

# TESIS DOCTORAL

PROGRAMA DE DOCTORADO EN LENGUAS, TEXTOS Y  
CONTEXTOS

TRADUCCIÓN E INTERPRETACIÓN



UNIVERSIDAD  
DE GRANADA

*Traducción al inglés y al español de textos árabes en las Naciones  
Unidas: homogeneidad de entidades nombradas*

**Carmen Sainz Quinn**

**Director: Manuel Feria García**

**Granada, 2022**

Editor: Universidad de Granada. Tesis Doctorales  
Autor: Carmen Sainz Quinn  
ISBN: 978-84-1117-550-0  
URI: <https://hdl.handle.net/10481/77679>

# **Agradecimientos**

Quisiera dedicar unas palabras de agradecimiento a todas las personas que han contribuido a que esta tesis llegue a su fin. De manera directa e indirecta, he recibido opinión experta y consejo, pero también mucho cariño y ánimos de numerosas personas. Gracias a todos.

Muchas gracias también a toda mi familia, por confiar en mí y en mi trabajo.

Por último, agradecer enormemente al Dr. Manuel Feria García por su confianza, apoyo, ánimos y enseñanzas durante todos estos años.



# Índice de contenidos

<i>Índice de contenidos</i> .....	<i>iii</i>
<i>Índice de tablas</i> .....	<i>vii</i>
<i>Índice de gráficos</i> .....	<i>ix</i>
<i>Índice de ilustraciones</i> .....	<i>xi</i>
<i>Listado de siglas y abreviaturas</i> .....	<i>xiii</i>
<b>Capítulo I: Introducción</b> .....	<b>1</b>
<b>1. Presentación</b> .....	<b>1</b>
<b>2. Justificación y objetivos</b> .....	<b>1</b>
<b>3. Contextualización</b> .....	<b>3</b>
a) Las Naciones Unidas: sus órganos principales y sistema de protección .....	3
b) Órganos basados en la Carta.....	6
c) Órganos creados en virtud de tratados .....	9
d) Los ciclos de informes ( <i>reporting cycles</i> ) .....	12
e) El multilingüismo en las Naciones Unidas .....	15
f) Los servicios de traducción de las Naciones Unidas .....	19
<b>4. Planteamiento del problema, hipótesis, metodología y corpus</b> .....	<b>22</b>
a) Intertextualidad y homogeneidad .....	22
b) Hipótesis .....	24
c) Metodología.....	25
d) Estructura de la tesis doctoral .....	27
<b>5. Consideraciones ortotipográficas y formales</b> .....	<b>28</b>
<b>Capítulo II: Estado de la cuestión</b> .....	<b>31</b>
<b>1. Traductología</b> .....	<b>31</b>
a) Traducción institucional en organismos internacionales .....	31
b) Traducción en o para la Organización de las Naciones Unidas.....	35
c) Corpus de textos de la Organización de las Naciones Unidas .....	38
d) Homogeneidad en la traducción .....	41
<b>2. Informes, orientaciones y herramientas institucionales</b> .....	<b>45</b>
a) Informes institucionales.....	45
b) Herramientas de apoyo a la traducción para la Organización de las Naciones Unidas .....	47

c) Orientaciones para la traducción en la ONU .....	51
<b>3. Lingüística computacional .....</b>	<b>55</b>
a) Recuperación y Extracción de Información .....	55
b) Reconocimiento de Entidades Nombradas .....	56
c) Reconocimiento de Entidades Nombradas Árabes .....	57
<b>Capítulo III: Corpus y metodología .....</b>	<b>61</b>
<b>1. Corpus.....</b>	<b>61</b>
a) El corpus original.....	61
b) Selección del subcorpus.....	75
c) Selección de la muestra .....	76
d) Compilación del corpus paralelo .....	82
<b>2. Metodología .....</b>	<b>84</b>
a) Unidades de análisis .....	84
b) Extracción y organización de los datos .....	91
c) Análisis de los datos .....	96
<b>Capítulo IV: Datos y análisis de los resultados .....</b>	<b>105</b>
<b>1. Datos.....</b>	<b>105</b>
a) Análisis intratextual.....	105
b) Análisis intertextual 1 .....	109
c) Análisis intertextual 2.....	113
<b>2. Análisis de los resultados.....</b>	<b>116</b>
a) Relevancia de los resultados.....	116
b) Aportaciones al estado de la cuestión.....	120
<b>Capítulo V: Conclusiones.....</b>	<b>123</b>
<b>1. Descripción general de la investigación .....</b>	<b>123</b>
<b>2. Resumen de los capítulos.....</b>	<b>124</b>
<b>3. Refutación de las hipótesis y consecución de los objetivos .....</b>	<b>125</b>
<b>4. Principales resultados.....</b>	<b>127</b>
<b>5. Líneas futuras de investigación.....</b>	<b>128</b>
<b>Conclusiones en lengua inglesa para la obtención de la mención internacional en el título de Doctora .....</b>	<b>131</b>
<b>1. General description of the research .....</b>	<b>131</b>
<b>2. Brief description of the chapters .....</b>	<b>132</b>

<b>3. Refutation of hypotheses and achievement of objectives .....</b>	<b>133</b>
<b>4. Main results .....</b>	<b>134</b>
<b>5. Future research lines .....</b>	<b>136</b>
<i><b>Bibliografía .....</b></i>	<i><b>137</b></i>
<i><b>Anexos .....</b></i>	<i><b>159</b></i>
<b>1. Anexo 1. Organigrama del Sistema de las Naciones Unidas.....</b>	<b>161</b>
<b>2. Anexo 2: Organigrama del Departamento de la Asamblea General y de Gestión de Conferencias.....</b>	<b>162</b>
<b>3. Anexo 3. Ejemplo de la portada de un informe periódico de Estado.....</b>	<b>163</b>
<b>4. Anexo 4. Tratados internacionales y comités de expertos.....</b>	<b>167</b>





# Índice de tablas

Tabla 1: Comités de expertos, tratados y sus protocolos, año de aprobación y año de entrada en vigor .....	11
Tabla 2: Informes iniciales y periódicos de Estado del subcorpus candidatos a formar parte de la muestra, con datos desglosados por órgano receptor, número de palabras, ciclo y año de presentación .....	80
Tabla 3: Informes incluidos en la muestra .....	81
Tabla 4: Denominación simplificada de los comités e informes de la muestra .....	82
Tabla 5: Corpus paralelo, con datos desglosados por número de palabras .....	83
Tabla 6: Número de unidades de análisis, con datos desglosados por informe y categoría de EN .....	105
Tabla 7: Unidades de análisis traducidas de manera no homogénea en el análisis intratextual, con datos desglosados por informe, categoría de EN y lengua, y expresados en términos totales y porcentuales .....	107
Tabla 8: Resultados del análisis de disimilitud intratextual, con datos desglosados por lengua e informe. Entre paréntesis, desviación estándar (SD) .....	108
Tabla 9: Unidades de análisis localizadas en el análisis intertextual 1, con datos desglosados por categoría de EN .....	109
Tabla 10: Unidades de análisis traducidas de manera no homogénea en el plano intracomité, con datos desglosados por categoría, lengua y comité, expresados en términos totales y porcentuales .....	110
Tabla 11: Unidades de análisis traducidas de manera no homogénea en el plano intradocumental, con datos desglosados por categoría, lengua y por informe, expresados en términos totales y porcentuales .....	111
Tabla 12: Heterogeneidad comparada de los planos intracomité e intradocumental, con datos desglosados por lengua y categoría, y expresados en términos porcentuales y totales .....	112

Tabla 13: Resultados del análisis de la disimilitud intracomité, con datos desglosados por lengua y comité. Entre paréntesis, desviación estándar (SD).....	112
Tabla 14: Resultados de disimilitud en el plano intradocumental, con datos desglosados por lengua y categoría. Entre paréntesis, desviación estándar (SD) .....	113
Tabla 15: Número de unidades de análisis identificadas para el estudio del efecto de la variable temática, con datos desglosados por categoría y por lengua.....	114
Tabla 16: Heterogeneidad comparada de los planos intracomité e intercomité, con datos desglosados por lengua y categoría, y expresados en términos porcentuales y totales .....	114
Tabla 17: Resultados del análisis de la disimilitud en el plano intracomité e intercomité, desglosados por lengua. Entre paréntesis, desviación estándar (SD, según sus siglas en inglés) .....	116
Tabla 18: Ejemplos de paralelismos entre el español y el inglés en los informes del comité A118	
Tabla 19: Ejemplos de discrepancias con el original repetidas en la traducción al español y al inglés (resaltadas en negrita).....	118
Tabla 21: Relación completa de los tratados internacionales, y los respectivos años de aprobación y comités de expertos.....	167

# Índice de gráficos

Gráfico 1: Evolución anual del volumen de palabras traducidas del árabe en el ACNUDH entre 2000 y 2016, con datos desglosados por número de palabras y porcentaje de variación interanual .....	51
Gráfico 2: Descripción del corpus con datos desglosados por Estado, número de informes y lengua .....	66
Gráfico 3: Descripción del corpus con datos desglosados por Estado, número de palabras y lengua .....	69
Gráfico 4: Volumen de palabras por año y por lengua (barras verticales) y tendencia del uso de las lenguas (línea de puntos) .....	71
Gráfico 5: Número de informes totales y por lengua presentados a cada órgano.....	71
Gráfico 6: Distribución de palabras totales y por lengua dirigidas a cada órgano.....	73
Gráfico 7: Número de palabras por tipo de informe, con datos desglosados por lengua, en el corpus original .....	74
Gráfico 8: Número de palabras presentadas por la República del Yemen que integran el subcorpus, con datos desglosados por órgano receptor .....	78
Gráfico 9: Número de informes presentados por la República del Yemen que integran el subcorpus, con datos desglosados por órgano receptor .....	78



## Índice de ilustraciones

Ilustración 1: Flujo documental antes y después del período de sesiones del EPU.....	8
Ilustración 2: Resumen de la documentación intercambiada entre el CDESCR y la República del Yemen en el segundo ciclo de diálogo.....	15
Ilustración 3: Representación gráfica del cómputo de caracteres coincidentes para el cálculo de la distancia de Jaro entre las cadenas LEMAS y LEAMOS.....	100
Ilustración 4: Organigrama del Sistema de las Naciones Unidas .....	161
Ilustración 5: Organigrama del Departamento de la Asamblea General y de Gestión de Conferencias.....	162
Ilustración 6: Ejemplo de la portada de un informe periódico del Estado (CCPR/C/YEM/5) en su versión en español.....	163
Ilustración 7: Ejemplo de la portada de un informe del EPU (A/HRC/WG.6/32/YEM/1 en su versión en inglés.....	165



## Listado de siglas y abreviaturas

<i>Sigla / Forma abreviada</i>	<i>Forma completa</i>
ACE	Automatic Content Extraction
ACNUDH	Alto Comisionado de las Naciones Unidas para los Derechos Humanos
CAT	Comité contra la Tortura
CCPR	Comité de Derechos Humanos
CE	Comisión Europea
CED	Comité contra la Desaparición Forzada
CEDAW	Comité para la Eliminación de la Discriminación contra la Mujer
CERD	Comité para la Eliminación de la Discriminación Racial
CESCR	Comité de Derechos Económicos, Sociales y Culturales
CMW	Comité de Protección de los Derechos de Todos los Trabajadores Migratorios y de Sus Familiares
CoNLL	Conference on Computational Natural Language Learning
CRC	Comité de los Derechos del Niño
CRPD	Comité sobre los Derechos de las Personas con Discapacidad
DGACM	Departamento de la Asamblea General y de Gestión de Conferencias
DCI	Dependencia Común de Inspección
DGT	Dirección General de Traducción
DIDH	Derecho Internacional de los Derechos Humanos
DPI	Departamento de Información Pública
EI	Extracción de Información
EN	Entidad Nombrada
EPU	Examen Periódico Universal
FMI	Fondo Monetario Internacional
MUC	Message Understanding Conference
TFM	Trabajo fin de máster
OMC	Organización Mundial del Comercio
OMPI	Organización Mundial de la Propiedad Intelectual
ONU	Organización de las Naciones Unidas / la Organización
ONUG	Oficina de las Naciones Unidas en Ginebra

## Listado de siglas y abreviaturas

OTAN	Organización del Tratado del Atlántico Norte
REN	Reconocimiento de Entidades Nombradas
RENA	Reconocimiento de Entidades Nombradas Árabes
RI	Recuperación de Información
SMT	Traducción automática estadística
SPT	Subcomité para la Prevención de la Tortura y Otros Tratos o Penas Cruels, Inhumanos o Degradantes
TA	Traducción Automática
TAO	Traducción Asistida por Ordenador
UE	Unión Europea



# Capítulo I: Introducción

## 1. Presentación

La tesis doctoral titulada *Traducción del árabe al inglés y el español en las Naciones Unidas: homogeneidad en entidades nombradas* se presenta en el marco del Programa de Doctorado *Lenguas, Textos y Contextos* de la Escuela Doctoral de Humanidades, Ciencias Sociales y Jurídicas (EDHCSJ), dependiente de la Escuela Internacional de Posgrado de la Universidad de Granada. Este trabajo se adscribe a la línea de investigación «Traducción e interpretación», a la que la doctoranda accedió tras cursar el Máster Universitario de Traducción Profesional, impartido por la misma universidad, en el itinerario de traducción árabe-español. Esta tesis doctoral se realizó bajo la dirección del Doctor Manuel Feria García, Profesor Titular del Departamento de Traducción e Interpretación de la Universidad de Granada.

## 2. Justificación y objetivos

Esta investigación surgió del interés académico suscitado en el Dr. Feria por la traducción en los organismos internacionales tras su labor como traductor del árabe y el inglés al español en la Organización de las Naciones Unidas (en adelante, ONU o «la Organización»).

