablo son muy diferentes*”.

Aunque esta propuesta puede generar dudas por limitar la riqueza léxica, tiene sentido aclarar que se puede preferir el adjetivo *diferente* antes que *distinto* solamente cuando se refieren en personas de sexo diferente. Cuando se refiere a dos

personas de género masculino, se puede preservar el adjetivo *distintos*.

- **Desviación semántica:** Esta técnica puede ser útil para ampliar el uso del lenguaje inclusivo. Por ejemplo, la frase “Los españoles tienen una renta per cápita superior a...” puede cambiarse a “La renta per cápita en España es superior a...”.

Es una estrategia que se puede aplicar en contextos específicos. Al utilizarla, se cambia ligeramente el enfoque de la oración al modificar su sujeto. Desde una perspectiva computacional, implementar esta estrategia es más complejo mediante el reconocimiento de expresiones lingüísticas. Se podría desarrollar utilizando redes neuronales, lo cual requeriría un enfoque técnico diferente al que se ha elegido para este trabajo.

- **Utilización de sendos géneros gramaticales.** Es decir, no es una repetición nombrar en femenino y masculino cuando se representa a grupos mixtos. Utilizar por ejemplo frases como la siguiente: “*Futbolistas, tanto mujeres como hombres, se reunieron el sábado para debatir sobre el tema.*”

Para emplear esta estrategia, es necesario conocer con precisión el contexto semántico al que se refiere el hablante. Desde un enfoque computacional, es sencillo crear una herramienta que reconozca este tipo de palabras y sugiera la inclusión de las mujeres. Sin embargo, es probable que esto resulte repetitivo y no sea práctico o conveniente.

- **Estrategias sintácticas con el uso de “se” impersonal, gerundios y pasivas reflejas.** Ejemplos: “*en la prehistoria el hombre vivía en cuevas*” y “*en la prehistoria se vivía en cuevas*”, “*si los políticos trabajan adecuadamente, se puede lograr un cambio*” cambiado por “*trabajando adecuadamente desde la política se puede lograr un cambio*”, y por último “*los diputados están a punto de debatir estas cuestiones*” y “*estas cuestiones están a punto de debatirse*”.

Estas estructuras no solo favorecen la neutralidad de género, sino que también aportan un tono de formalidad y objetividad. Aunque son fácilmente comprensibles y no afectan la claridad o la eficiencia comunicativa, es importante notar que su uso puede ser más frecuente en ciertos tipos de discursos formales o académicos. Es interesante esta estrategia, pero implementarla puede resultar costoso desde el punto de vista del procesamiento del lenguaje natural (PLN), a menos que se ajuste utilizando un modelo de lenguaje grande (LLM) con técnicas de “few-shot learning”. Este enfoque permite al modelo aprender a aplicar la estrategia con solo unos pocos ejemplos, reduciendo así la necesidad de grandes cantidades de datos específicos o reglas complejas de programación. No obstante, ajustar un LLM de esta manera requiere acceso a modelos avanzados y puede implicar un uso significativo de recursos computacionales y de entrenamiento especializado.

Para concluir: “El lenguaje inclusivo no es poner todo en femenino, ni cambiar los morfemas flexivos de género. Tampoco es duplicar continuamente hasta hacer

incomprensible lo que se quiere expresar. El lenguaje inclusivo aspira a plasmar la realidad- realidad que se compone de hombres y de mujeres- y ayuda a tomar conciencia de que no nombrar a la mitad de la población perpetúa discriminaciones.” (Martin, 2019)

Como se ha mencionado previamente, en este trabajo, se ha optado por las dos primeras estrategias: el uso de términos colectivos neutros y el empleo de abstractos. Las palabras a incluir se han seleccionado a base de su frecuencia de uso y los corpus disponibles para el entrenamiento. A continuación, se muestran las palabras iniciales y sus propuestas alternativas:

Palabra inicial	Propuesta alternativa
los alumnos	el alumnado
los profesores	el profesorado
los investigadores	el equipo de investigación
los habitantes	la población
los camareros	el personal del servicio
los ingenieros	el equipo de ingeniería
los hombres de negocios	las personas de negocios
los clientes	la clientela

Tabla 1. Palabras elegidas relacionadas al género y sus alternativas. Fuente propia.

Es importante mencionar que la herramienta objeto de desarrollo en este trabajo no busca imponer una forma de hablar, sino busca dar la oportunidad que los usuarios del español reflexionen sobre su forma de estar en el mundo mediante las palabras.

6.2. Selección y justificación de expresiones para el uso del lenguaje inclusivo en relación con el grupo social de personas con discapacidad

Según la COCEMFE, “Palabras anticuadas, inexactas o inapropiadas pueden servir para perpetuar imágenes sociales negativas y estigmatizar a personas o grupos. Así mismo, palabras con significados imprecisos y/o abstractos pueden invisibilizar a las personas con discapacidad, difuminándolas con términos que persiguen obviar o negar la diferencia, pretendiendo una homogeneización de toda la sociedad que no tiene en cuenta las necesidades específicas que tienen determinados grupos que también forman parte de la sociedad y son ciudadanas y ciudadanos con derechos. Por ello, consideramos sumamente importante cuidar el lenguaje que se utiliza, en nuestro caso concreto, al referirnos a las personas con discapacidad.”

Frases no inclusivas	Frase inclusivas	Explicación
----------------------	------------------	-------------

Discapacitado(s) / Discapacitada(s)	Persona con Discapacidad	En lugar de sustantivar adjetivos, es preferible utilizar otros términos absolutos, poniendo la palabra "persona" delante. Por ello, debemos intentar de mencionar a las personas con discapacidad como "el discapacitado, la ciega, los epilépticos etc. y hacerlo de la siguiente manera: persona con discapacidad, con ceguera, con epilepsia, con tetraplejia.
Minusválido(s) / Minusválida(s)	Persona con Discapacidad	
Invalido(s) / Inválida(s)	Persona con Discapacidad	
Disminuido(s) / Disminuida(s)	Persona con Discapacidad	
Retrasado(s) / Retrasada(s)	Persona con Discapacidad	
Tetrapléjico(s) / Tetrapléjica(s)	Persona con Tetraplejia	
Sufrir discapacidad	Tener discapacidad	Evitar descripciones negativas o sensacionalistas. Es decir, una discapacidad no se sufre. Por ello, se debería eliminar las descripciones negativas o sensacionalistas de la discapacidad de una persona. No diremos "sufre de" "una víctima de", "padece de" o "persona afectada" Tampoco se deben utilizar expresiones como "postrado en una silla de ruedas", "condenado a...", sino "persona en cama" o "persona usuaria de silla de ruedas" o que "utiliza silla de ruedas".
Padece discapacidad	Tener discapacidad	

Tabla 2. Palabras elegidas relacionadas a la diversidad funcional, sus alternativas y justificación.
Fuente propia.

6.3 Desarrollo

6.3.1. Datos y herramientas disponibles

Para realizar este estudio se han recopilado varios textos de dos diferentes tipos de fuentes para tener un corpus representativo. También se ha utilizado una herramienta de anotación en línea de uso libre para la validación de las etiquetas ([spaCy NER Annotator](#)) y otras herramientas como las librerías y modelos adaptados para el reconocimiento de expresiones lingüísticas.

Se han realizado búsquedas exhaustivas para explorar las características y opciones de exportación de los corpus disponibles. Tras evaluar diversas opciones, se decidió utilizar el corpus CORPES XXI (Real Academia de la lengua, 2024) de la Real Academia Española de la Lengua para la mayoría de las etiquetas. Según la RAE, el CORPES XXI es un corpus de dominio general (incluye novelas, obras de teatro, guiones de cine, noticias de prensa, ensayos, transcripciones de noticiarios radiofónicos o televisivos, transcripciones de conversaciones, discursos etc.) de español contemporáneo de todo el mundo hispanohablante.

Asimismo, se ha recurrido al corpus esTenTen (TenTen Corpus Family, 2018) un recurso creado a partir de datos web, ya que el CORPES XXI no proporcionaba suficientes resultados para cumplir con el número requerido para la muestra.

Para asegurar un entrenamiento efectivo del modelo, se ha dividido el corpus recopilado en tres conjuntos: un conjunto de entrenamiento con 150 ocurrencias, 25 para la validación y 25 adicionales para la prueba. Además, se incluyeron fragmentos de texto que no contienen ninguna entidad, lo cual es fundamental para que el modelo aprenda a distinguir entre textos relevantes e irrelevantes y mejore su precisión en el reconocimiento de expresiones lingüísticas.

	Tokens	Type (palabras únicas)	Oraciones	Longitud media de oraciones
Entrenamiento	75992	14006	2734	27.7
Validación	13404	4075	491	27.29
Test	11044	3563	543	20.33
Total	100440	21644	3768	25.10

Tabla 3. Datos del Corpus. Fuente propia.

Palabra inicial	Corpus	Tokens	Type (palabras únicas)	Oraciones
alumnos	Corpes XXI	5888	4861	176
profesores	Corpes XXI	5863	4776	165
investigadores	Corpes XXI	4346	3683	142
habitantes	Corpes XXI	5242	4349	150
camareros	Corpes XXI	3397	2816	114
ingenieros	Corpes XXI	4334	3559	132
hombres de negocios	Corpes XXI	3644	2995	103
clientes	Corpes XXI	4531	3760	136

discapacitado(s)/discapacitada(s)	Corpes XXI	5488	4519	161
minusválido(s)/minusválida(s)	Corpes XXI	3466	2899	107
invalido(s)/inválida(s)	Corpes XXI	3989	3337	132
disminuido(s)/disminuida(s)	esTenTen	4071	3308	111
retrasado(s)/retrasada(s)	Corpes XXI	4044	3407	143
tetrapléjico(s)/tetrapléjica(s)	esTenTen	4092	3564	144
sufrir discapacidad	esTenTen	4229	3682	150
padecer discapacidad	esTenTen	3848	3287	130

Tabla 4. Datos del entrenamiento. Fuente propia.

Palabra inicial	Corpus	Tokens	Type (palabras únicas)	Oraciones
alumnos	Corpes XXI	811	711	28
profesores	Corpes XXI	706	610	26
investigadores	Corpes XXI	539	472	23
habitantes	Corpes XXI	940	747	25
camareros	Corpes XXI	530	432	20
ingenieros	Corpes XXI	724	583	24
hombres de negocios	Corpes XXI	767	621	19
clientes	Corpes XXI	695	615	25
discapacitado(s)/discapacitada(s)	Corpes XXI	954	752	24
minusválido(s)/minusválida(s)	Corpes XXI	1019	850	30
invalido(s)/inválida(s)	Corpes XXI	738	579	18
disminuido(s)/disminuida(s)	esTenTen	872	716	25
retrasado(s)/retrasada(s)	Corpes XXI	503	354	24

tetrapléjico(s)/ tetrapléjica(s)	esTenTen	709	618	25
sufrir discapacidad	esTenTen	683	610	25
padece discapacidad	esTenTen	1233	1058	42

Tabla 5. Datos de validación. Fuente propia.

Palabra inicial	Corpus	Tokens	Type (palabras únicas)	Oraciones
alumnos	Corpes XXI	840	725	32
profesores	Corpes XXI	525	454	22
investigadores	Corpes XXI	502	445	23
habitantes	Corpes XXI	550	473	24
camareros	Corpes XXI	559	489	23
ingenieros	Corpes XXI	443	385	19
hombres de negocios	Corpes XXI	444	399	22
clientes	Corpes XXI	468	424	22
discapacitado(s)/ /discapacitada(s)	Corpes XXI	629	566	26
minusválido(s)/ minusválida(s)	Corpes XXI	658	542	28
invalido(s)/ inválida(s)	Corpes XXI	605	532	24
disminuido(s)/ disminuida(s)	esTenTen	395	362	21
retrasado(s)/ retrasada(s)	Corpes XXI	597	539	25
tetrapléjico(s)/ tetrapléjica(s)	esTenTen	579	518	24
sufrir discapacidad	esTenTen	478	442	19
padece discapacidad	esTenTen	376	348	31

Tabla 6. Datos del conjunto de prueba. Fuente propia.

6.3.2. Preparación de los datos y Pre-procesamiento

Se ha procedido con la exportación del corpus en formato .txt para que se pueda procesar.

Las librerías y las herramientas utilizadas para los experimentos, se detallan a continuación:

- La librería de uso abierto spaCy (Honnibal, M., & Montani, I. 2024) junto a la herramienta spaCy NER Annotator (Murugavel, M. 2024) que permite anotar manualmente las expresiones lingüísticas que interesan y también permite entrenar un modelo de reconocimiento.
- El modelo fundacional preentrenado BERT disponible en el repositorio de HuggingFace (Hugging Face, 2025)
- Por último, el interfaz conversacional ChatGPT del modelo GPT4 (Open AI, 2025).

Proceso de anotación:

Las etiquetas se dividen en dos categorías principales:

- Etiquetas para expresiones de género (gen)
- Etiquetas para expresiones de grupos con discapacidad (disc)

Categoría	Ejemplo	Solución alternativa de lenguaje inclusivo
género	los alumnos	el alumnado
	los profesores	el profesorado
	los investigadores	el equipo de investigación
	los habitantes	la población
	los camareros	el personal del servicio
	los ingenieros	el equipo de ingeniería
	los hombres de negocios	las personas de negocios
	los clientes	la clientela
discapacidad	discapacitado/a(s)	persona con discapacidad
	minusválido/a(s)	persona con discapacidad
	invalido/a(s)	persona con discapacidad
	disminuido/a(s)	persona con discapacidad
	retrasado/a(s)	persona con discapacidad
	tetrapléjico/a(s)	personas con tetraplejia
	sufrir discapacidad	tener discapacidad

	padecer discapacidad	tener discapacidad
--	----------------------	--------------------

Tabla 7. Ejemplo del etiquetario. Fuente propia.

- Se anotaron las frases en las que, si la palabra en cuestión se reemplaza por la frase alternativa, la oración sigue teniendo sentido, y no se ve afectada por ningún cambio semántico.

Ejemplo: “Aterrizar en un asteroide requerirá de toda la pericia de los ingenieros de la NASA, en una maniobra que durará más de cinco horas antes de frenar a la nave, de 30 kilómetros por hora a sólo 10.”

En esta frase no se cambia el sentido si se reemplaza la palabra *ingenieros* por *equipo de ingeniería*.

- No se anotaron frases en las que ya preexiste la frase alternativa.

Ejemplo: “No hay que perder de vista, por otra parte, la escasa **población** de la Caracas de los años treinta, que no pasaba entonces de cien mil **habitantes**.”

Aquí, no tiene sentido reemplazar la palabra *habitantes* por *población*, ya que se crearía una repetición innecesaria.

- No se anotaron frases en las que la palabra elegida tiene otra carga semántica.