La investigación se ha desarrollado en tres fases: trabajo de fin de máster (en adelante, TFM); elaboración con el director de esta tesis de un capítulo de libro, y elaboración de esta tesis doctoral. A continuación se describen esas tres fases.

El TFM consistió en un estudio exploratorio en el que se testó un sistema de medición de la homogeneidad en la traducción al español y al inglés de dos categorías de entidades

nombradas (en adelante, EN): leyes, en adelante «textos normativos», y entidades corporativas nacionales árabes. Las unidades de análisis se extrajeron de una muestra procedente del corpus elaborado por Roldán y Feria (en prensa) que describiremos más adelante. La muestra consistió en informes presentados en árabe al Comité contra la Tortura (en adelante, CAT, según sus siglas en inglés).

El TFM se tituló *Coherencia en la traducción de textos árabes de la ONU: entidades corporativas y leyes* y se defendió en julio de 2015 ante el tribunal evaluador designado por la coordinación del Máster Universitario de Traducción Profesional. Los resultados demostraron la validez del sistema de medición a juicio de los miembros de ese tribunal, que animaron a la doctoranda a proseguir la investigación.

Tras ello, la doctoranda comenzó su tesis doctoral bajo la dirección del Dr. Feria. En primer lugar, se amplió la investigación empírica con una muestra mayor que la utilizada en el TFM y que incluía informes presentados en árabe al Comité de Derechos Económicos, Sociales y Culturales (en adelante, CESCR, según sus siglas en inglés) y al CAT.

Los resultados de esa segunda fase de la investigación se plasmaron en el trabajo *Translating Arabic named entities into English and Spanish: translation consistency at the United Nations* (Sainz-Quinn & Feria, 2020), que tras revisión por pares ciegos se publicó como capítulo de la monografía titulada *Routledge Handbook of Arabic Translation* (Hanna et al., 2020).

La presente tesis doctoral culmina la investigación descrita, profundiza su descripción del estado de la cuestión, amplía la muestra analizada, mejora el análisis de los datos, extrae conclusiones más sólidas, plantea nuevas perspectivas de investigación y propone posibles actuaciones institucionales encaminadas a prevenir la falta de homogeneidad en la traducción del árabe en la ONU.

Los objetivos principales de esta investigación son:

- i. Contribuir a un mejor conocimiento de la traducción institucional, en particular en organismos internacionales;
- ii. Contribuir a los estudios de traducción del árabe al español y el inglés;
- iii. Contribuir al desarrollo del procesamiento automático de la lengua árabe, y

- iv. Tender puentes entre la investigación en Traductología y el mundo profesional de la traducción.

Con todo ello, la presente investigación pretende, desde una perspectiva multidisciplinar, contribuir al desarrollo de ámbitos de investigación inexplorados en Traductología.

### **3. Contextualización**

A continuación se describe brevemente el contexto institucional en que se plantea el problema de investigación abordado en esta tesis doctoral, con particular atención a la documentación generada y las cuestiones comunicativas.

#### **a) Las Naciones Unidas: sus órganos principales y sistema de protección**

El Derecho Internacional de los Derechos Humanos (en adelante, DIDH) se define como el

s[S]istema de principios y normas [...] cuyo objeto es el fomento del respeto de los derechos humanos y libertades fundamentales universalmente reconocidos, así como el establecimiento de mecanismos para la garantía y la protección de tales derechos y libertades [...]. (Villán Durán & Faleh Pérez, 2016, p. 13)

El DIDH surge como disciplina jurídica independiente tras la Segunda Guerra Mundial y coincide con la constitución de la ONU y la aprobación de la Carta de las Naciones Unidas (San Francisco, 1945) (Bou Franch & Castillo Daudí, 2010, p. 46; Villán Durán & Faleh Pérez, 2016, p. 11), ya que

the United Nations Charter itself first gave formal and authoritative expression to the human rights regime that began at the end of the Second World War. Since its birth in 1945, the UN has served as a vital institutional spur to the development of the regime, as well as serving as a major forum for many-sided debates about it. (Alston & Goodman, 2013, p. 140)

Al amparo del DIDH se han desarrollado dos tipos de sistemas de protección (Villán Durán & Faleh Pérez, 2016, p. 12):

- Los sistemas regionales (como los dependientes del Consejo de Europa, la Organización de los Estados Americanos o la Organización para la Unidad Africana), y
- El sistema de protección de la ONU, integrado por sus mecanismos y por las dependencias subsidiarias de sus órganos principales.

En la Carta de las Naciones Unidas quedan recogidos los propósitos, principios e instituciones de la ONU. Su artículo 7 dispone que sus órganos principales son la Asamblea General, el Consejo de Seguridad, el Consejo Económico y Social, el Consejo de la Administración Fiduciaria, la Corte Internacional de Justicia y la Secretaría. A continuación los describiremos de forma muy breve:

- Asamblea General (Naciones Unidas, 1945, Capítulo IV): es el órgano representativo, normativo y deliberativo de la ONU. Los 193 Estados Miembros actuales se reúnen cada año en Nueva York para debatir y tomar decisiones o adoptar resoluciones relativas a la seguridad, la paz o la admisión de nuevos Estados Miembros;
- Consejo de Seguridad (Naciones Unidas, 1945, Capítulo V): mantiene y asegura la paz y la seguridad internacionales. Está formado por 15 miembros (5 permanentes con derecho de veto y 10 no permanentes), promueve la resolución pacífica de los conflictos y, si es el caso, impone sanciones;
- Consejo Económico y Social (Naciones Unidas, 1945, Capítulo X): trata los asuntos económicos, sociales y medioambientales y supervisa los cuerpos subsidiarios y de expertos;
- Consejo de la Administración Fiduciaria (Naciones Unidas, 1945, Capítulo XIII): supervisaba la administración de los territorios en fideicomiso para facilitar su transición al estatus de Estados soberanos, y se disolvió en noviembre de 1994 por no existir ya territorios en fideicomiso;
- Corte Internacional de Justicia (Naciones Unidas, 1945, Capítulo XIV): principal órgano judicial de la ONU, resuelve disputas legales entre los Estados y emite dictámenes, u opiniones consultivas, en relación con cualesquiera cuestiones jurídicas planteadas por los órganos de la ONU;
- Secretaría (Naciones Unidas, 1945, Capítulo XV): órgano administrativo de la ONU, su titular es el máximo representante diplomático de la Organización. Dispone de departamentos y oficinas en la Sede Central de las Naciones Unidas en Nueva York

(en adelante, la Sede), Ginebra, Nairobi y Viena, y de comisiones regionales en Bangkok, Beirut, Addis Abeba, Ginebra y Santiago de Chile.

El [Anexo 1](#) presenta los órganos principales de la ONU y sus respectivos órganos subsidiarios, departamentos y oficinas.

El sistema de protección de los derechos humanos de la ONU se basa en la Carta Internacional de Derechos Humanos. La Carta comprende los siguientes documentos:

- La Declaración Universal de Derechos Humanos (Asamblea General de las Naciones Unidas, 1948), cuyos 30 artículos definen los derechos humanos fundamentales;
- Los dos pactos internacionales (el Pacto Internacional de Derechos Civiles y Políticos<sup>1</sup> y el Pacto Internacional de Derechos Económicos, Sociales y Culturales<sup>2</sup>), y
- Los protocolos facultativos de los dos pactos internacionales

Los protocolos son documentos que complementan los tratados. Los pactos señalados disponen de un primer protocolo facultativo, en virtud del cual los órganos competentes de la ONU pueden recibir denuncias de violación de cualquiera de los derechos enunciados en sus pactos de referencia<sup>3</sup>. El Pacto Internacional de Derechos Civiles y Políticos, por su parte, dispone de un segundo protocolo facultativo<sup>4</sup> en virtud del cual los Estados se comprometen a eliminar la pena de muerte en su jurisdicción nacional. La Carta Internacional de Derechos Humanos se completa con el resto de los tratados internacionales que se detallan más adelante ([I.3.c](#), [tabla 1](#)). Todos esos documentos pueden consultarse en línea.

El principal órgano de la ONU para la promoción de los derechos humanos es la Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos (en adelante,

---

<sup>1</sup> Véase <https://www.ohchr.org/SP/ProfessionalInterest/Pages/CCPR.aspx> para consultar el Pacto Internacional de Derechos Civiles y Políticos [Última consulta: 8 de enero de 2022]

<sup>2</sup> Véase <https://www.ohchr.org/SP/ProfessionalInterest/Pages/CESCR.aspx> para consultar el Pacto Internacional de Derechos Económicos, Sociales y Culturales [Última consulta: 8 de enero de 2022]

<sup>3</sup> Véase <https://www.ohchr.org/SP/ProfessionalInterest/Pages/OPCCPR1.aspx> para el Protocolo Facultativo del Pacto Internacional de Derechos Civiles y Políticos y <https://www.ohchr.org/sp/professionalinterest/pages/opcescr.aspx> para el Protocolo Facultativo del Pacto Internacional de Derechos Económicos, Sociales y Culturales [Última consulta: 8 de enero de 2022]

<sup>4</sup> Véase <https://www.ohchr.org/SP/ProfessionalInterest/Pages/2ndOPCCPR.aspx> para consultar del Segundo Protocolo Facultativo del Pacto Internacional de Derechos Civiles y Políticos [Última consulta: 8 de enero de 2022]

ACNUDH) (Asamblea General de las Naciones Unidas, 1993). Esta entidad depende de la Secretaría, tiene su sede en Ginebra y una oficina en Nueva York.

El ACNUDH promueve los mecanismos de derechos humanos que vigilan el cumplimiento por los Estados de los tratados internacionales en que son parte (Manhire, 2018, p. 229). Entre esos mecanismos se distinguen los órganos basados en la Carta de las Naciones Unidas y los órganos creados en virtud de tratados.

## **b) Órganos basados en la Carta**

Los órganos basados en la Carta son el Consejo de Derechos Humanos, los Procedimientos Especiales y el Examen Periódico Universal.

El Consejo de Derechos Humanos es un foro político para la promoción y la protección de los derechos humanos en todo el mundo (Asamblea General de las Naciones Unidas, 2006a). Cuenta con representantes de 47 Estados<sup>5</sup> elegidos por la Asamblea en votación directa y secreta para un período de tres años. Este órgano sustituyó en 2006 a la Comisión de Derechos Humanos, en funcionamiento desde 1947.

El Consejo de Derechos Humanos desarrolla la mayor parte de su labor a través de sus mecanismos dependientes. Entre esos mecanismos conviene destacar los Procedimientos Especiales y el Grupo de Trabajo sobre el Examen Periódico Universal.

Los Procedimientos Especiales investigan la situación de los derechos humanos en el mundo y toman medidas ante las denuncias de violación de derechos civiles, culturales, económicos, sociales, políticos o de otra índole (United Nations Coordination Committee of Special Procedures, 2008). Existen dos tipos de Procedimientos Especiales: a) los temáticos, encargados de vigilar una cuestión determinada de derechos humanos en todo el mundo, y b) los dedicados a países concretos. En noviembre de 2018 desarrollaban su labor 44 mandatos temáticos y 12 mandatos de país (Manhire, 2018, p. 41).

El titular de un Procedimiento Especial se denomina «Relator Especial». Los relatores Especiales son expertos independientes que visitan los Estados, dirigen comunicaciones

---

<sup>5</sup> 13 de África, 13 de Asia, 6 de Europea Oriental, 8 de América Latina y el Caribe y 7 de Europa Occidental y otros Estados, entre otros los Estados Unidos, Australia, Canadá, Nueva Zelanda e Israel.

a los Estados y las partes interesadas, realizan estudios temáticos o regionales y organizan consultas de expertos (Manhire, 2018, p. 41). El resultado de la labor de los Relatores Especiales se plasma en sus informes. Aunque algunos Relatores tengan el árabe como lengua materna<sup>6</sup>, no parece que hasta ahora se haya presentado en árabe informe alguno de un Relator Especial<sup>7</sup>.

Desde 2006, el Consejo de Derechos Humanos cuenta también con un mecanismo denominado Examen Periódico Universal (en adelante, EPU) (Asamblea General de las Naciones Unidas, 2006a; Consejo de Derechos Humanos de las Naciones Unidas, 2007; Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos, s. f.), que proporciona de forma periódica a los 193 Estados Miembros la oportunidad de comunicar las medidas que adoptaron para mejorar la situación general de los derechos humanos (Manhire, 2018, p. 32). Este mecanismo está a cargo del llamado «Grupo de Trabajo sobre el Examen Periódico Universal», en el que se integran representantes de los 47 Estados miembros del Consejo. El carácter general de este examen es el que diferencia la revisión del EPU del seguimiento específico que llevan a cabo los comités de tratados, sobre los que volveremos más abajo (véase [I.3.c](#)).

Cada Estado Miembro es examinado por el EPU una vez cada cuatro años. Cada período de revisión del conjunto de los Estados Miembros se denomina «ciclo». Hasta la fecha se han completado dos ciclos e iniciado el tercero: 2009 a 2011, 2012 a 2016 y 2017 a 2021 (Consejo de Derechos Humanos de las Naciones Unidas, s. f.).

El examen de cada Estado en cada ciclo del EPU genera seis documentos (Sección de Traducción al Español - Paginilla, s. f.-b):

- i. Informe nacional presentado por el Estado (EPU 1), en que este expone las medidas adoptadas para proteger y promover los derechos humanos en general;
- ii. Compilación de la información contenida en los informes presentados a los órganos de tratados (EPU 2). En este documento preparado por el ACNUDH se resumen las obligaciones internacionales contraídas por el Estado, las

---

<sup>6</sup> Véase el listado de titulares de mandato actualizado a febrero de 2021 en el siguiente enlace: [https://www.ohchr.org/Documents/HRBodies/SP/VisualDirectoryFebruary2021\\_en.pdf](https://www.ohchr.org/Documents/HRBodies/SP/VisualDirectoryFebruary2021_en.pdf) [Última consulta: 2 de octubre de 2021]. Como puede comprobarse, de los 79 titulares, 2 declaran que cuentan el árabe entre sus lenguas de trabajo, y por sus nombres podría deducirse que efectivamente el árabe es su lengua materna.