Ejemplo: “El resultado de las votaciones sorprendió tanto a Maduro que organizó una campaña tremenda denunciando fraude de la oposición por medio de una compra de votos y una situación irregular por la abundancia de votos **inválidos** registrados”

En este ejemplo no tiene sentido anotar la palabra *inválido*, ya que se refiere a votos y no a una persona.



Figura 3. Ejemplo de la herramienta spaCy NER Annotator. Fuente propia.

Una vez finalizado el proceso de anotación y validación, la herramienta mencionada, se exporta los datos en un formato que requiere procesamiento adicional para convertirlo en formato adecuado para spaCy. Por lo tanto, se ha realizado una limpieza y transformación del archivo exportado mediante código.

El proceso del desarrollo de la posible herramienta ha comenzado con la creación de un esquema inicial del flujo de trabajo previsto. Siguiendo el esquema establecido durante la fase de planificación. Este enfoque sistemático asegura una implementación más eficiente y organizada del proyecto de reconocimiento de expresiones lingüísticas.

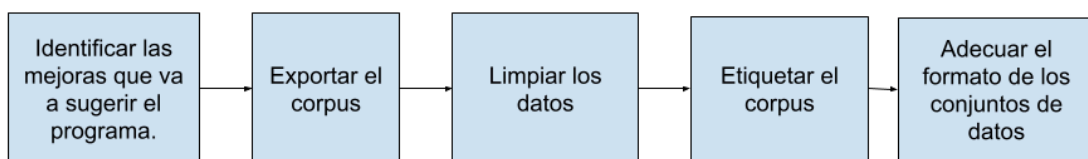


Figura 4. Esquema de preprocesamiento. Fuente propia.

6.3.3. Entrenamiento

SpaCy

Se ha utilizado spaCy para llevar a cabo el primer experimento, entrenando el modelo

mediante una función de entrenamiento sencilla, estableciendo un número fijo de iteraciones en 10. Esto significa que el modelo ha sido ajustado y refinado durante diez ciclos completos de entrenamiento, permitiendo al algoritmo mejorar su capacidad de reconocimiento y procesamiento del lenguaje natural con cada iteración. Una vez completado este proceso, la función entrenada ha sido ejecutada y sus resultados se almacenaron en la memoria para su posterior utilización.

El siguiente paso en el desarrollo ha sido evaluar la efectividad del entrenamiento realizado. Esta evaluación es fundamental para determinar si el modelo ha aprendido correctamente a identificar y procesar los patrones lingüísticos según las expectativas establecidas y para medir la precisión, y la eficiencia del modelo a nuevos datos no vistos durante el entrenamiento.

Para facilitar la interacción con los usuarios y permitir la personalización del texto, se ha creado una simple interfaz de usuario, conocida como *input frame*. Este componente solicita un texto al usuario. La interfaz ha sido diseñada para ser intuitiva, permitiendo a los usuarios seleccionar fácilmente archivos de texto o escribir directamente en un campo designado.

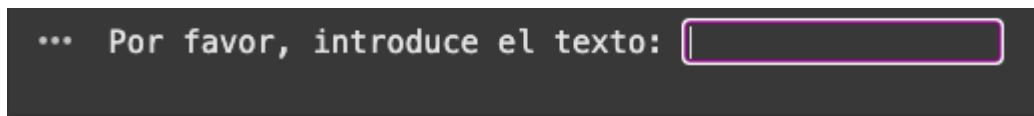


Figura 5. *Input frame*. Fuente propia.

Finalmente, se ha desarrollado una simple función adicional que proporciona sugerencias de un lenguaje más inclusivo basadas en las palabras y expresiones identificadas dentro del texto importado. Si durante el análisis no se encuentra ninguna palabra o frase que coincide con los patrones reconocidos, se responde con un mensaje adecuado, indicando que no se habían encontrado coincidencias o sugerencias relevantes. Este enfoque aseguraba que el sistema fuera útil y respondiera de manera coherente, incluso en casos donde las sugerencias no fueran aplicables.

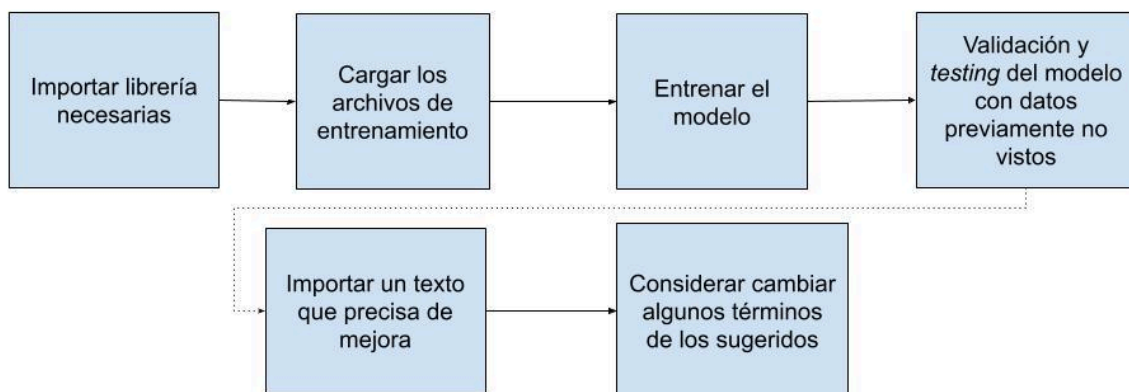


Figura 6. Esquema de entrenamiento. Fuente propia.

Modelo BERT

Para el segundo experimento, en el cual se utilizó el modelo del lenguaje preentrenado BERT ([HuggingFace](#), modelo: dccuchile/bert-base-spanish-wwm-cased). Se ha adoptado un enfoque metodológico similar al empleado previamente con spaCy. Se ha procurado minimizar las modificaciones en los datos y en el formato utilizado para el entrenamiento de spaCy, asegurando así la coherencia y la comparabilidad entre ambos experimentos.

BERT (*Bidirectional Encoder Representations from Transformers*) es un modelo de lenguaje con arquitectura Transformer. Fue desarrollado por *Google AI* en 2018, y es un avance significativo en el procesamiento del lenguaje natural, ya que debido a su preentrenamiento bidireccional tiene la capacidad para comprender el contexto de las palabras analizando tanto su relación con las palabras anteriores como con las posteriores en un texto.

En la etapa de preprocesamiento, se ha desarrollado una función con el objetivo de tokenizar los distintos conjuntos de datos utilizando el tokenizador propio del modelo. Posteriormente, se configuró el modelo para la tarea de reconocimiento de expresiones lingüísticas no inclusivas, recopilando las etiquetas únicas presentes en los datos y transformándolas en una lista ordenada con el fin de garantizar consistencia en su tratamiento.

Como se mencionó previamente, el modelo preentrenado seleccionado es la versión española 'dccuchile/bert-base-spanish-wwm-cased', debido a su adecuación para trabajar con texto en español y su libre acceso. Además, ha sido necesario convertir los textos en un formato de dataset compatible con los requisitos de entrada del modelo.

Configuración del entrenamiento

Se han ajustado parámetros clave como la tasa de aprendizaje del optimizador, la frecuencia de evaluación, el número de épocas y los mecanismos de regularización, con el fin de evitar el sobreajuste.

Parámetros de entrenamiento	
modelo	BertForTokenClassification.from_pretrained('dccuchile/bert-base-spanish-wwm-cased', num_labels=len(id2label), id2label=id2label, label2id=label2id)
train_dataset	dataset["train"]
eval_dataset	dataset["validation"]

evaluation_strategy	"epoch"
logging_steps	10
learning_rate	2e-5
per_device_train_batch_size	8
per_device_eval_batch_size	8
num_train_epochs	3
weight_decay	0.001

Tabla 8. Parámetros de entrenamiento. Fuente propia.