<sup>7</sup> Comunicación personal de Jaime Sánchez Ratia y Manuel Fera García, correo electrónico de 11 de enero de 2019.

recomendaciones que los diversos organismos de la Organización le formularon, la cooperación del Estado con los mecanismos de derechos humanos y el cumplimiento de sus obligaciones pertinentes;

- iii. Compilación, también elaborada por el ACNUDH, de las comunicaciones presentadas por las organizaciones no gubernamentales en relación con la situación de los derechos humanos en el Estado (EPU 3);
- iv. Informe del Grupo de Trabajo sobre el EPU, que comprende recomendaciones dirigidas al Estado;
- v. Informe de respuesta del Estado examinado, y
- vi. Decisión adoptada por el Consejo de Derechos Humanos sobre el resultado del EPU.

La siguiente infografía muestra el flujo documental de dicho procedimiento:



Ilustración 1: Flujo documental antes y después del período de sesiones del EPU

Fuente: <https://www.ohchr.org/sp/hrbodies/upr/pages/uprmain.aspx> [Última consulta: 2 de octubre de 2021]

Estos documentos tienen una extensión limitada a entre 10 y 20 páginas, según el caso (Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos, s. f., p. 14; Villán Durán, 2016, p. 136). En la práctica, solo los documentos presentados por el Estado (es decir, EPU 1 y respuesta del Estado al informe del Grupo de Trabajo sobre



el EPU) podrían haberse redactado en árabe<sup>8</sup> y, por tanto, resultarían de interés para esta investigación.

### c) **Órganos creados en virtud de tratados**

Cada instrumento internacional de derechos humanos tiene asignado un órgano que se ocupa de supervisar su aplicación. Estos órganos, denominados comités, están compuestos por expertos voluntarios independientes (Alto Comisionado de las Naciones Unidas para los Derechos Humanos, s. f.). Un comité puede disponer de un subcomité para el seguimiento de uno de sus protocolos.

Los comités realizan un seguimiento periódico de la situación en cada Estado de los derechos humanos por él tutelados y le formula recomendaciones al respecto. Los comités no realizan un seguimiento general de la situación de los derechos humanos en el Estado, como hace el EPU. Por ejemplo, el Comité para la Eliminación de la Discriminación Racial (en adelante, CERD, según sus siglas en inglés) está compuesto por 18 expertos independientes y solo supervisa la aplicación de los derechos consagrados en la Convención Internacional sobre la Eliminación de Todas las Formas de Discriminación Racial (Asamblea General de las Naciones Unidas, 1965, pt. II, 8.1).

El comité mantiene un diálogo con el Estado parte de modo similar a como lo hace el Grupo de Trabajo sobre el EPU; en realidad, el segundo se inspira en los ciclos de diálogo de los comités. Sin embargo, la extensión de los documentos que los Estados presentan a los comités es considerablemente mayor y puede superar las 100 páginas, a pesar de las limitaciones estipuladas en las directrices (hasta las 80 páginas, según el tipo de informe) (Naciones Unidas, 2009, p. 7, párr. 19).

Por otro lado, mientras que el EPU funciona desde 2006, año en que se estableció el Consejo de Derechos Humanos, los comités suelen contar con una trayectoria más dilatada (Villán Durán & Faleh Pérez, 2016, p. 38). Por ejemplo, la Convención Internacional sobre la Eliminación de Todas las Formas de Discriminación Racial entró en vigor en 1969 (Asamblea General de las Naciones Unidas, 1965, art. 8) y desde entonces el comité correspondiente recibe informes de seguimiento. En algunos pocos

---

<sup>8</sup> Comunicación personal de Jaime Sánchez Ratia, correo electrónico de 11 de enero de 2019.

casos, los comités o subcomités comenzaron a celebrar sesiones después de 2006. Ejemplo de ello es el Comité sobre los Derechos de las Personas con Discapacidad (en adelante CRPD, según sus siglas en inglés), que celebró su primera sesión a finales de 2008<sup>9</sup>.

Por tanto, para formar un corpus, dispondremos de un número mayor de palabras, y con mayor variedad sincrónica y diacrónica, si utilizamos los documentos relativos a la comunicación entre los Estados y los comités que si utilizamos los relativos a la comunicación con el EPU, ya que:

- El EPU es un mecanismo único, mientras que existen 10 comités y subcomités (véase tabla 1);
- La labor de los comités suele retrotraerse en el tiempo más que la labor del EPU, y
- Los documentos presentados a los comités suelen ser más extensos que los presentados al EPU.

A efectos de compilación de un corpus de documentos de la ONU es importante, por tanto, considerar el número de tratados y protocolos existente, el año de aprobación y entrada en vigor de cada uno y el número de sus Estados partes. A continuación presentamos (tabla 1) los 10 órganos creados en virtud de tratados con esa información adicional desglosada.

Comité <sup>10</sup>	Tratado y protocolos	Año de aprobación	Año de entrada en vigor
Comité de Derechos Humanos (CCPR)	Pacto Internacional de Derechos Civiles y Políticos	Pacto: 1966	Pacto: 1976
	Protocolo Facultativo del Pacto Internacional de Derechos Civiles y Políticos	Primer protocolo facultativo: 1966	Primer protocolo facultativo: 1976
	Segundo Protocolo Facultativo del Pacto Internacional de Derechos Civiles y Políticos, destinado a abolir la pena de muerte	Segundo protocolo facultativo: 1989	Segundo protocolo facultativo: 1991
Comité de Derechos Económicos, Sociales y Culturales (CESCR)	Pacto Internacional de Derechos Económicos, Sociales y Culturales	Pacto: 1966	Pacto: 1976
	Protocolo Facultativo del Pacto Internacional de Derechos Económicos, Sociales y Culturales	Protocolo facultativo: 2008	Protocolo facultativo: 2013

<sup>9</sup> <http://www.un.org/spanish/disabilities/default.asp?navid=19&pid=1445> [Última consulta: 12 de octubre de 2020].

<sup>10</sup> Al nombre de cada comité le siguen sus siglas en inglés.

Comité para la Eliminación de la Discriminación Racial (CERD)	Convención Internacional sobre la Eliminación de Todas las Formas de Discriminación Racial	1965	1969
Comité para la Eliminación de la Discriminación contra la Mujer (CEDAW)	Convención sobre la Eliminación de Todas las Formas de Discriminación contra la Mujer	Convención 1979	Convención 1981
	Protocolo Facultativo de la Convención sobre la Eliminación de Todas las Formas de Discriminación contra la Mujer	Protocolo facultativo: 1999	Protocolo facultativo: 2000
Comité contra la Tortura (CAT)	Convención contra la Tortura y Otros Tratos o Penas Crueles, Inhumanos o Degradantes	1984	1987
Comité de los Derechos del Niño (CRC)	Convención sobre los Derechos del Niño	Convención: 1989  Protocolo facultativo 1 y 2: 2000  Tercer protocolo facultativo: 2011	Convención: 1990
	Protocolo Facultativo de la Convención sobre los Derechos del Niño relativo a la participación de niños en los conflictos armados		Protocolo facultativo 1 y 2: 2002
	Protocolo Facultativo de la Convención sobre los Derechos del Niño relativo a la venta de niños, la prostitución infantil y la utilización de niños en la pornografía		Tercer protocolo facultativo: 2014
Protocolo Facultativo de la Convención sobre los Derechos del Niño relativo a un procedimiento de comunicaciones			
Comité de Protección de los Derechos de Todos los Trabajadores Migratorios y Sus Familiares (CMW)	Convención Internacional sobre la Protección de los Derechos de Todos los Trabajadores Migratorios y de Sus Familiares	1990	2003
Comité sobre los Derechos de las Personas con Discapacidad (CRPD)	Convención Internacional sobre los Derechos de las Personas con Discapacidad	Convención: 2006	Convención: 2008
	Protocolo Facultativo de la Convención sobre los Derechos de las Personas con Discapacidad	Protocolo facultativo: 2006	Protocolo facultativo: 2008
Comité contra la Desaparición Forzada (CED)	Convención Internacional para la Protección de Todas las Personas contra las Desapariciones Forzadas	2006	2010
Subcomité para la Prevención de la Tortura y Otros Tratos o Penas Crueles, Inhumanos o Degradantes (SPT)	Protocolo Facultativo de la Convención contra la Tortura y Otros Tratos o Penas Crueles, Inhumanos o Degradantes	2002	2006

Tabla 1: Comités de expertos, tratados y sus protocolos, año de aprobación y año de entrada en vigor

Fuente: elaboración propia, basada en Órganos de Derechos Humanos, Los órganos de tratados, [www.ohchr.org/SP/HRBodies/Pages/HumanRightsBodies.aspx](http://www.ohchr.org/SP/HRBodies/Pages/HumanRightsBodies.aspx) [Última consulta: 2 de octubre de 2021]

Se pueden consultar los Estados que han ratificado las distintas convenciones y protocolos facultativos en el mapa interactivo disponible en el siguiente sitio web<sup>11</sup>: <https://indicators.ohchr.org>.

<sup>11</sup> Última consulta del sitio web realizada el 2 de octubre de 2021.

Por último, el seguimiento de la aplicación de los tratados por los Estados se completa con los procedimientos de comunicaciones o denuncias, que son mecanismos complementarios de la labor de los comités. Por ejemplo, según el artículo 1 del Protocolo Facultativo de la Convención sobre los Derechos de las Personas con Discapacidad (Asamblea General de las Naciones Unidas, 2006b):

Todo Estado Parte en el presente Protocolo ("Estado Parte") reconoce la competencia del Comité sobre los Derechos de las Personas con Discapacidad ("el Comité") para recibir y considerar las comunicaciones presentadas por personas o grupos de personas sujetos a su jurisdicción que aleguen ser víctimas de una violación por ese Estado Parte de cualquiera de las disposiciones de la Convención, o en nombre de esas personas o grupos de personas.

El artículo 3 añade:

[...] el Comité pondrá en conocimiento del Estado Parte, de forma confidencial, toda comunicación que reciba con arreglo al presente Protocolo. En un plazo de seis meses, ese Estado Parte presentará al Comité por escrito explicaciones o declaraciones en las que se aclare la cuestión y se indiquen las medidas correctivas que hubiere adoptado el Estado Parte, de haberlas.

Seguidamente, el comité examina la información recibida y puede invitar al Estado a que presente observaciones. Por motivos de espacio no conviene extendernos en los procedimientos de denuncia ya que, si bien esos mecanismos generan documentación multilingüe, los textos suelen ser confidenciales e inaccesibles para la investigación. Sí conviene, sin embargo, describir con mayor amplitud el procedimiento de comunicación entre los comités y los Estados. Eso hacemos en el apartado siguiente.

#### **d) Los ciclos de informes (*reporting cycles*)**

Como se señaló en el apartado anterior, los comités evalúan la situación de los derechos humanos en cada Estado mediante ciclos periódicos de seguimiento. El ciclo comprende habitualmente una serie de documentos previos al examen oficial del Estado y un

documento final, posterior a ese examen. Toda esa documentación es accesible en la página web del ACNUDH<sup>12</sup>, lo que permite reconstruir el ciclo de diálogo con detalle.

Los documentos previos al examen oficial del Estado son:

- Informe de Estado (*State party's report*), denominado «informe inicial de Estado» en el primer ciclo, en el que el Estado expone las medidas tomadas para aplicar el tratado;
- Documentación remitida por las organizaciones no gubernamentales, en la que estas exponen sus inquietudes y solicitan al comité que formule recomendaciones;
- Lista de cuestiones, documento en el que el comité solicita al Estado aclaraciones o ampliación de información y le proporciona «información adicional sobre el sentido de las preguntas a las que sus representantes tendrán que responder durante el examen oficial del informe» (Villán Durán y Faleh Pérez 2016, p. 46);
- Respuesta del Estado a la lista de cuestiones.

Nos interesan en particular el informe inicial de Estado y la respuesta del Estado, ya que son los documentos que pueden haber sido redactados en árabe. Esos dos documentos serían el equivalente al EPU 1 y a la respuesta del Estado al informe del Grupo de Trabajo sobre el Examen Periódico Universal.

Tras el intercambio de documentación descrita tiene lugar el examen oficial del Estado. Se trata de una sesión pública en la que el comité, generalmente con la mediación de intérpretes, examina con los representantes del Estado los documentos presentados. Los debates de esa sesión pública se publican en forma de acta resumida (*summary record*). Una vez finalizada la sesión pública, el comité emite el último documento del ciclo:

- Observaciones finales del comité (*concluding observations*), en el que el comité evalúa la situación y formula sugerencias y recomendaciones.

Seguidamente, tras un periodo determinado por el tratado, se inicia un nuevo ciclo de seguimiento mediante el envío del informe periódico de Estado. Por ejemplo, la Convención de los Derechos del Niño dispone lo siguiente:

Los Estados Partes se comprometen a presentar al Comité [de los Derechos del Niño (CRC, según sus iniciales en inglés)], por conducto del Secretario General de las

---

<sup>12</sup> <https://www.ohchr.org/EN/HRBodies/Pages/HumanRightsBodies.aspx> [Última consulta: 11 de enero de 2019].