Al terminar el proceso del entrenamiento, se guarda el modelo para su posterior evaluación. mediante las métricas de precisión, cobertura y las medidas F1.

A continuación, se ha cargado el modelo guardado y se le pide al usuario que importe, el texto que precisa de mejora. El texto importado se tokeniza, y a través de una función de predicción se asignan las posibles etiquetas a los tokens. Por último, se muestran los resultados y las recomendaciones.

6.3.4. Retos, dificultades y excepciones

Formato de los datos

Uno de los principales desafíos identificados ha sido la adecuación del formato de una gran cantidad de datos al esquema requerido por la fase del entrenamiento. En el caso de spaCy, es necesario un formato específico, donde los índices de inicio, fin y la etiqueta de cada entidad se agrupan entre paréntesis; las listas de expresiones lingüísticas o etiquetas se organizan entre corchetes; los diccionarios de texto y anotaciones se definen entre llaves; y el texto y las etiquetas se delimitan con comillas. Aunque se ha empleado una herramienta para facilitar la anotación, fue necesario realizar un procesamiento adicional de los datos para cumplir con los requisitos formales del modelo.

En el caso de BERT, se ha buscado mantener al máximo la estructura de los datos preparada previamente para spaCy, evitando modificaciones innecesarias. No obstante, debido a las exigencias propias del modelo, fue imprescindible convertir los datos a un formato de dataset compatible para los procesos de entrenamiento y evaluación.

Validación de ejemplos

Se necesitaba una validación exhaustiva de manera manual con numerosos ejemplos para garantizar la precisión del modelo. También es importante mencionar que durante el

proceso de etiquetar se han encontrado oraciones en las que aparece la entidad lingüística entidad en cuestión pero no se puede sustituir con la entidad colectiva dado el contexto. Por ejemplo en la frase “los profesores James P. Allison y Tasuku Honjo desarrollaron las terapias de bloqueo de puntos de control inmunitarios que enfrenta la capacidad de las células cancerígenas de evitar el sistema inmunitario”. No se ha anotado la entidad “profesores” ya que efectivamente todos los profesores son de género masculino, y no tendría sentido sustituirlo por la palabra “el profesorado”.

6.4. Análisis de los resultados

6.4.1 Análisis de resultados

Antes de proceder haciendo varias pruebas se han examinado los resultados de la función de evaluación, donde se ve la evaluación de cada modelo. Los resultados obtenidos son:

Resultados spaCy	
Precisión	0,9536
Cobertura (<i>recall</i>)	0,9507
<i>F1-score</i>	0,9508
<i>F1- macro</i>	0,9507
<i>F1- micro</i>	0,9507

Tabla 9. Métricas de los resultados spaCy. Fuente propia.

Resultados BERT	
Precisión	1
Cobertura (<i>recall</i>)	1
<i>F1-score</i>	1
<i>F1- macro</i>	1
<i>F1- micro</i>	1

Tabla 10. Métricas de los resultados BERT. Fuente propia.

Este análisis proporciona una base para evaluar ambos experimentos. La precisión mide la exactitud de las predicciones del modelo, es decir, la proporción de expresiones lingüísticas correctamente identificadas entre todas las expresiones que el modelo ha identificado. La cobertura mide la capacidad del modelo para detectar todas las expresiones relevantes, es decir, la proporción de expresiones verdaderas que el modelo ha

identificado correctamente. El F1-score es una métrica que equilibra ambos aspectos de la precisión y la cobertura, y proporciona una única métrica que equilibra ambos aspectos, e indica si puede detectar expresiones lingüísticas correctas y evitar predicciones incorrectas. Finalmente, la F1 - macro mide la media de los resultados de F1 global considerando todas las etiquetas juntas, favoreciendo clases frecuentes, es decir las etiquetas que aparecen más veces en el conjunto de datos. Y la F1-macro calcula el F1 por etiqueta y luego promedia los valores, dando igual peso a todas las clases.

La comparación entre spaCy y BERT en el reconocimiento muestra resultados elevados en ambos modelos, aunque con ciertas diferencias. BERT demuestra mayor consistencia, alcanzando valores cercanos o iguales al 100% en precisión, recall y F1-score para la mayoría de las etiquetas. Por otro lado, spaCy también presenta resultados satisfactorios, aunque con una mayor variabilidad entre las etiquetas.

Por último, las etiquetas de la categoría de género están reconocidas ligeramente con mayor facilidad, probablemente debido a la mayor amplitud de corpus. Sin embargo, el desempeño en las etiquetas relacionadas de la categoría de discapacidad es también muy competitivo, destacando algunas etiquetas con un reconocimiento perfecto (como *dis6* y *vdis1*). Sin embargo, en algunos casos como la expresión “inválido” puede ser más ambigua, ya que se puede utilizar en otros contextos (Ej. un método inválido), y por esto la precisión en estos ejemplos baja a (0.81). En cambio, otras expresiones se detectan perfectamente obteniendo un grado de precisión casi perfecto debido a que no son ambiguas como el “tetrapléjico”. Esto refleja que el modelo tiene un buen balance general, aunque podría beneficiarse de un refuerzo en las etiquetas con menor recall.

6.4.2 Análisis de resultados de pruebas prácticas

Para las pruebas prácticas se han introducido algunos textos que incluyen las palabras para evaluar si el modelo entrenado ha podido identificar las expresiones en cuestión:

En una universidad, los **alumnos** se preparan para un futuro prometedor mientras asisten a clases impartidas por los **profesores**. Estos docentes colaboran a menudo con los **investigadores** para desarrollar nuevos proyectos y avanzar en diferentes campos del conocimiento. Al finalizar sus jornadas de estudio y trabajo, tanto alumnos como **investigadores** suelen visitar cafeterías locales donde los **camareros** les sirven con amabilidad. En la misma ciudad, los **habitantes** llevan a cabo sus actividades cotidianas, entre las cuales se incluyen los **ingenieros** que diseñan infraestructuras innovadoras para mejorar la calidad de vida. También se encuentran los **hombres de negocios**, quienes buscan oportunidades para invertir y expandir sus empresas. En estos encuentros, siempre están presentes los clientes, fundamentales para el éxito de cualquier empresa, quienes valoran la calidad y el servicio que reciben.

Después de ejecutarlo se observa que reconoce correctamente las expresiones lingüísticas que está entrenado a reconocer y da sugerencias siguientes:

En una universidad, ** los alumnos alu ** se preparan para un futuro prometedor mientras asisten a clases impartidas por ** los profesores profs **. Estos docentes colaboran a menudo con ** los investigadores inv ** para desarrollar nuevos proyectos y avanzar en diferentes campos del conocimiento. Al finalizar sus jornadas de estudio y trabajo, tanto alumnos como investigadores suelen visitar cafeterías locales donde ** los camareros cam ** les sirven con amabilidad. En la misma ciudad, ** los habitantes hab ** llevan a cabo sus actividades cotidianas, entre las cuales se incluyen ** los ingenieros ing ** que diseñan infraestructuras innovadoras para mejorar la calidad de vida. También se encuentran ** los hombres de negocios bus **, quienes buscan oportunidades para invertir y expandir sus empresas. En estos encuentros, siempre están presentes ** los clientes cl **, fundamentales para el éxito de cualquier empresa, quienes valoran la calidad y el servicio que reciben.