Naciones Unidas, informes sobre las medidas que hayan adoptado para dar efecto a los derechos reconocidos en la Convención y sobre el progreso que hayan realizado en cuanto al goce de esos derechos:

- a) En el plazo de dos años a partir de la fecha en la que para cada Estado parte haya entrado en vigor la presente Convención;
- b) En lo sucesivo, cada cinco años. (Asamblea General de las Naciones Unidas, 1989, art. 44).

A modo de ejemplo, presentamos a continuación una captura de pantalla en la que se lista la documentación intercambiada en el transcurso del segundo ciclo de informes sobre la aplicación del Pacto Internacional de Derechos Económicos, Sociales y Culturales por la República del Yemen. Como puede observarse, entre esos documentos se incluye la lista de participantes en el examen oficial y el discurso original en árabe del representante del Estado. El primer ciclo y el tercer ciclo de informes también pueden abrirse mediante desplegable. La accesibilidad a toda esta información fue fundamental para la elaboración del corpus (véase [III.1](#)).

CESCR - Pacto Internacional de Derechos Económicos, Sociales y Culturales					
09 feb. 1987 (a)					
Ciclo de presentación de informes			Sesión (año)		
II			67 (2020)		
II			46 (2011)		
Tipo de documento	Símbolo/Título	Fecha de vencimiento	Fecha de entrega	Fecha de publicación	Descargar
Informe del Estado parte	E/C.12/YEM/2	30 jun. 2008	03 jul. 2008	08 dic. 2009	Visualizar el documento
Lista de cuestiones	E/C.12/YEM/Q/2			14 jun. 2010	Visualizar el documento
Respuestas a la lista de cuestiones	E/C.12/YEM/Q/2/ADD.1		28 dic. 2010	20 ene. 2011	Visualizar el documento
Info de las organizaciones de la sociedad civil	Equality Now				Visualizar el documento
Info de las organizaciones de la sociedad civil	International Disability Alliance (IDA)				Visualizar el documento
Info de las organizaciones de la sociedad civil	The All Youth Network for Society Development				Visualizar el documento
Info de las organizaciones de la sociedad civil	The Global Initiative to End All Corporal Initiative of Children				Visualizar el documento
Info de las organizaciones de la sociedad civil	The National Organization for Defending Rights and Freedoms				Visualizar el documento
Declaración					Visualizar el documento
Lista de participantes/delegación					Visualizar el documento
Observaciones finales	E/C.12/YEM/CO/2			01 jun. 2011	Visualizar el documento
Actas resumidas	E/C.12/2011/SR.14			31 may. 2011	Visualizar el documento
Actas resumidas	E/C.12/2011/SR.29			17 ene. 2012	Visualizar el documento
Actas resumidas	E/C.12/2011/SR.12			26 may. 2011	Visualizar el documento

Ilustración 2: Resumen de la documentación intercambiada entre el CESCR y la República del Yemen en el segundo ciclo de diálogo

Fuente: Reporting Status for Yemen. Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos [https://tbinternet.ohchr.org/\\_layouts/15/TreatyBodyExternal/countries.aspx?Lang=sp](https://tbinternet.ohchr.org/_layouts/15/TreatyBodyExternal/countries.aspx?Lang=sp) [Última consulta: 11 de enero de 2019].

La documentación descrita se completa en ocasiones con informes de adiciones o correcciones presentados por los Estados, que en general son de corta extensión.

### e) El multilingüismo en las Naciones Unidas

En las páginas anteriores hemos presentado algunas pistas sobre la complejidad de la comunicación multilingüe en la ONU. A continuación, y considerando el contexto institucional descrito, definiremos el contexto comunicativo que nos conducirá al problema de investigación.

La ONU es un organismo supranacional que, como se reconoce en la Resolución núm. 71/328 de la Asamblea General<sup>13</sup>, propugna el multilingüismo como valor básico para

<sup>13</sup> <http://undocs.org/es/A/RES/71/328> [Última consulta: 11 de enero de 2019].

preservar la unidad en la diversidad y fomentar la eficiencia, la transparencia, el funcionamiento de la Organización y el logro de sus propósitos.

El Secretario General es responsable de transversalizar el multilingüismo en toda la Secretaría y sus actividades y la Asamblea General, mediante resolución, solicitó la designación de un Coordinador para el multilingüismo. El Coordinador «deberá canalizar las cuestiones y preguntas planteadas por los Estados Miembros de las Naciones Unidas y las entidades de la Secretaría, facilitar un enfoque del multilingüismo coordinado, consecuente y coherente en la Secretaría, e inspirar a todos los departamentos y oficinas iniciando y proponiendo soluciones innovadoras que promuevan una cultura institucional propicia para el multilingüismo» (Asamblea General de las Naciones Unidas, 2017, p. 2). El problema de investigación que se plantea en esta tesis doctoral carecería de sentido si el marco institucional descrito fuera monolingüe.

En la ONU se distingue entre lenguas oficiales y lenguas de trabajo. Sin embargo, no existe una definición común de «lengua de trabajo» para todo el sistema de las Naciones Unidas (Dependencia Común de Inspección de las Naciones Unidas, 2020, p. v)<sup>14</sup>. La Carta de las Naciones Unidas reconoció en 1945 cinco lenguas oficiales: chino, español, francés, inglés y ruso (Naciones Unidas, 1945, Capítulo XIX, art. 111). Esas lenguas se emplean para la comunicación externa de la ONU, es decir, para la difusión de información y documentación oficial y para la recepción de documentos. Tafalla Plana

---

<sup>14</sup> En su informe más reciente, la Dependencia Común de Inspección (en adelante, DCI) insta a los jefes y jefas ejecutivos de las organizaciones del sistema de las Naciones Unidas a que, en el marco del Comité de Alto Nivel sobre Gestión de la JJE, aborden la posibilidad de adoptar una definición común de «idioma oficial», traten esta cuestión en el marco del Comité de Alto Nivel sobre Gestión de la JJE y, de ser necesario, se someta esa definición a la aprobación de los órganos legislativos o rectores de las respectivas organizaciones.

La DCI, consciente de la clara demanda de esas definiciones en el contexto del multilingüismo por la mayoría de los interlocutores implicados en el examen, elaboró las siguientes definiciones funcionales a efectos únicamente de ese informe:

Considera «idiomas oficiales» de las organizaciones del sistema de las Naciones Unidas el conjunto de idiomas seleccionados y reconocidos por los Estados Miembros como los principales idiomas de comunicación en el marco de la adopción de decisiones de los órganos legislativos o rectores de las organizaciones, utilizados para los intercambios orales y escritos entre los delegados y delegadas de los Estados Miembros, así como para la documentación obligatoria acordada, los documentos normativos y parlamentarios y las publicaciones dirigidas a ellos y a las personas de todo el planeta para las que trabajan las organizaciones.

Considera que los «idiomas de trabajo» del sistema de las Naciones Unidas son los idiomas oficiales de cada organización que se utilizan: a) para las comunicaciones internas —tanto orales como escritas— entre los funcionarios y funcionarias y los demás miembros del personal de la secretaría respectiva de cada organización en sus actividades cotidianas; y b) entre los delegados y delegadas de los Estados Miembros, en todos los casos en que las normas o prácticas no los obliguen a utilizar los demás idiomas oficiales. (Dependencia Común de Inspección de las Naciones Unidas, 2020)



(2010, p. 141) coincide con la definición utilizada por la DCI para «lenguas de trabajo», que son las empleadas para la comunicación interna en la Organización, inclusive para todas las interpretaciones, y en las cuales se redactan las actas literales y las actas resumidas (Tabory, 1980, p. 6), a las que hicimos referencia antes (véase [I.3.d](#)).

En 1973 se adoptó el árabe como lengua oficial de la Asamblea General (Asamblea General de las Naciones Unidas, 1973).

The first practical step in the adoption of the Arabic language in the General Assembly was taken in 1954, when the General Assembly decided to publish in Arabic a limited quantity of documents of the Assembly, its committees and subcommittees, together with reports of other United Nations organs dealing with problems of interest to Arabic-Speaking regions. (Tabory 1980, p. 9)

Las resoluciones 3190 (XXVIII) and 3191 (XXVIII) declararon el árabe lengua oficial y de trabajo de la Asamblea General. El árabe fue, por tanto, la primera «non-Charter language holding official and working language status in the General Assembly» (1980, p. 11) y su adopción constituyó una medida de profundo significado simbólico. Sin embargo,

el reconocimiento de un idioma como oficial implica la posibilidad de la utilización de esta lengua en todas las reuniones y conferencias convocadas por las Naciones Unidas y sus órganos y la garantía de que se haga la traducción a las otras lenguas oficiales de todos los discursos y los documentos oficiales presentados. (Tafalla Plana, 2010, p. 141)

En efecto, la medida acarreó complicaciones presupuestarias y de reclutamiento y capacitación de personal, en particular de traductores e intérpretes, que Tabory (1980) explica con detalle y se extienden hasta la actualidad. Esas complicaciones serán muy pertinentes al examinar los resultados de nuestro análisis (véase [IV](#)).

Aunque el origen de la distinción entre lenguas oficiales y lenguas de trabajo es incierto (Dependencia Común de Inspección de las Naciones Unidas, 2011, p. 7), parece que se sitúa en la Conferencia de San Francisco y sentó el precedente «for linguistic practices at the United Nations, with each organ adopting procedures regarding languages to suit its individual requirements» (Tabory, 1980, p. 7). De este modo, la distribución concreta de lenguas de trabajo y lenguas oficiales en los distintos órganos de la ONU ha ido variando

con el tiempo y puede consultarse en los informes de la Dependencia Común de Inspección de las Naciones Unidas (en adelante, DCI).

Pese a la complejidad del panorama, al menos en teoría todas las lenguas oficiales son también lenguas de trabajo en la mayor parte del sistema: en la Asamblea General, por ejemplo, la distinción desapareció a partir de diciembre de 1978 (Tabory, 1980, p. 7). En la práctica, sin embargo, se emplean como lenguas de trabajo el francés y, cada vez en mayor medida, el inglés (Tafalla Plana, 2010, p. 146). Según Guzmán y Prieto Ramos (2021, p. 4) 9 de cada 10 documentos se redactan en inglés. En esta investigación nos interesa en particular la Oficina de las Naciones Unidas en Ginebra (en adelante, ONUG), donde las lenguas de trabajo son el inglés y el francés (Oficina de las Naciones Unidas en Ginebra, s. f.-b) y donde, como veremos ([III.1.a](#)), el volumen de documentación que los Estados presentan en lengua árabe es notable.

Por otro lado, conviene distinguir entre multilingüismo y paridad lingüística. La vocación multilingüe de la Organización ha sido reiterada en multitud de ocasiones mediante resoluciones de la Asamblea General<sup>15</sup> en las que también se recordaba la falta de paridad lingüística real en la ONU. Las inquietudes en la materia movieron a que la DCI elaborara diversos informes de situación entre los que destacan los de 2003, 2011 y 2020 (Dependencia Común de Inspección de las Naciones Unidas, 2003, 2011, 2020). En los dos primeros exámenes se pone de relieve el uso hegemónico del inglés, las limitaciones de financiación, la carencia de planes estratégicos, la dispar capacidad lingüística de los agentes y los problemas relacionados con el entorno de trabajo, el acceso al empleo en la Organización y la traducción o la interpretación, entre otros la carencia de personal. Por su parte, el informe emitido en 2020 constata que casi diez años después del último examen realizado, la mayoría de los órganos del sistema de las Naciones Unidas no habían aplicado medidas para el cumplimiento de las recomendaciones recogidas en el último examen. «Por lo tanto, la mayoría de las organizaciones participantes en la DCI no cuentan actualmente con un marco estratégico que regule y proporcione orientación sobre el uso equitativo general de los idiomas y la puesta en práctica del multilingüismo

---

<sup>15</sup> La primera de ellas es la Resolución número 50/11 de la Asamblea General, de 2 de noviembre de 1995, adoptada por la Asamblea General con motivo del cincuentenario del establecimiento de la ONU, a las que siguieron entre otras las resoluciones de la Asamblea General 52/23, de 25 de noviembre de 1997, 54/64, de 6 de diciembre de 1999, 56/262, de 15 de febrero de 2002, 59/309, de 22 de junio de 2005, 61/121 B, de 14 de diciembre de 2006, 61/236, 61/244, de 22 de diciembre de 2006, 61/266, de 8 de junio de 2007, y 71/328, de 11 de septiembre de 2017. Todas ellas pueden consultarse en línea.

en la Organización» (Dependencia Común de Inspección de las Naciones Unidas, 2020, p. vi).

Lo que algunos autores denominan «unilingüismo anglófono» de la ONU (Tafalla Plana, 2010, p. 137) condujo también a la adopción de iniciativas entre los Estados, en particular en la esfera de la difusión pública de la información<sup>16</sup>. Como respuesta, la ONU formuló una serie de normas mínimas que buscan «eliminar la disparidad entre la utilización del inglés y la de los otros cinco idiomas oficiales y asegurar que todos los idiomas oficiales de las Naciones Unidas reciban un trato cabal y equitativo» (Departamento de Información Pública de las Naciones Unidas, s. f.) en relación con la difusión de información pública. Según la DCI, las limitaciones presupuestarias son la causa de la publicación exclusiva en inglés de ciertos documentos o sitios web (Dependencia Común de Inspección de las Naciones Unidas, 2011, p. 28, 2020, p. 44). En la actualidad, «la inmensa mayoría de las secretarías de las organizaciones participantes en la DCI de todo el sistema dejan en un segundo plano los demás idiomas de trabajo y privilegian la utilización del inglés en sus operaciones y programas» (2020, p. 31).

El multilingüismo es el corolario de la vocación universal de la ONU, pero su aplicación práctica no está exenta de obstáculos y genera inquietud a la Organización, a los Estados Miembros y a la literatura especializada. El problema de investigación que se plantea en esta tesis doctoral debe entenderse en ese marco y considerando que el multilingüismo solo es sostenible gracias a un esfuerzo continuado y costoso para mantener servicios de traducción e interpretación eficientes y de alto nivel. A continuación presentaremos, muy brevemente, esos servicios.

#### **f) Los servicios de traducción de las Naciones Unidas**

Los Servicios de Traducción de la ONU se integran en el Departamento de la Asamblea General y de Gestión de Conferencias (en adelante, DGACM, según sus siglas en inglés).

---

<sup>16</sup> En 2001, por ejemplo, los Estados hispanohablantes dirigieron una carta al Secretario General para solicitar el incremento del material de información pública de la Organización en español (documento A/56/50), a la que respondió el Secretario General (documento A/56/176) recordando las dificultades presupuestarias de lo solicitado. Los documentos citados pueden consultarse en <http://undocs.org/es/A/56/176> [Última consulta: 11 de enero de 2019].

El DGACM es el Departamento más grande de la Secretaría de las Naciones Unidas (véase [Anexo II](#)). Su misión consiste en

prestar servicios de reuniones y documentación de alta calidad en los idiomas oficiales de la Organización a todos los Estados Miembros en Nueva York, Ginebra, Viena, Nairobi y dondequiera que las Naciones Unidas celebren conferencias y reuniones internacionales. El Departamento de la Asamblea General y de Gestión de Conferencias desempeña un papel fundamental en el fomento del diálogo y la cooperación multilingüe entre los Estados Miembros de las Naciones Unidas, las organizaciones internacionales, los órganos intergubernamentales y la sociedad civil. (Naciones Unidas, s. f.-c)

Los Servicios de Interpretación se adscriben a la División de Reuniones y Publicaciones del DGACM y proporcionan servicios de interpretación de y a las seis lenguas oficiales. Generalmente las interpretaciones se ofrecen en la modalidad de simultánea, aunque en ocasiones también se hace interpretación consecutiva (Departamento de la Asamblea General y de Gestión de Conferencias de las Naciones Unidas, s. f.-b).

Los Servicios de Traducción, que son los que interesan en esta investigación, dependen de la División de Documentación. Gracias a su labor,

los delegados de los Estados Miembros y otros participantes en las reuniones intergubernamentales pueden acceder en el idioma que prefieran a los informes oficiales, los documentos organizativos y de trabajo y las resoluciones y decisiones, desde que se elabora el primer proyecto hasta que pasan a formar parte de una recopilación de textos aprobados. Sin esos documentos que se usan a diario, los delegados no podrían trabajar ni se podrían celebrar reuniones. (Departamento de la Asamblea General y de Gestión de Conferencias de las Naciones Unidas, s. f.-a)

La División de Documentación dispone de seis servicios de traducción, uno para cada lengua oficial, cuyo cometido es traducir

[...] todos los documentos oficiales de las Naciones Unidas, las actas de las sesiones y la correspondencia de la Sede al árabe, el chino, el español, el francés, el inglés y el ruso y algunos documentos oficiales también se traducen al alemán<sup>17</sup>. El Servicio de Traducción al Inglés y Edición también prepara las actas resumidas de las sesiones celebradas por los

---

<sup>17</sup> La ONU dispone de una Sección de Traducción al alemán. Se estableció en 1975 y está financiada a través de un fondo fiduciario al que realizan aportaciones Alemania, Austria, Liechtenstein y Suiza. Está presente únicamente en Nueva York (Asamblea General de las Naciones Unidas, 1974).

órganos intergubernamentales, en las que se hace una síntesis de las reuniones, incluidas las declaraciones realizadas y las medidas adoptadas. Una vez que se han redactado, las actas resumidas se traducen a los demás idiomas oficiales. (Departamento de la Asamblea General y de Gestión de Conferencias de las Naciones Unidas, s. f.-c)

Según la DCI (2011, pp. 27-28), «los servicios [de traducción] prestados al ACNUDH, al Consejo de Derechos Humanos y a los órganos de tratados de derechos humanos representaron el 60 % del volumen de trabajo de los servicios de traducción de la ONUG en 2010» y el volumen de trabajo generado por esas instancias aumentó el 40 % entre 2008 y 2010. El volumen de palabras traducidas del árabe también aumenta sin cesar. En el caso de lo presentado al ACNUDH de acceso no restringido, se tradujeron casi medio millón de palabras árabes en el año 2010 y entre 2013 y 2016 más de 400.000 palabras anuales<sup>18</sup>.

Para aumentar la productividad o suplir la falta de personal en ciertas lenguas, ante una demanda creciente y con recursos limitados, la ONU recurre en general a tres procedimientos: traducción por relé, traducción externa y autorrevisión. En situaciones extraordinarias, a esos procedimientos pueden sumarse otros de carácter específico<sup>19</sup>.

No disponemos de datos oficiales sobre el porcentaje de relé y autorrevisión. La traducción externa, por su parte, era responsabilidad de la Dependencia de Traducciones por Contrata, órgano actualmente extinto. Los últimos datos disponibles señalan que la traducción externa canalizaba hasta 25 % de la carga de trabajo<sup>20</sup>. Para controlar la calidad de esas traducciones, los revisores de plantilla examinan un porcentaje del texto traducido (Departamento de la Asamblea General y de Gestión de Conferencias de las Naciones Unidas, s. f.-c). La propia Organización ha reconocido que, aunque pueda afectar a la calidad, la contratación externa «se ha convertido en una fuente sistemática

---

<sup>18</sup> Manuel Fera García y Juan José Roldán Verdejo, comunicación personal, 24 de febrero de 2019. En el Capítulo 2 se describe la información proporcionada con mayor detalle.

<sup>19</sup> A modo de ejemplo, durante el confinamiento decretado a raíz de la crisis sanitaria producida por la pandemia del SARS-CoV-2, el personal temporal de la ONUG asumió tareas de apoyo imprevistas. Los intérpretes, por ejemplo, «shifted their linguistic skills from the spoken to the written form to support the translation of ad hoc documents and written materials, including the UN Geneva Annual Report 2019, COVID-19-related information, and web pages» (Oficina de las Naciones Unidas en Ginebra, 2020, p. 75).

<sup>20</sup> Dependencia de Traducciones por Contrata. Departamento de la Asamblea General y de Gestión de Conferencias. Recuperado en <http://www.un.org/es/hq/dgacm/ctu.shtml> [Última consulta: 26 de abril de 2017]

de los servicios de idiomas» (Dependencia Común de Inspección de las Naciones Unidas, 2011, pp. 31-33)<sup>21</sup>.

#### **4. Planteamiento del problema, hipótesis, metodología y corpus**

##### **a) Intertextualidad y homogeneidad**

Como cabe deducir de la exposición anterior, la labor de los servicios de traducción de la ONU es muy compleja y se desarrolla en un marco de gestión terminológica y documental muy amplio.

La traducción es solamente un eslabón de la cadena de producción y gestión de documentos, en la que intervienen muchos agentes y comprende las tres fases principales siguientes: a) redacción, edición y aprobación del texto en el idioma original por las dependencias que se elaboran; b) traducción, revisión y procesos conexos; y c) impresión y distribución dentro de los plazos prescritos. (Dependencia Común de Inspección de las Naciones Unidas, 2011, p. 21)

En esta investigación nos interesa en particular la comunicación en los ciclos de seguimiento que se describieron en la sección [I.3.d](#). En el marco de un solo comité, el proceso abarca los siguientes actos de traducción e interpretación cuando el Estado se expresa en lengua árabe <sup>22</sup>:

- i. El informe inicial de Estado en árabe se traduce al resto de las lenguas oficiales en los servicios de traducción de la ONUG o mediante traducción externa. La traducción se realiza directamente desde el árabe o mediante relé. Hasta la fecha, en las secciones de ruso y chino siempre se traduce del árabe mediante relé, ya que carecen de traductores de árabe;
- ii. La documentación remitida por las organizaciones no gubernamentales generalmente se traduce del inglés (en ocasiones del francés) al resto de lenguas oficiales, inclusive el árabe;
- iii. La lista de cuestiones se traduce del inglés al resto de lenguas oficiales, inclusive el árabe;

---

<sup>21</sup> También la Comisión Europea externaliza hasta un 25 % del trabajo para reducir costes (Comisión Europea, 2012, p. 9)

<sup>22</sup> La descripción que sigue es una reconstrucción del Dr. Manuel Feria.

- iv. La respuesta del Estado se traduce del árabe al resto de lenguas, al igual que el resto de los documentos adicionales enviados por el Estado, en las condiciones descritas en i.;
- v. En el examen público del Estado median intérpretes y se emplean usualmente el árabe, lengua de los representantes del Estado, el francés, el inglés y cualesquiera otras lenguas oficiales solicitadas;
- vi. Partiendo de esas interpretaciones se redactan en inglés las actas resumidas, que posteriormente son traducidas al resto de los idiomas oficiales, inclusive al árabe;
- vii. Las observaciones finales del comité se traducen del inglés al resto de lenguas oficiales, inclusive el árabe.

En el ciclo subsiguiente, el proceso se reinicia con un nuevo informe nacional en árabe (informe periódico de Estado) en el que el Estado alude a las conclusiones y las recomendaciones que el comité formuló en inglés en las observaciones finales al ciclo anterior.

Este proceso se repite en paralelo en diferentes comités y en el EPU. Recordemos que en la ONU hay 193 Estados Miembros y 2 Estados observadores<sup>23</sup> y que, por tanto, el conjunto documental acumulado es formidable y aumenta a un ritmo vertiginoso.

Por otro lado, los conjuntos documentales relativos a un solo Estado conforman una unidad y mantienen relaciones complejas de intertextualidad interna (ya que unos documentos del Estado remiten, mediante citas expresas u ocultas, a otros documentos de ese Estado y a los tratados a los que se da seguimiento) y externa (ya que remiten, en particular, a la legislación nacional del Estado).

Esta situación exige un alto nivel de coordinación entre los agentes implicados en el proceso comunicativo. Los textos normativos y las entidades corporativas son elementos muy repetidos en los informes iniciales y periódicos de Estado, como veremos más adelante, y, por tanto, también en su traducción se espera la máxima coordinación y homogeneidad.

---

<sup>23</sup> Los dos Estados observadores no miembros de la ONU son la Santa Sede y el Estado de Palestina, desde 1964 y 2012 (A/RES/67/19), respectivamente. Participan como observadores en los períodos de sesiones y trabajos de la Asamblea General, pero no tienen derecho a voto (Naciones Unidas, s. f.-b).

La importancia de esos elementos es tal que los tratados los citan de forma expresa cuando determinan el contenido deseable de los informes. A modo de ejemplo, las directrices pertinentes disponen que los informes iniciales y periódicos de los Estados dirigidos al CCPR deben

abordar específicamente cada uno de los artículos de las partes I, II y III del Pacto; no bastará con describir las normas jurídicas: deberán explicarse la situación de hecho y la disponibilidad práctica, los efectos y el ejercicio de los recursos en caso de violación de los derechos del Pacto. (Naciones Unidas, 2009, p. 46)

Para describir las normas jurídicas es imprescindible nombrarlas y para describir las situaciones de hecho lo es citar las entidades corporativas nacionales que hacen efectivo el derecho humanitario. Esos elementos son cambiantes y es ese devenir y su adecuación a los propósitos del derecho humanitario lo que los comités vigilan.

La falta de homogeneidad en la traducción de esos y otros elementos podría dificultar la vigilancia del cumplimiento de las obligaciones humanitarias que incumben a los Estados, así como la recuperación de información (en adelante, RI), la extracción de información (en adelante, EI) y, en particular, el reconocimiento de entidades nombradas (en adelante, REN) en esos conjuntos documentales.

## **b) Hipótesis**

A la vista del contexto y del problema descritos, en la presente investigación se plantea la siguiente hipótesis nula principal:

*Hipótesis principal:* En las traducciones del árabe al español y el inglés realizadas en o para la ONU no se registra falta de homogeneidad en las denominaciones de los textos normativos y las entidades corporativas nacionales.

De refutarse la hipótesis nula principal se plantearían dos hipótesis nulas secundarias:

*Hipótesis secundaria 1:* En la traducción de los informes dirigidos a un mismo comité, esa falta de homogeneidad no es mayor en la traducción de varios documentos que en la traducción de un solo documento.



*Hipótesis secundaria 2:* Esa falta de homogeneidad no es mayor en la traducción de los informes dirigidos a varios comités que en la traducción de los informes dirigidos a un solo comité.

La hipótesis secundaria 1 apunta a la variable diacrónica y su posible efecto puesto que, por la dinámica descrita de los ciclos de seguimiento ([I.3.d](#)), entre un informe y otro dirigido al mismo comité habrá distancia temporal pero no diferencias esenciales de temática o intertextualidad.

La hipótesis secundaria 2 no apunta necesariamente a la variable diacrónica, puesto que los ciclos de seguimiento de dos comités pueden ser sincrónicos. Esta hipótesis apunta, por tanto, a la potencial influencia sobre la homogeneidad del cambio de temática y de referentes intertextuales.

### **c) Metodología**

Esta investigación de carácter empírico y cuantitativo (Hernández Sampieri et al., 2006) tiene como objetivo medir el grado de homogeneidad en la traducción como producto. Para ello estaba previsto utilizar metodología de corpus para la compilación del corpus, metodología de REN para la extracción automática de las unidades de análisis y metodología de lingüística computacional para el análisis de los datos.

El corpus original, compilado por Roldán y en el que se basan Roldán y Fera (en prensa), estaba compuesto por los informes dirigidos por los Estados árabes entre 2000 y 2018 al Grupo de Trabajo sobre el Examen Periódico Universal y a los comités de tratados. Se trata de un corpus trilingüe (árabe, francés e inglés) de unos 9 millones de palabras ([III.1.a](#)). Puesto que esta investigación se centra en la traducción del árabe, se descartaron los informes presentados en inglés y francés. El resultado fue un subcorpus ([III.1.b](#)) de aproximadamente 7 millones de palabras formado exclusivamente por originales árabes.