Figura 7. Ejemplo de reconocimiento de expresiones relacionadas a la inclusión de género en spaCy. Fuente propia.

```

Considere cambiar los alumnos por 'el alumnado'.
Considere cambiar los profesores por 'el profesorado'.
Considere cambiar los investigadores por 'el equipo de investigación'.
Considere cambiar los camareros por 'el personal de servicio'.
Considere cambiar los habitantes por 'la población'.
Considere cambiar los ingenieros por 'el equipo de ingeniería'.
Considere cambiar los hombres de negocios por 'las personas de negocios'.
Considere cambiar los clientes por 'la clientela'

```

Figura 8. Ejemplo I de sugerencias spaCy. Fuente propia.

```

⇒ alumnos: alu
   profesores: profs
   investigadores: inv
   alumnos: alu
   investigadores: inv
   habitantes: hab
   ingenieros: ing
   hombres: bus
   de: bus
   negocios: bus
   clientes: cl

```

Figura 9. Ejemplo de reconocimiento de expresiones relacionadas a la inclusión de género en BERT. Fuente propia.

```
Considere cambiar 'los alumnos' por 'el alumnado'.
Considere cambiar 'los profesores' por 'el profesorado'.
Considere cambiar 'los investigadores' por 'el equipo de investigación'.
Considere cambiar 'los alumnos' por 'el alumnado'.
Considere cambiar 'los investigadores' por 'el equipo de investigación'.
Considere cambiar 'los habitantes' por 'la población'.
Considere cambiar 'los ingenieríeros' por 'el equipo de ingeniería'.
Considere cambiar 'los hombres de negocios' por 'las personas de negocios'.
Considere cambiar 'los hombres de negocios' por 'las personas de negocios'.
Considere cambiar 'los hombres de negocios' por 'las personas de negocios'.
Considere cambiar 'los clientes' por 'la clientela'
```

Figura 10. Ejemplo I de sugerencias BERT. Fuente propia.

Para la prueba de la inclusión respecto la diversidad funcional se ha utilizado el siguiente texto:

En el ámbito de la atención sanitaria, es fundamental brindar apoyo adecuado a quienes enfrentan diversas condiciones de salud. Por ejemplo, el **discapacitado** puede requerir acceso a terapias de rehabilitación especializadas, mientras que el **minusválido** a menudo necesita adaptaciones en su entorno para mejorar su movilidad y calidad de vida. De manera similar, el **inválido** puede beneficiarse de dispositivos de asistencia tecnológica que le permitan realizar tareas cotidianas de forma más independiente.

Existen condiciones más específicas, como la de un **tetrapléjico**, que generalmente requiere cuidados médicos continuos y soporte especializado para manejar su situación. Además, quienes **padecen discapacidad** a veces deben enfrentar barreras en el acceso a la educación y el empleo, lo cual limita sus oportunidades de desarrollo personal y profesional.

Por otro lado, algunos individuos que **sufren discapacidad** debido a accidentes o enfermedades pueden experimentar un impacto significativo en su vida emocional y social, lo que hace necesario el apoyo psicológico para adaptarse a su nueva realidad. También, términos como **el disminuido** o **el retrasado** se utilizan en ciertos contextos médicos o educativos para describir grados específicos de limitaciones físicas o cognitivas, y sus necesidades particulares deben ser atendidas de manera personalizada para garantizar su bienestar y desarrollo integral.

En el ámbito de la atención sanitaria, es fundamental brindar apoyo adecuado a quienes enfrentan diversas condiciones de salud. Por ejemplo, el **discapacitado dis1** puede requerir acceso a terapias de rehabilitación especializadas, mientras que el **minusválido dis2** a menudo necesita adaptaciones en su entorno para mejorar su movilidad y calidad de vida. De manera similar, el **inválido dis3** puede beneficiarse de dispositivos de asistencia tecnológica que le permitan realizar tareas cotidianas de forma más independiente. Existen condiciones más específicas, como la de un **tetrapléjico dis6**, que generalmente requiere cuidados médicos continuos y soporte especializado para manejar su situación. Además, quienes **padecen discapacidad vdis2** a veces deben enfrentar barreras en el acceso a la educación y el empleo, lo cual limita sus oportunidades de desarrollo personal y profesional. Por otro lado, algunos individuos que **sufren discapacidad vdis1** debido a accidentes o enfermedades pueden experimentar un impacto significativo en su vida emocional y social, lo que hace necesario el apoyo psicológico para adaptarse a su nueva realidad. También, términos como el **disminuido dis4** o el **retrasado dis5** se utilizan en ciertos contextos médicos o educativos para describir grados específicos de limitaciones físicas o cognitivas, y sus necesidades particulares deben ser atendidas de manera personalizada para garantizar su bienestar y desarrollo integral.

Figura 11. Ejemplo de reconocimiento de expresiones relacionadas a la diversidad funcional en spaCy. Fuente propia.

Igualmente, se observa que se reconoce correctamente las expresiones lingüísticas de las que se ha entrenado y crea las sugerencias adecuadas.

```
Considere cambiar discapacitado por 'la persona con discapacidad'.
Considere cambiar minusválido por 'la persona con discapacidad'.
Considere cambiar inválido por 'la persona con discapacidad'.
Considere cambiar tetrapléjico por 'la persona con tetrapléjia'.
Considere cambiar padecen discapacidad por 'tener discapacidad'
Considere cambiar sufren discapacidad por 'tener discapacidad'
Considere cambiar disminuido por 'la persona con discapacidad'.
Considere cambiar retrasado por 'la persona con discapacidad'.
```

Figura 12. Ejemplo II de sugerencias spaCy. Fuente propia.

```
Token: discapacitado, Label: dis1
Token: minusválido, Label: dis2
Token: inválido, Label: dis3
Token: tetrapléjico, Label: dis6
Token: padecen, Label: vdis2
Token: discapacidad, Label: vdis2
Token: sufren, Label: vdis1
Token: discapacidad, Label: vdis1
Token: retrasado, Label: dis5
```

Figura 13. Ejemplo de reconocimiento de expresiones relacionadas a la diversidad funcional en BERT. Fuente propia.

```

Considere cambiar 'el discapacitado' por '(persona) con discapacidad'.
Considere cambiar 'el discapacitado' por '(persona) con discapacidad'.
Considere cambiar 'el minusválido' por '(persona) con discapacidad'.
Considere cambiar 'el minusválido' por '(persona) con discapacidad'.
Considere cambiar 'el inválido' por '(persona) con discapacidad'.
Considere cambiar 'el inválido' por '(persona) con discapacidad'.
Considere cambiar 'el inválido' por '(persona) con discapacidad'.
Considere cambiar 'el tetrapléjico' por 'la persona con tetraplégia'.
Considere cambiar 'el tetrapléjico' por 'la persona con tetraplégia'.
Considere cambiar 'el tetrapléjico' por 'la persona con tetraplégia'.
Considere cambiar 'el tetrapléjico' por 'la persona con tetraplégia'.
Considere cambiar 'padecer discapacidad' por 'tener discapacidad'.
Considere cambiar 'padecer discapacidad' por 'tener discapacidad'.
Considere cambiar 'sufir discapacidad' por 'tener discapacidad'.
Considere cambiar 'sufir discapacidad' por 'tener discapacidad'.
Considere cambiar 'el retrasado' por '(persona) con discapacidad'.

```

Figura 14. Ejemplo II de sugerencias BERT. Fuente propia.

6.4.3 Análisis y comparación de resultados con otras herramientas

Para obtener una visión más completa de la cuestión inicial planteada en este estudio, se ha optado por hacer adicionalmente una prueba empleando la técnica de ajuste mediante instrucciones "prompt-tuning". Estas instrucciones se han introducido al interfaz conversacional ChatGPT (OpenAI) del modelo GPT 4. Para esta prueba, se ha realizado un experimento introduciendo la instrucción sin proporcionar ningún ejemplo, es decir "zero-shots". También, se ha realizado otra prueba introduciendo instrucciones, pero proporcionando algunos ejemplos a seguir "few-shots".