Las unidades de análisis elegidas eran entidades nombradas (en adelante, EN; véase [III.2.a](#)). Se eligieron las EN para tal propósito porque como nombres propios, por definición, deben traducirse de forma homogénea. En esta investigación se utilizan dos categorías de EN: textos normativos y entidades corporativas nacionales. La categoría «texto normativo» no se contempla en las taxonomías actuales de EN (véase [II.3.b](#)), por

lo que su determinación como EN debe subrayarse como aporte de esta investigación. La categoría «entidad corporativa nacional» constituye una subcategoría de «organización».

Durante la fase exploratoria (elaboración del TFM) estaba previsto emplear para la extracción automática de esas unidades de análisis metodología y herramientas informáticas de REN. Sin embargo, ninguna herramienta de las sondeadas parecía permitir la extracción automatizada de nuestras unidades de análisis. Por ello, y considerando la dificultad que supone el procesamiento de la lengua árabe, se procedió a extraer una muestra suficiente para un análisis sólido.

Puesto que la hipótesis principal solo podría refutarse considerando como universo el Estado individual, ya que una misma unidad de análisis no tendría por qué traducirse de forma homogénea para dos Estados diferentes, el primer criterio de selección de la muestra consistió en identificar el Estado con mayor presencia en términos cuantitativos (número de palabras). Ese Estado es la República del Yemen ([III.1.c.i](#)).

Para seleccionar los informes yemeníes más válidos para el análisis fue preciso aplicar una serie de criterios relacionados con las hipótesis general y secundarias 1 y 2, respectivamente:

- i. Extensión: Los documentos elegidos debían tener la mayor extensión posible para que se garantizara la disponibilidad de suficientes unidades de análisis;
- ii. Variable diacrónica: En igualdad de condiciones debían priorizarse informes separados por el mayor tiempo posible;
- iii. Variable temática y referencial: Debían priorizarse los informes dirigidos a comités diferentes.

Tras la aplicación de los criterios señalados se eligieron 10 documentos dirigidos a 5 comités (2 documentos por comité)<sup>24</sup>, lo que sumó 357 509 palabras originales árabes.

---

<sup>24</sup> Comité contra la Tortura: CAT/C/16/Add.10 y CAT/C/YEM/2; Comité de Derechos Humanos: CCPR/C/YEM/2001/3 y CCPR/C/YEM/5; Comité para la Eliminación de la Discriminación Racial: CERD/C/362/Add.8 y CERD/C/YEM/17-18; Comité de los Derechos Económicos, Sociales y Culturales: E/1990/5/Add.54 y E/C.12/YEM/3; Comité de los Derechos del Niño: CRC/C/129/Add.2 y CRC/C/YEM/4.

Por último, se localizaron las traducciones oficiales de esos informes al español y al inglés (368 948 palabras en español y 328 112 palabras en inglés) y se formó un corpus paralelo de 1 054 569 palabras (véase [III.1.d](#)).

Del corpus paralelo se extrajeron de manera prácticamente manual las unidades de análisis. Para analizar los datos se cuantificó:

- el grado de heterogeneidad en la traducción en español e inglés de las EN;
- el grado de disimilitud en español e inglés, mediante la fórmula de Jaro, un algoritmo bien conocido en lingüística computacional que permite medir la distancia de edición entre dos cadenas de caracteres.

El análisis midió la homogeneidad intratextual y el posible impacto de las potenciales variables diacrónica y temática (véase [III.2.c.ii](#), [-iii](#) y [-iv](#)).

Las conclusiones extraídas de los análisis demostraron la viabilidad de la metodología. Téngase en cuenta que las repercusiones de una diferencia computacional como la descrita no tienen necesariamente un correlato en términos de significación; dicho de otro modo, una falta de homogeneidad mayor en términos computacionales no implica una diferencia de significado mayor, o viceversa. Por ejemplo, «Ley de los menores» y «Ley de las menores» implica una falta de homogeneidad mínima en términos computacionales, pero quizás máxima en términos de significado. Este inconveniente es ineludible en términos computacionales, que son los únicos aplicables a la RI, la EI y el REN.

Hasta donde se nos alcanza, no existen investigaciones previas que midan la homogeneidad en traducción en estos términos computacionales concretos. También consideramos muy novedoso la aportación de conclusiones empíricas a la investigación sobre la traducción en organismos internacionales y, en general, sobre la traducción del árabe.

#### **d) Estructura de la tesis doctoral**

Esta tesis doctoral abarca cinco capítulos y una sección de referencias:

- El Capítulo I («Introducción») justifica y describe la investigación;

- El Capítulo II («Estado de la cuestión») ofrece una panorámica sobre las áreas de investigación relacionadas;
- El Capítulo III («Corpus y metodología») describe el corpus, justifica la selección de la muestra y detalla la metodología aplicada;
- El Capítulo IV («Datos y análisis de los resultados») expone los resultados obtenidos y analiza los datos en relación con el estado de la cuestión;
- El Capítulo V («Conclusiones») extrae conclusiones y presenta vías futuras de investigación.

Tras esos cinco capítulos y las referencias bibliográficas figura un resumen y conclusiones en lengua inglesa, ya que esta tesis aspira a la mención internacional en el título de Doctora, así como varios anexos.

## **5. Consideraciones ortotipográficas y formales**

Cuando en esta tesis doctoral se haga referencia a cuestiones propias de la ONU (en particular nombres de mecanismos o tratados y nombres de Estado), se empleará la recomendación de UNTERM en caso de discrepancia ortotipográfica entre esta y las normas de la Real Academia Española de la Lengua. En particular, conviene señalar lo siguiente:

- Estados Miembros son solo los de la Asamblea General de las Naciones Unidas;
- En esta tesis se emplea «Yemen» exclusivamente como forma abreviada de «República del Yemen». Para «República Árabe de Yemen» o «República Democrática Popular de Yemen» utilizaremos estas denominaciones, sin hacer uso de forma abreviada alguna;
- Se emplea Estado(s) parte(s), salvo en los nombres propios y en citas literales de tratados cuando su original porte mayúscula inicial en el adjetivo.

En el cuerpo del texto en grafía latina se utiliza en punto 12; en las citas de extensión superior a 40 palabras, punto 11; en las notas a pie de página, punto 10; y en las tablas, los gráficos y sus respectivos títulos, punto 9.

En el cuerpo del texto en grafía árabe se emplea *Traditional Arabic* punto 16 y en las

tablas, punto 12. Los ejemplos aparecen en grafía árabe, transcritos y acompañados de traducción literal entre corchetes, seguidos de traducción oficial al inglés y al español en columnas aparte. Las transliteraciones se realizaron con el programa desarrollado por Medea-García (2017), que proporciona las transliteraciones de manera automática de acuerdo con las convenciones establecidas por la ISO 233: 1984 de la Organización Internacional de Normalización.

En las referencias bibliográficas e inserción de citas en texto se siguen las normas de la American Psychological Association (APA).



## Capítulo II: Estado de la cuestión

La interdisciplinariedad de esta investigación exige abordar el estado de la cuestión en los estudios de traducción, la literatura institucional y la lingüística computacional. En primer lugar, describimos y analizamos la literatura sobre la traducción en los organismos supranacionales, los avances académicos e institucionales en la compilación y el análisis de corpórea de textos de la ONU y la literatura sobre la homogeneidad en traducción. A continuación, recopilamos la información pertinente de los informes, herramientas de apoyo a la traducción y orientaciones de la propia Organización. Seguidamente describimos los principales avances en relación con la recuperación y extracción de información, en particular el reconocimiento de entidades nombradas árabes. En cada sección se pone de relieve nuestra originalidad y aportaciones. No dedicamos un apartado a los estudios de corpus y a los sistemas de medición de similitud y disimilitud de cadenas de caracteres, dos ámbitos fundamentales para esta investigación, por economía de espacio, ya que la metodología pertinente (III.2) es bien conocida y no comprende avances recientes.

### 1. Traductología

#### a) Traducción institucional en organismos internacionales

Kang, con una definición redundante, afirma que la traducción institucional consiste en «translating in or for a specific organisation» (2009, p. 141). Koskinen, traductora del finés en la Comisión Europea (en adelante, CE) añade que

[...] we are dealing with institutional translation in those cases when an official body (government agency, multinational organization or a private company, etc.; also an individual person acting in an official status) uses translation as a means of ‘speaking’ to

a particular audience. Thus, in institutional translation, the voice that is to be heard is that of the translating institution. As a result, in a constructivist sense, the institution itself gets translated. (2008, p. 22)

Koskinen afirmó unos años después que la traducción institucional es un tipo de «autotraducción» porque «institutions produce documents, and in multilingual contexts these documents also need to be translated. Significantly, in institutional translation, the institution is typically the author of both the source text and its translation(s). Thus, institutional translation is self-translation» (2011, p. 57). Para la autora, los entornos de la traducción institucional son: i) servicios públicos nacionales (*public services*), que afrontan los retos lingüísticos generados por los flujos migratorios; ii) organismos nacionales bilingües o multilingües, y iii) organismos supranacionales multilingües (2008, pp. 28-32).

A nuestro juicio, la descripción de la traducción institucional como «autotraducción» no puede aplicarse al entorno i) puesto que los documentos los produce por definición un Estado diferente al Estado de acogida que promueve la traducción y pueden reflejar criterios tan distintos a los del segundo que incluso se discute en ocasiones su constitucionalidad y aplicabilidad. En España, por ejemplo, solo con un concepto amplio de orden público en la doctrina se aceptan los repudios (Cervilla Garzón, 2018, p. 152). Nada más opuesto, pues, a la autotraducción. En ii), sin embargo, la traducción institucional sí es por definición una forma de autotraducción.

En cualquier caso, aquí nos interesa el entorno iii). Entre los organismos supranacionales multilingües destacan la ONU, la Unión Europea (en adelante, UE) (24 lenguas oficiales), el Fondo Monetario Internacional (en adelante, FMI) (8 lenguas<sup>25</sup>) y la Organización del Tratado del Atlántico Norte (inglés y francés [North Atlantic Treaty Organization, 2010]).

Antes de avanzar en la literatura sobre traducción institucional, nos interesa hacer un paréntesis y abordar la posible equiparación de la traducción en sus dos principales entornos institucionales: la ONU y la UE.

La ONU y la UE tienen numerosos puntos pertinentes de coincidencia: ambas procuran equilibrar la observancia del multilingüismo con sus cargas financieras y cuentan entre

---

<sup>25</sup> Alemán, árabe, chino, español, francés, inglés, portugués y ruso. <https://www.imf.org/external/np/adm/rec/job/langsvc.htm> [Última consulta: 27 de junio de 2018]



sus idiomas oficiales lenguas semíticas (árabe en la ONU y maltés en la UE) y de otras familias no indoeuropeas (lenguas urálicas en la UE y sinotibetanas en la ONU, en concreto chino), lo que implica problemas similares de reclutamiento de personal y retos ligados al empleo de sistemas de escritura no latinos (alefato, alfabeto cirílico o pictogramas).

La UE, sin embargo, reconoce 24 lenguas oficiales, frente a las 6 de la ONU, lo que implica 552 combinaciones lingüísticas. En términos cuantitativos, por tanto, el multilingüismo es un fenómeno mucho más complejo en la UE que en la ONU. Esto explica que desde que en 2004 se adoptaron como lenguas oficiales el checo, el eslovaco, el esloveno, el estonio, el húngaro, el letón, el lituano, el maltés y el polaco se utilice un sistema de lenguas pivote consistente en traducir primero al inglés, el francés o el alemán, y de esa lengua a las demás<sup>26</sup>. En la ONU, sin embargo, solo se recurre al relé en casos imprescindibles y nunca como política oficial.

Por otro lado, los Estados miembros de la UE están unidos por lazos culturales, económicos, jurídicos y políticos mucho más estrechos que los que unen a los 193 Estados Miembros de la ONU, cuya diversidad es abrumadora. En este sentido, la traducción en la ONU es más compleja. El reto que supone la traducción del árabe en la ONU, por el número de Estados, la tradición jurídica diferente y el volumen de documentación, tampoco es equiparable a la traducción del maltés en la UE. Resulta difícil, pues, extrapolar de manera directa las conclusiones de la literatura previa sobre la UE a nuestra investigación.

Tras este breve análisis comparativo, conviene que consideremos la adecuación de la definición de Koskinen al entorno iii). La información que figura en la literatura institucional, que revela que el 68 % de la documentación traducida en la ONUG es externa, y el corpus compilado por Roldán (véase [II.1.c](#)) demuestran que no necesariamente es así. Si lo fuera, nuestra investigación carecería de sentido, puesto que el árabe no es lengua de trabajo y las EN se adscribirían a la propia Organización, por lo que difícilmente no figurarían en las herramientas de traducción a disposición de los traductores para la ONU y difícilmente no se traducirían de forma homogénea.

---

<sup>26</sup> [http://www.europarl.europa.eu/about-parliament/files/organisation-and-rules/multilingualism/es-ep\\_translators.pdf](http://www.europarl.europa.eu/about-parliament/files/organisation-and-rules/multilingualism/es-ep_translators.pdf) [Última consulta: 11 de enero de 2019].

En base a lo que precede, en el entorno iii) de Koskinen (2008), el organismo se autotraduce unas veces y otras, traduce documentación externa. En la ONU, que es lo que aquí nos interesa, el mecanismo de seguimiento y diálogo entre los comités y los Estados (I.3) se basa justamente en que se traduzcan «para la Organización» informes nacionales que no son «de la Organización». En este sentido, un elemento fundamental de distinción en el seno de un único organismo supranacional que no parece tener en cuenta Koskinen es la combinación lingüística y su direccionalidad. En la ONU, la traducción del inglés al español puede ser una forma de autotraducción (así es en la Sede), pero nunca lo es cuando se traduce del árabe al español.