Few-shots

El prompt inicial que se ha utilizado es el siguiente:

Eres un recomendador de formas de lenguaje más inclusivo. Vas a recibir un texto con algunas expresiones que se pueden modificar para que el lenguaje sea más inclusivo. Deberías detectar estas expresiones y sugerir una modificación. Te voy a proporcionar 10 ejemplos.

Los diez ejemplos que se han elegido como prompt son los siguientes:

Ejemplo 1.

Texto de entrada: El rector interrogó al portero, a los alumnos, a los profesores, y así se enteró de que el extravagante personaje era uno de los más destacados.

Expresiones a detectar: alumnos, profesores

Sugerencias para modificaciones: alumnado, profesorado

Ejemplo 2.

Texto de entrada: El deterioro en la educación ha llegado a límites descorazonadores, salvado por las espléndidas élites que hay, que siempre ha habido en la UNAM, de profesores, investigadores y alumnos, que no existen en tal número en otras universidades.

Expresiones a detectar: profesores, alumnos, investigadores

Sugerencias para modificaciones: profesorado, alumnado, equipo de investigación

Ejemplo 3.

Texto de entrada: Y que a los habitantes de la ciudad cualquier lugar les parecería perfecto.

Expresiones a detectar: habitantes

Sugerencias para modificaciones: población

Ejemplo 4.

Texto de entrada: Los ingenieros construyeron dos puentes sobre las aguas grises, ya a punto de congelarse y arrastrando pesados pedruscos de hielo.

Expresiones a detectar: ingenieros

Sugerencias para modificaciones: equipo de ingeniería

Ejemplo 5.

Texto de entrada: Allí no entraban los clientes a darle palique, tampoco los camareros o el resto del personal.

Expresiones a detectar: clientes, camareros

Sugerencias para modificaciones: clientela, personal de servicio

Ejemplo 6.

Texto de entrada: El problema consistía, con todo, en que los vendedores de granos exigían dinero metálico, y los hombres de negocios apenas concedían préstamos.

Expresiones a detectar: hombres de negocios

Sugerencias para modificaciones: personas de negocios

Ejemplo 7.

Texto de entrada: Protección de la familia, derechos de la mujer, derechos de los niños y adolescentes, ancianos y discapacitados, a la seguridad social, a la salud y al saneamiento ambiental.

Expresiones a detectar: discapacitados

Sugerencias para modificaciones: personas con discapacidad

Ejemplo 8.

Texto de entrada: El abatimiento era por acumulación, sin embargo, porque a la vez me daba cuenta de que nada de lo que allí desfilaba me producía, a pesar de todo, tanto ni tan dañino efecto como lo que había tenido lugar ante mis ojos sin la mediación de una pantalla, en el lavabo de los minusválidos.

Expresiones a detectar: minusválidos

Sugerencias para modificaciones: personas con discapacidad

Ejemplo 9.

Texto de entrada: El 10% de los diabéticos españoles sufre discapacidad visual.

Expresiones a detectar: sufre discapacidad
Sugerencias para modificaciones: tiene discapacidad

Ejemplo 10.

Texto de entrada: Establecer períodos de veda, particularmente en épocas de migraciones, con el fin de asegurar el desove de los animales y recuperar así las densidades poblacionales de la especie, aparentemente disminuidas.

Expresiones a detectar: -

Sugerencias para modificaciones: -

A continuación, se le proporcionó el mismo texto que se utilizó previamente en las pruebas prácticas. La respuesta de ChatGPT ha sido la siguiente:

Expresiones detectadas y sugerencias de modificaciones:

1. **los alumnos, tanto alumnos**
Sugerencia: el alumnado, **el estudiantado**
2. **los profesores**
Sugerencia: el profesorado
3. **los investigadores, tanto investigadores**
Sugerencia: el equipo de investigación
4. **los camareros**
Sugerencia: el personal de servicio
5. **los habitantes**
Sugerencia: la ciudadanía
6. **los ingenieros**
Sugerencia: **las personas ingenieras**, el equipo de ingeniería
7. **los hombres de negocios**
Sugerencia: las personas de negocios
8. **los clientes**
Sugerencia: las personas clientas, la clientela

Texto revisado:

En una universidad, el **alumnado** se prepara para un futuro prometedor mientras asiste a clases impartidas por el **profesorado**. Estas **personas docentes** colaboran a menudo con el **equipo de investigación** para desarrollar nuevos proyectos y avanzar en diferentes campos del conocimiento. Al finalizar sus jornadas de estudio y trabajo, tanto el **alumnado** como el **equipo de investigación** suelen visitar cafeterías locales donde el

personal de servicio les atiende con amabilidad. En la misma ciudad, la **ciudadanía** lleva a cabo sus actividades cotidianas, entre las cuales se incluyen **las personas ingenieras** que diseñan infraestructuras innovadoras para mejorar la calidad de vida. También se encuentran las **personas de negocios**, quienes buscan oportunidades para invertir y expandir sus empresas. En estos encuentros, siempre están presentes **las personas clientas**, fundamentales para el éxito de cualquier empresa, quienes valoran la calidad y el servicio que reciben.

Este texto ahora es más inclusivo y mantiene la coherencia en el uso del lenguaje. 😊

Se puede observar que, en la mayoría de los casos, GPT sigue las instrucciones proporcionadas en los *prompts* iniciales. No obstante, también introduce cambios que no se ajustan a las instrucciones proporcionadas por el usuario. En concreto:

Términos como *alumnos*, *profesores*, *camareros*, *hombres de negocios* e *investigadores* fueron reemplazados de manera correcta por *profesorado*, *personal de servicio*, *personas de negocios* y *equipo de investigación*, respectivamente, sin inconvenientes.

- Por otro lado, las palabras *ingenieros*, *habitantes* y *clientes* fueron sustituidas por las expresiones *las personas ingenieras*, *ciudadanía* y *personas clientas*. Cabe destacar que estos cambios no cumplen las instrucciones proporcionadas. En el caso de *ciudadanía*, aunque no se ajusta a lo solicitado, no altera de manera significativa el sentido de la frase. Sin embargo, en los casos de *las personas ingenieras* y *personas clientas*, la naturalidad del texto resulta afectada.
- Por último, se observa que el término *personas docentes* fue marcado y considerado como cambiado, a pesar de no requerir ninguna modificación ni haber sufrido alteración alguna.

Zero-shots

En la prueba de *zero-shots*, se ha elegido trabajar con el texto de las pruebas prácticas sobre diversidad funcional. El *prompt* que se proporcionó es “¿Puedes cambiar este texto para que sea más inclusivo la diversidad funcional?”, y se le proporcionó directamente el texto que se utilizó en las pruebas prácticas anteriores. El resultado es el siguiente:

En el ámbito de la atención sanitaria, es fundamental brindar apoyo adecuado a quienes enfrentan diversas condiciones de salud. Por ejemplo, las **personas con discapacidad** pueden requerir acceso a terapias de rehabilitación especializadas, mientras que quienes tienen **movilidad reducida** a menudo necesitan adaptaciones en su entorno para mejorar su calidad de vida. De manera similar, las **personas con limitaciones físicas** pueden beneficiarse de dispositivos de asistencia tecnológica que les permitan realizar tareas cotidianas de forma más independiente.