A finales del decenio de 1990, Koskinen insistía en la escasa atención prestada a la traducción institucional y en la orientación puramente jurídica de la literatura (2000, p. 50). Hoy, más de veinte años después, esa misma afirmación sigue siendo cierta (Prieto Ramos, 2018, pp. 1-2). A continuación presentamos en orden cronológico el estado de la cuestión sobre la traducción en esos organismos, con exclusión de la investigación centrada en la traducción jurídica. La literatura específicamente sobre la ONU se aborda en el apartado siguiente.

Koskinen (2000) analiza el multilingüismo, el proceso de traducción y la figura del traductor en la UE y critica la escasa atención que la traducción recibe en la institución. La autora afirma que «translations are not really translations, but language versions» (p. 54) y que «texts are written and translated by the Commission» (p. 60). Es revelador, por ejemplo, que se invisibilizara la traducción en una campaña europea por la claridad en el lenguaje jurídico<sup>27</sup>. La institución no valora la traducción como parte fundamental del proceso comunicativo ni su potencial para mejorar la calidad del diálogo (p. 54).

Posteriormente, Koskinen propone la definición de «traducción institucional» que ya hemos analizado (2008, p. 22), profundiza en cuestiones de sociología de la traducción y autoría, examina el entorno laboral y compara los cambios observados en los borradores, las versiones finales y las dos traducciones al finés de la Comunicación COM(2001)678 para determinar los factores institucionales o interpersonales que generaron esos cambios.

Wagner, Bech & Martínez (2002, 2012), quienes describen el multilingüismo en las instituciones de la UE, recuerdan que solo se traducen a todas las lenguas oficiales los

---

<sup>27</sup> *Fight the FOG campaign*. Se inició en 1998 por traductores de la CE (Wagner et al., 2012, pp. 72-75).

documentos procedentes de las instituciones europeas, ya que los remitidos por los ciudadanos o los Estados miembros se traducen exclusivamente al inglés y el francés (2012, pp. 9-10). Esta última afirmación refuerza la noción de que no toda la traducción institucional es una forma de «autotraducción».

En particular, Wagner, Bech y Martínez describen los servicios de traducción de la UE, el porcentaje de externalización, la contratación y los perfiles disponibles, así como los problemas que afronta el proceso, como la baja calidad de los originales, la falta de claridad del lenguaje empleado en la CE y los ajustados plazos de entrega, que se justifican por la convicción de que «an imperfect translation delivered on time is much better than a perfect one delivered too late» (2012, p. 80).

Las últimas aportaciones en la materia se han producido fundamentalmente en el seno del proyecto LETRINT (*Legal Translation in International Institutional Settings: Scope, Strategies and Quality Markers*)<sup>28</sup>, que está dirigido por Fernando Prieto Ramos. Este proyecto tiene como objetivo profundizar en el estudio de la traducción jurídica e institucional desde un punto de vista empírico. Sus estudios se centran fundamentalmente en la homogeneidad terminológica como indicador de la calidad de la traducción jurídica en entornos de producción textual multilingüe, en particular en la ONU, la UE y la OMC (Guzmán & Prieto Ramos, 2021; Prieto Ramos, 2020; Prieto Ramos & Guzmán, 2018, 2021).

A continuación exponemos las principales contribuciones académicas sobre la traducción en o para las Naciones Unidas.

## **b) Traducción en o para la Organización de las Naciones Unidas**

Cao y Zhao (2008) (el segundo, traductor en la ONUG) subrayan la escasa atención prestada a la traducción en la Organización, («despite the long history of translation and multilingual practice at the UN, very little has been studied and written as to the nature and difficulties of translating documents at the UN in translation studies» (p. 39)) y su carácter institucional («all documents at the UN are produced in accordance with an institutional system of standards and criteria» (p. 44)). Los autores describen el proceso

---

<sup>28</sup> <https://transius.unige.ch/en/research/letrint>

documental, el reclutamiento y las dificultades que afrontan los traductores: «technical demands, consistency, working as a team, turn over time (short deadlines), and the need to keep abreast of world changes, which are invariably reflected in the discussions at the UN» (p. 52) y «[D]ifficulties due to the multi-racial and multilingual characteristics of UN work» (p. 46). Nos interesa aquí en particular su insistencia en la necesidad de preservar la homogeneidad de las denominaciones oficiales.

Nóbrega (2008), jefa del Servicio de Traducción al Español en la Sede, señala que el volumen documental es mayor y más variado en la Sede, aunque los textos traducidos en la ONUG son más especializados, y resalta el papel del español como lengua de traducción en la Organización. A diferencia de lo que sucede en las instituciones europeas, en la ONU «conviven las variantes del español de la Península con las de los países de América Latina» (p. 138), lo que no plantea grandes conflictos. Debe mantenerse el «estilo ONU» y sus convenciones. Nóbrega subraya la dificultad para reclutar traductores de árabe y el problema que ello supone, ya que «hay numerosos documentos originales en ese idioma, algunos de gran complejidad» (p. 135).

García (2010), traductor de la ONUG y también de la CE, el Parlamento Europeo y la Organización Mundial del Comercio (en adelante, OMC), repasa la historia del multilingüismo en la Organización, describe el proceso de contratación y subraya la importancia de preservar la coherencia y la calidad, en particular gracias al «acceso a las referencias, la idoneidad de los medios técnicos utilizados y una adecuación de las expectativas de los clientes a las condiciones en que se realice el trabajo» (p. 402).

Lafeber (2012) destaca que la escasa investigación en la materia se basa en «expert opinion [...] or controlled observations [...] rather than on empirical research» (p. 22). Explora cómo mejorar la validez y la fiabilidad de los exámenes de reclutamiento de traductores mediante un estudio empírico basado en dos cuestionarios en los que participaron 320 traductores y revisores de unos 24<sup>29</sup> organismos internacionales. Concluye que las instituciones, si bien valoran en particular las habilidades lingüísticas, «also need research, computer, analytical and interpersonal skills, as well as extensive general knowledge and, possibly, specialized subject knowledge» (p. 134), y resalta las

---

<sup>29</sup> La autora no tiene constancia del número de organismos internacionales exacto debido al distinto grado de especificidad de las respuestas obtenidas.

diferencias interinstitucionales en el reclutamiento y en la valoración de habilidades, en particular entre la UE y la ONU.

Feria (2013), traductor en la Sede y la ONUG y profesor de la Universidad de Granada, analiza los obstáculos para la traducción del árabe en la ONU (la traducción de conceptos islámicos y la variación terminológica) y subraya que, frente al discurso jurídico islámico, los discursos científico-técnico, administrativo y jurídico no islámico presentan en árabe un alto grado de variación diatópica.

Fernández-Vitores (2014) analiza la incorporación del español como lengua oficial debido a la implicación de los Estados latinoamericanos en la Conferencia de San Francisco y señala que, en la práctica, el árabe tiene mayor reconocimiento como lengua oficial que el español (80 % frente a 77,6 %) y menor como lengua de trabajo (33 % versus 50 %) (p. 7)<sup>30</sup>. El español ocupa la tercera posición como lengua de trabajo en reuniones formales e informales, en particular en la ONUG, y en la redacción de documentos originales (aunque en comparación con el inglés su uso es mínimo), es prácticamente inexistente en la comunicación institucional interna (p. 15) y en la externa es la tercera lengua más utilizada en los sitios web de la ONU. Por todo ello, el español es «lengua de traducción, ya que la mayoría de los textos en español de las Naciones Unidas son traducciones de otras lenguas, normalmente del inglés» (p. 10).

Finalmente, Roldán y Feria (en prensa) analizan los cambios en las estructuras oracionales presentes en las traducciones al español de documentos de la ONUG redactados en árabe. Para ello, parten de un corpus multilingüe previamente compilado por Roldán (véase [III.1.a](#)) y utilizan un corpus paralelo árabe-español de 8 681 110 palabras gráficas compuesto por todos los informes presentados entre 2000 y 2018 por los Estados Miembros pertenecientes a la Liga de los Estados Árabes. Hasta donde se nos alcanza, es la única investigación sobre el producto en el ámbito de la traducción del árabe en organismos supranacionales. Además, los autores nos han confirmado que están

---

<sup>30</sup> Estos datos se calcularon a partir de la información obtenida por la DCI (2011, pp. 55-56) mediante cuestionarios enviados «a los coordinadores administrativos sobre las cuestiones del multilingüismo de las organizaciones participantes del sistema de las Naciones Unidas, y también a otras partes interesadas pertinentes, tales como las asociaciones del personal de los servicios de idiomas y otras organizaciones internacionales con dimensiones multilingüísticas pertinentes» (p. 2).

trabajando en la misma línea con otro corpus paralelo derivado del multilingüe, en esta ocasión árabe-inglés<sup>31</sup>.

Nuestra investigación, en definitiva, es junto con la de Roldán y Feria (en prensa) de las primeras en abordar la traducción del árabe en organismos supranacionales y en estudiar la traducción de documentos de seguimiento presentados por los Estados a la ONUG. La combinación lingüística, el entorno y el tipo documental permanecían hasta ahora inexplorados en los estudios de traducción.

A continuación exponemos las principales contribuciones académicas a la compilación de corpora de documentos de la Organización que comprendan textos en árabe.

### **c) Corpus de textos de la Organización de las Naciones Unidas**

Xu, Fraser y Weischedel (2001) compilaron el primer corpus paralelo árabe-inglés de textos de la ONU que, con un total de 38 000 pares de documentos (más de 50 millones de palabras en inglés), permitió realizar experimentos pioneros, en particular de *stemming* en árabe y de recuperación de información multilingüe.

Investigadores de la Universidad Autónoma de Madrid compilaron desde 2004 dos corpora con textos procedentes de diversos organismos de la ONU (Consejo de Seguridad, Consejo Económico y Social de la ONU, Asamblea General, UNESCO y otros). El primero, un corpus paralelo bilingüe español-árabe de 2 millones de palabras (Samy et al., 2004) y el segundo, un corpus paralelo multilingüe árabe-español-inglés de 3 millones de palabras (Samy et al., 2006).

Con el corpus paralelo bilingüe realizaron experimentos sobre alineado (Samy et al., 2004) en los que, para lograr una mayor eficacia, utilizaron dos clases de EN como «anchor points»: fechas y nombres propios. Etiquetaron las fechas mediante reconocimiento de patrones; para los nombres propios, primero los localizaron en el texto español mediante búsqueda de mayúsculas y posteriormente en el árabe con un lexicón bilingüe. También utilizaron este corpus para proponer un etiquetador de EN árabes (Samy et al., 2005).

---

<sup>31</sup> Juan José Roldán Verdejo, comunicación personal, 27 de octubre de 2021.

Samy (2005) utilizó una muestra de 300 oraciones en español (de un total de 1200 pares de oraciones disponibles) para llevar a cabo experimentos sobre patrones en la traducción de EN y estrategias para su traducción. Esta muestra se extrajo de un corpus paralelo multilingüe (2 millones de palabras, árabe-español-inglés) en fase de compilación por el Laboratorio de Lingüística Computacional. El etiquetado se realizó por lengua. Además, a pesar de contar con un etiquetador para el español<sup>32</sup> y la posibilidad de adaptarlo al inglés, fue necesario recurrir al etiquetado manual (de manera parcial para estas dos lenguas y total para el texto árabe).

En el año 2006 compilaron el primer corpus paralelo multilingüe en la combinación árabe-español-inglés. Su objetivo era alinear este corpus y, para lograr mejores resultados, etiquetaron de nuevo las EN para utilizarlas como «anchor points». Los autores afirman que la información lingüística que estas aportan a los modelos estadísticos hace que los resultados mejoren considerablemente. Finalmente, Samy y González-Ledesma (2008) utilizaron una muestra de este corpus para realizar experimentos de procesado automático multilingüe de marcadores discursivos.

A pesar del protagonismo de las EN en esos estudios tan novedosos y la gran contribución que supusieron, no nos consta que se le haya dado continuidad.

Rafalovitch y Dale (2009) compilaron un corpus paralelo multilingüe de dominio público y fines de investigación<sup>33</sup>. El corpus está formado por 2100 resoluciones de la Asamblea General en las seis lenguas oficiales. La dimensión total del corpus es de 3 millones de *tokens* de media por lengua. Los autores destacan entre las dificultades afrontadas la incomodidad del formato (PDF con dos columnas de texto y, en ocasiones, solo imagen). Desconocemos que este corpus diera lugar a investigaciones empíricas.

En 2010, Hammouda Salhi presentó un corpus paralelo bilingüe inglés-árabe (EAPCOUNT, 7 550 418 *tokens*<sup>34</sup> en total) de resoluciones de la Asamblea General y el Consejo de Seguridad y publicaciones de UNICEF y del FMI y sus traducciones al árabe. El objetivo era promover la lingüística de corpus en los estudios de traducción (Salhi,

---

<sup>32</sup> Etiquetador para el español desarrollado en el Laboratorio de Lingüística Computacional de la Universidad Autónoma de Madrid. Este etiquetador solo detecta nombres propios y fechas (Samy, 2005, p. 6).

<sup>33</sup> <http://opus.nlpl.eu/UN.php> [Última consulta: 05 de agosto de 2019].

<sup>34</sup> Véase [https://en.wikipedia.org/wiki/English-Arabic\\_Parallel\\_Corpus\\_of\\_United\\_Nations\\_Texts](https://en.wikipedia.org/wiki/English-Arabic_Parallel_Corpus_of_United_Nations_Texts) [Última consulta: 05 de agosto de 2019].