Existen condiciones más específicas, como la de una **persona con tetraplejia**, que generalmente requiere cuidados médicos continuos y soporte especializado para manejar su situación. Además, quienes **viven con alguna discapacidad** a veces deben enfrentar barreras en el acceso a la educación y el empleo, lo cual limita sus oportunidades de desarrollo personal y profesional.

Por otro lado, algunas personas que **tienen una discapacidad** debido a accidentes o enfermedades pueden experimentar un impacto significativo en su vida emocional y social, lo que hace necesario el apoyo psicológico para adaptarse a su nueva realidad. También, se utilizan términos en ciertos contextos médicos o educativos para describir grados específicos de limitaciones físicas o cognitivas, y sus necesidades particulares deben ser atendidas de manera personalizada para garantizar su bienestar y desarrollo integral.

Como ocurre con la prueba anterior, algunas expresiones lingüísticas no cambian de la misma forma.

- Las palabras *discapacitado* y *tetrapléjico* se cambian correctamente a *personas con discapacidad* y *personas con tetraplejia*.
- La palabra *minusválido* se convierte en *persona con movilidad reducida*. Aunque *movilidad reducida* es un término políticamente correcto, semánticamente no lo es en este caso, porque una persona con discapacidad no tiene necesariamente movilidad reducida.
- La palabra *inválido* se reemplaza por la frase *personas con limitaciones físicas*. Aun esa frase es correcta, como ocurre con el caso anterior, una persona con discapacidad no tiene por qué tener siempre limitaciones físicas. Hay personas con discapacidades orgánicas y a veces ni siquiera visibles.
- Los verbos *sufrir* y *padecer una discapacidad*, se han cambiado por *tener una discapacidad* y vivir con una discapacidad lo cual está correcto.
- Por último, los términos: *disminuido* y *retrasado*, se han omitido directamente.

En ambas pruebas, se observa que GPT realiza la tarea requerida con un grado considerable de éxito, aunque existen ciertos aspectos susceptibles de mejora. En el caso de la prueba de *few-shots*, resulta sorprendente que, a pesar de proporcionarle instrucciones precisas, el modelo ha efectuado modificaciones que no se le solicitaron, ajustándose en cierta medida a los patrones adquiridos durante su entrenamiento. Por otro lado, en la prueba de *zero-shots*, los resultados han sido bastante satisfactorios, si bien se han identificado algunas carencias semánticas, como en el ejemplo relacionado con la movilidad reducida.

SpaCy, BERT y la IA generativa ofrecen enfoques diferentes para tareas de reconocimiento de expresiones lingüísticas, cada uno con sus características distintivas.

En concreto, spaCy es una herramienta robusta y eficiente para tareas de procesamiento de lenguaje natural (NLP), especialmente cuando se requiere un enfoque rápido sin alto

consumo de recursos computacionales. Sus modelos preentrenados de aprendizaje supervisado de redes neuronales simples, pueden identificar expresiones con alta precisión en textos de dominios específicos, aunque sus resultados pueden bajar en casos complejos o con un contexto ambiguo, razón por la cual ha dado bajos de precisión, cobertura y F1 en algunos casos.

Por otro lado, BERT es un modelo de lenguaje basado en la arquitectura de transformers, con más capacidades para comprender el contexto de una palabra en una oración. Este modelo se entrenó sobre grandes cantidades de datos y puede capturar relaciones complejas en el lenguaje. Sin embargo, requiere más recursos computacionales.

Finalmente, la IA generativa, como los modelos GPT, son modelos que están diseñados para generar texto de forma coherente y fluida. ChatGPT puede realizar tareas de NER de manera indirecta pero flexible, generando respuestas basadas en patrones aprendidos. Su enfoque generativo lo hace adecuado para tareas que requieren una mayor comprensión y adaptación al contexto,. En este caso, el resultado dado es satisfactorio, incluso en el caso de *zero-shots*, es decir sin tener ejemplos previos de la tarea en cuestión.

7. Conclusiones

El presente trabajo subraya la importancia del lenguaje como un medio poderoso en la configuración de la realidad y la percepción social. A pesar de los esfuerzos de varios organismos en las últimas décadas para concienciar sobre el uso inclusivo del lenguaje y combatir estereotipos de género y otras formas de discriminación, persisten prácticas lingüísticas que perpetúan sesgos y desigualdades. A través del desarrollo de una herramienta que identifica esos sesgos y realiza recomendaciones. Se han realizado tres experimentos con tres tecnologías diferentes (spaCy, ajustando un modelo pre entrenado, BERT y a través de instrucciones a un interfaz de IA generativa). Esta investigación ha intentado demostrar que es posible fomentar una comunicación más inclusiva y representativa de la sociedad actual, abarcando tanto cuestiones de género como de diversidad funcional.

Los resultados obtenidos a través de la combinación de enfoques computacionales y cuantitativos pueden confirmar la eficacia de dos aproximaciones diferentes, para mejorar la inclusión en textos escritos. Sin embargo, es importante reconocer las limitaciones de estas aproximaciones incluyendo la necesidad de una mayor amplitud de datos y la adaptación a diferentes contextos lingüísticos y culturales. En este sentido, futuras investigaciones podrían centrarse en ampliar el alcance de la herramienta y en su integración con tecnologías de procesamiento de lenguaje natural más avanzadas.

En conclusión, el trabajo realizado no solo aporta una herramienta práctica para mejorar la inclusión lingüística, sino que también explora diferentes metodologías para conseguir el objetivo propuesto, y refuerza la idea de que el lenguaje inclusivo es una vía fundamental para construir una sociedad más justa y equitativa. Al cuestionar y modificar nuestras prácticas lingüísticas, se avanza hacia una comunicación que refleja verdaderamente la diversidad y la dignidad de todas las personas.

De cara al futuro, sería una buena adición ampliar el conjunto de datos empleado, integrando textos provenientes de diversas fuentes, géneros discursivos y contextos socioculturales. Esto permitiría entrenar de manera que se refleje la diversidad y la riqueza del lenguaje. Asimismo, introducir ejemplos más variados y ambiguos podría fortalecer la capacidad de los modelos a detectar sesgos lingüísticos en situaciones complejas y contextos menos explícitos.

Por último, sería interesante desarrollar funcionalidades que permitan personalizar las recomendaciones según diferentes normativas de los diversos organismos e instituciones o preferencias específicas de los usuarios, fomentando así su aplicabilidad en distintos entornos.

8. Bibliografía

- Alonso Torío, E. (2015). *El género y el sexismo en la clase de ELE*. Instituto Cervantes de Mánchester.
https://cvc.cervantes.es/ensenanza/biblioteca_ele/publicaciones_centros/PDF/manchester_2015-2016/04_alonso.pdf
- Bolívar, A. (2019). *Una introducción al análisis crítico del 'lenguaje inclusivo'* [Birmingham University].
<https://www.scielo.cl/pdf/lyl/n40/0716-5811-lyl-40-355.pdf>
- Cañete, J., Chaperon, G., Fuentes, R., Ho, J., Kang, H., & Pérez, J. (2020). *Spanish Pre-trained BERT Model and Evaluation Data*. Recuperado 25 noviembre 2024:
<https://github.com/dccuchile/beto>
- COCEMFE. (s. f.). *Manual de lenguaje Inclusivo*.
https://www.cocemfe.es/wp-content/uploads/2019/02/20181010_COCEMFE_Lenguaje_inclusivo.pdf
- Council of Europe. (2024). Guidelines for the use of language as a driver of inclusivity. Council of Europe.
<https://rm.coe.int/guidelines-for-the-use-of-language-as-a-driver-of-inclusivity/1680aec235>
- Devlin, J., Chang, M. W., & Lee, K. (s. f.). *BERT: Pre-training of deep bidirectional transformers for language understanding*. [Software]. Recuperado 26 de junio de 2024, de <https://arxiv.org/abs/1810.04805>
- Forgas i Berdet, E., Lledó, E., Guirre Romero, J., Gómez Sánchez, M. E., Igualada Belchi, D. A., Aramburuzabala, I., Rodríguez González, F., Simonis, A., & Vígara Tauste, A. M. (2009). *De igualdad y diferencias: Diez estudios de género* (A. M. Vígara Tauste, Ed.). Huerga y Fierro editores.

- García Meseguer, A. (1996). *¿Es sexista la lengua española? Una investigación sobre el género gramatical* (1. ed., 1. reimpr). Ed. Paidós.
- García Meseguer, A. (1985). Lenguaje y discriminación sexual.
- Guerrero Salazar, S. (2020). “Los dardos” de Lázaro Carreter al lenguaje de género. Textos en Proceso.
- Goddard, A. (with Patterson, L. M.). (2005). *Lenguaje y género* (1a ed). Ediciones de la Universidad de Castilla-La Mancha.
- Google Developers. (n.d.). (s. f.). *Google Drive API* [Software]. Recuperado 29 de junio de 2024, de <https://developers.google.com/drive>
- Grijelmo, A. (2019). *Propuesta de acuerdo sobre el lenguaje inclusivo: Una argumentación documentada para acercar posturas muy distantes* (Primera edición). Taurus.
- Guía de buenas prácticas contra el estigma*. (2020). Repro-Expres S.L.
- Guía de comunicación y trato inclusivo*. (2021). EDICIONES COMPLUTENSE.
- Honnibal, M., & Montani, I. (s. f.). *spaCy* [Python]. Recuperado 29 de junio de 2024, de <https://spacy.io/usage/spacy-101/>
- Hugging Face. (n.d.). bert-base-uncased. Hugging Face. Recuperado 15 de diciembre 2024, <https://huggingface.co/bert-base-uncased>
- Informe de la Real Academia Española sobre el lenguaje inclusivo y cuestiones conexas*. (s. f.). Real Academia Española. Recuperado 1 de agosto de 2024, de https://www.rae.es/sites/default/files/Informe_lenguaje_inclusivo.pdf
- Jeleton, G. (2019). *Género Neutro Y Lenguaje Inclusivo: Una Pequeña Guía Práctica* (1a ed.). Lauredal Ediciones.
- JSON. n.d. (s. f.). *Introducing JSON*. [Software]. Recuperado 10 de julio de 2024, de <https://www.json.org>
- Kilgarrieff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P.,

- & Suchomel, V. (2014). *The Sketch Engine: TenTen Corpus Family*. Lexical Computing Ltd. Recuperado 15 de diciembre 2024 de <https://www.sketchengine.eu>
- La Moncloa, C. de M. (s. f.). *Aprobado el Texto Refundido de la Ley General de Derechos de las Personas con Discapacidad y de su Inclusión Social*. Sanidad, Servicios Sociales e Igualdad. Recuperado 10 de julio de 2024, de <https://www.lamoncloa.gob.es/consejodeministros/paginas/enlaces/291113EnlaceDiscapacidad.aspx>
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., & Chen, D. (s. f.). *RoBERTa: A robustly optimized BERT pretraining approach*. [Software]. Recuperado 26 de junio de 2024, de <https://arxiv.org/abs/1907.11692>
- Márquez, M. (2013). *Género gramatical y discurso sexista*. Síntesis.
- Martín Barranco, M. (2019). *Ni por favor ni por favora: Cómo hablar con lenguaje inclusivo sin que se note (demasiado)*. Catarata.
- Martín, M. (2022). *Punto en boca: Esto no es un manual de lenguaje inclusivo*. Los Libros de la Catarata.
- Martínez, J. J. (2024, enero 27). España se une para eliminar la palabra «disminuidos» de la Constitución. *France 24*. <https://www.france24.com/es/europa/20240127-espa%C3%B1a-se-une-para-eliminar-la-palabra-disminuidos-de-la-constituci%C3%B3n-es-el-mayor-consenso-en-las-cortes-de-los-%C3%BAltimos-a%C3%B1os>
- McIntosh, C. & Cambridge University Press (Eds.). (2013). *Cambridge advanced learner's dictionary: With CD-ROM* (Fourth edition). Cambridge University Press.
- Ministerio de Sanidad, Servicios Sociales e Igualdad, I. de las mujeres. (s. f.). *GUÍAS PARA EL USO NO SEXISTA DEL LENGUAJE*.

https://www.inmujeres.gob.es/servRecursos/formacion/GuiasLengNoSexista/docs/Guiaslenguajenosexista_.pdf

Murugavel, M. (s. f.). *Spacy-ner-annotator* [Software]. Recuperado 29 de junio de 2024, de

<https://manivannanmurugavel.github.io/annotating-tool/spacy-ner-annotator/>

Naciones Unidas. (2019). *LISTA DE VERIFICACIÓN PARA USAR EL ESPAÑOL DE FORMA INCLUSIVA EN CUANTO AL GÉNERO* (p. 1).

<https://www.un.org/es/gender-inclusive-language/guidelines.shtml>

OpenAI. (2025). *ChatGPT (modelo GPT-4)*. Recuperado 5 de enero 2024, de

<https://openai.com/chatgpt>

Parlamento Europeo. (s. f.). GENDER-NEUTRAL LANGUAGE in the European Parliament.

https://www.europarl.europa.eu/cmsdata/151780/GNL_Guidelines_EN.pdf

Python Software Foundation. (n.d.). (s. f.). *ast—Abstract Syntax Trees* [Software].

Recuperado 29 de junio de 2024, de <https://docs.python.org/3/library/ast.html>

Python Software Foundation. (n.d.). (s. f.-a). *os—Miscellaneous operating system interfaces* [Software]. Recuperado 29 de junio de 2024, de

<https://docs.python.org/3/library/os.html>

Python Software Foundation. (n.d.). (s. f.-b). *random—Generate pseudo-random numbers* [Software]. Recuperado 29 de junio de 2024, de

<https://docs.python.org/3/library/random.html>

Quilis Merín, M., Albelda Marco, M., Montañez Mesas, M. P., & Carcelén Guerrero, A. (2021). *Guía de comunicación no sexista* (A. Briz Gómez, Ed.; Primera edición). Debate.

Real Academia Española. (n.d.). *Corpus Presencial del Español Actual (COPRES XII)*. Recuperado 15 de noviembre 2024, de

<https://www.rae.es/recursos/banco-de-datos/copres>

RTVE.es/AGENCIAS. (2021, mayo 11). El Gobierno inicia los trámites para eliminar el término «disminuido» de la Constitución. *RTVE*.

<https://www.rtve.es/noticias/20210511/gobierno-elimina-termino-disminuido-constitucion/2089565.shtml>

Scikit-learn developers. (n.d.). (s. f.). *Scikit-learn: Machine learning in Python*

[Software]. Recuperado 9 de septiembre de 2024, de

<https://scikit-learn.org/stable/>

Unidad de Igualdad. (s. f.). *GUÍA BÁSICA PARA UN USO DEL LENGUAJE NO SEXISTA*. Universidad Complutense.

Vasallo, B. (2021). *Lenguaje inclusivo y exclusión de clase* (Primera edición).

Larousse.

9. ANEXO:

[Repositorio](#) en GitHub dónde se encuentran las herramientas.