2010, 2013). Tampoco en este caso nos consta que este corpus diera lugar a investigaciones empíricas.

MultiUN<sup>35</sup> es un corpus multilingüe que comprende documentos en las seis lenguas oficiales (más el 1 % en alemán), originales y traducidos, descargados del sistema documental de la ONU. El corpus comprende 463 406 documentos, con una media de 326 millones de *tokens* en cinco de las seis lenguas oficiales (Eisele & Chen, 2010), alineados en cada una de las 21 combinaciones lingüísticas. Dos años después se actualizó (MultiUN v2): aumentó el 9 % (513 091 documentos) y mejoró la alienación de frases hasta alcanzar un porcentaje de éxito del 40 % en las seis lenguas (Chen & Eisele, 2012).

En 2016 se hizo público el United Nations Parallel Corpus, el primer corpus paralelo compilado por la Organización (Ziemski et al., 2016). Está compuesto por documentos en las seis lenguas oficiales de la ONU fechados entre 1990 y 2014. La mitad del contenido está disponible en las seis lenguas oficiales, lo que proporciona material en combinaciones que, como árabe-chino, han recibido escasa o nula atención.

El United Nations Parallel Corpus fue el resultado del interés prestado por el DGACM a la traducción automática estadística (en adelante, SMT, del inglés *Statistical Machine Translation*) a partir de 2011, fecha en que la Organización Mundial de la Propiedad Intelectual<sup>36</sup> (en adelante, OMPI) desarrolló su prototipo de TAPTA, un sistema SMT para la combinación inglés-español. La ONU mostró interés por el proyecto y se desarrolló TAPTA4UN para 10 combinaciones (del inglés a las demás lenguas oficiales y viceversa) (Pouliquen et al., 2013), cuyos resultados mejoraban los de Google o Bing. TAPTA4UN se insertó como motor de traducción en eLUNa (véase [II.2.b.ii](#)).

Finalmente, como adelantamos en el epígrafe anterior, Roldán y Feria (en prensa) analizan los cambios en las estructuras oracionales presentes en las traducciones al español de documentos de la ONUG redactados en árabe. Para ello, utilizan un corpus paralelo árabe-español de 8 681 110 palabras gráficas compuesto por todos los informes presentados entre 2000 y 2018 por los Estados Miembros pertenecientes a la Liga de los Estados Árabes.

---

<sup>35</sup> Corpus disponible en <http://www.euromatrixplus.net/multi-un/> [Última consulta: 06 de agosto de 2019].

<sup>36</sup> En la actualidad la OMPI utiliza el sistema WIPO Translate [https://www.wipo.int/patentscope/es/news/pctdb/2015/news\\_0004.html](https://www.wipo.int/patentscope/es/news/pctdb/2015/news_0004.html) [Última consulta: 06 de agosto de 2019]. Para más información sobre WIPO Translate: <https://patentscope.wipo.int/translate/wtapta-user-manual-en.pdf> [Última consulta: 06 de octubre de 2019]



Así pues, los corpórea de documentos de la ONU que incluyan textos en árabe, salvo Roldán y Feria (en prensa), carecen de los requisitos para la investigación lingüística, ya que no describen los métodos de compilación o no distinguen entre textos escritos en árabe o traducidos al árabe. De hecho, en su explotación solo se alude a la direccionalidad para justificar algún resultado (cuestiones de etiquetado (Samy et al., 2005) o de alineado (Chen & Eisele, 2012; Eisele & Chen, 2010; Samy et al., 2006, 2004). En Xu et al. (2001), por ejemplo, resulta imposible saber la direccionalidad. Estas carencias podrían deberse a que sus compiladores solo tienen interés en la investigación informática o fueron las propias instituciones y los realizaron para fines exclusivamente profesionales. A los estudios derivados del proyecto LETRINT aludimos en el apartado siguiente.

En contraste, el de Roldán y Feria (en prensa) es un corpus paralelo que cumple los requisitos establecidos en la lingüística de corpus, sus originales son árabes y presentan una naturaleza uniforme (documentos nacionales sobre derechos humanos presentados a la ONUG). Este corpus, como adelantábamos, fue compilado a partir de uno de mayores dimensiones. Por ser el ideal para nuestra investigación lo describimos en detalle más adelante ([III.1.a](#)).

#### **d) Homogeneidad en la traducción**

El interés por la homogeneidad en la traducción crece desde mediados de 1990. Merkel (1996) remitió un cuestionario con ejemplos extraídos de dos manuales de software a 13 traductores y líderes de proyectos de las agencias autoras de las traducciones y representantes de los clientes. Los resultados revelaron el consenso sobre la necesidad de homogeneización, en particular en la variación léxica y terminológica, con alguna diferencia de criterio entre los clientes y el resto de actores (p. 142).

Melamed (1997) midió mediante un corpus paralelo inglés-francés la ambigüedad semántica partiendo de la hipótesis de que palabras con un alto grado de carga semántica suelen tener un equivalente único en otras lenguas y que, por tanto, suelen traducirse con mayor homogeneidad que las palabras con una carga semántica menor. Concluye que la ambigüedad semántica es inversamente proporcional al contenido informativo de la palabra, la carga semántica y la homogeneidad en su traducción (p. 46).

Sarairoh (2001) critica la falta de homogeneidad y la variación estilística en los textos técnicos: «I think that the reader is better off with a boring (by repetition) but accurate text rather than dealing with a stylistic but confusing translation» (p. 19).

Wade (2003) emplea la homogeneidad como criterio para medir la calidad en la traducción de textos religiosos y determinar el número de traductores involucrados en la traducción de ciertas partes de la Biblia.

Rogers (2008) analiza la homogeneidad terminológica en instrucciones de uso de material médico originalmente escritas en alemán y en sus traducciones al inglés y francés. En la línea de Sarairoh (2001, p. 104), considera que en los textos especializados la homogeneidad terminológica debe ser plena. Los resultados del análisis revelaron fenómenos de variación intratextual en el original alemán, incrementados en francés (sin coincidencia plena con la variación del original) y menores en inglés, lo que «raises questions about possible differences in particular languages with respect to the patterning of lexical chains» (Rogers, 2008, p. 107). Esa falta de homogeneidad conduce a la autora a cuestionar la calidad del original y las traducciones.

Guillou (2013) utiliza un corpus de textos (INTERSECT, sección inglés-francés) para analizar la homogeneidad en la traducción humana de palabras frecuentes, clasificadas según su categoría gramatical, la homogeneidad en una traducción realizada con un sistema SMT y comparar los resultados. Al igual que Melamed (1997), concluye que «human translators use lexical consistency to support what is important in a text» (Guillou, 2013, p. 10) y traducen de manera más homogénea los nombres que los verbos, tanto en francés como en inglés, si bien los verbos de uso menos frecuente se traducen de manera más uniforme que los de uso más frecuente. La traducción automática presentó mayor homogeneidad que la humana.

Xiao, Zhu, Yao y Zhang (2011), Carpaut y Simard (2012) e Itagaki, Aikawa y He (2013, 2007) se centran en la mejora de los sistemas de traducción automática. Los primeros estudian el modo de integrar contextos documentales en los sistemas de SMT con particular atención a la homogeneidad, que consideran «one of the most important issues in document-level MT» (Xiao et al., 2011, párr. 141).

Carpaut y Simard (2012) analizan si un sistema de SMT oracional traduce intratextualmente de manera uniforme y si las inconsistencias constituyen errores de

traducción. Los sistemas testados tradujeron con notable homogeneidad pese a carecer de información supraoracional, aunque homogeneidad no implicó necesariamente calidad, ya que «as can be expected in phrase-based SMT, weaker systems trained on less data produce translations that are more consistent than higher-quality systems trained on larger more heterogeneous data sets» (p. 448). Los autores explican ciertos fenómenos de heterogeneidad por errores semánticos o problemas sintácticos o estilísticos.

Itagaki, Aikawa y He (2013, 2007) miden la homogeneidad terminológica intratextual con un sistema que extrae nombres compuestos de un corpus bilingüe alineado inglés-japonés. Para ellos, la homogeneidad es el principal criterio de calidad en la traducción de texto especializado, ya que «inconsistent translations of terminology can cause confusion». No obstante, admiten que la terminología pueda variar en función del contexto.

Aljanati (2017) considera la homogeneidad en la terminología jurídica imprescindible para el correcto funcionamiento de la UE y destaca el papel que desempeña en esa esfera la *Guía práctica común del Parlamento Europeo, del Consejo y de la Comisión para la redacción de textos legislativos de la Unión Europea*. Concluye que para mantener la coherencia es necesario una buena base de datos terminológica y herramientas de traducción asistida por ordenador (en adelante, TAO).

Para Prieto Ramos, la homogeneidad es clave en la calidad de la traducción y, por tanto, esencial en la gestión terminológica orientada a la traducción (2020, p. 128). En el marco del proyecto LETRINT, estudia con su equipo la adecuación y consistencia de la toma de decisiones sobre terminología jurídica.

En ese sentido, Prieto Ramos y Guzmán (2018) analizaron la homogeneidad intratextual e intertextual en la traducción del término «due process» en tres subcórpora paralelos con documentos traducidos del inglés al español en la OMC, la UE y la ONU. La homogeneidad fue mayor en la OMC y menor en la UE, lo que suponen asociado a «the greater multiplicity of EU translation services [...] and the low adherence to the relevant IATE terminological recommendations» (p. 97). Los resultados para la ONU son similares a los de la OMC, pese al volumen documental y el reparto de tareas entre la Sede y la ONUG.

Poco después, el anterior estudio se amplió a otros términos considerados representativos (Prieto Ramos, 2020). Los hallazgos sugieren correlaciones significativas entre la asimetría entre sistemas jurídicos y los niveles de precisión de la traducción, y entre la consistencia intertextual y las fluctuaciones de precisión.

Un último estudio (Guzmán & Prieto Ramos, 2021) analizó la variación intratextual e intertextual de 30 denominaciones de tribunales nacionales en seis sistemas jurídicos distintos presentes en textos presentados ante los procedimientos de monitorización de derechos humanos de la ONU. Emplearon dos corpus de textos paralelos y comparables (inglés-español y francés-español) fechados entre 2000 y 2017. Los autores analizaron las decisiones terminológicas como indicadores de la competencia metodológica y la calidad en la traducción jurídica, en la misma línea que estudios previos. Los resultados revelaron que la mayoría de los elementos se tradujeron de manera heterogénea, aunque en muchos casos la inconsistencia era mínima. Cabe señalar que los autores no tienen en consideración las diferencias ortográficas o la flexión de número.

En algunos casos, las conclusiones de la literatura previa se nos antojan ingenuas: para Melamed (1997), por ejemplo, la homogeneidad es inversamente proporcional a la ambigüedad semántica y para Itagaki et al. (2007), constituye el parámetro máximo para la evaluación de la calidad. La literatura institucional de los organismos supranacionales sigue la misma línea. Desde luego no parece una noción muy realista del léxico jurídico, tan difícil de encasillar como cualquier otra manifestación lingüística de la vida social, y menos realista aún cuando se trata de lenguas asociadas a sistemas jurídicos varios.

Más realista es la insistencia de Lafeber en la necesidad de estandarización en los organismos supranacionales, pues «each transaltion sets a precedent for future language use» (2012, p. 19). En la ONU, esto es particularmente cierto desde la introducción de eLUNa (véase [II.2.b.ii](#)). En cualquier caso, aunque como afirma Koskinen (2000, p. 58) cambiar el estilo o la nomenclatura institucionales es muy delicado, pues provoca la coexistencia de distintas versiones y afecta a la homogeneidad intertextual, esa coexistencia es, en ocasiones, imprescindible por la propia evolución ideológica del organismo (un caso claro es el paso en la ONUG de «inválido» a «discapacitado» y, finalmente, «persona con discapacidad»).

Los estudios derivados del proyecto LETRINT (Guzmán & Prieto Ramos, 2021; Prieto Ramos, 2020; Prieto Ramos & Guzmán, 2018) y nuestra investigación son, que sepamos, los únicos trabajos que analizan empíricamente la homogeneidad en el producto, aunque difieren en la composición del corpus (UE, ONU y OMC frente a textos de la ONUG, a excepción de Guzmán y Prieto Ramos (2021)), el objeto (un número muy reducido términos de diversos Estados frente a cientos de EN de un único Estado), la combinación (inglés-español y francés-español frente a árabe-español-inglés) y la metodología en su conjunto.

## **2. Informes, orientaciones y herramientas institucionales**

Una fuente adicional y fundamental de información sobre la traducción institucional, aunque apenas explotada en la investigación previa, son los informes, recomendaciones y descripciones oficiales de las herramientas al servicio del traductor de la ONU. A continuación se describe el contenido de esa literatura, también en orden cronológico.

### **a) Informes institucionales**

La ONU ha afirmado el principio de multilingüismo y su inquietud por los obstáculos a su efectividad en numerosos documentos oficiales, entre otros los informes periódicos pertinentes de la Dependencia Común de Inspección (en adelante, DCI, del inglés *Joint Inspection Unit*), órgano independiente que supervisa y evalúa los recursos humanos y financieros de la ONU (Naciones Unidas, s. f.-a), y de la ONUG (Oficina de las Naciones Unidas en Ginebra, s. f.-a).

En su primer informe en la materia, la DCI subrayó la carencia de análisis institucionales previos sobre los servicios de traducción (1980, p. 4). En él también se insistió en que los recursos no corrían parejos al incremento de la producción documental y en la dificultad para cumplir los plazos manteniendo la calidad, así como para encontrar personal cualificado en ciertas lenguas.

La preocupación por que el incremento del volumen documental y la falta de recursos afecte a la calidad de las traducciones es constante (Dependencia Común de Inspección

de las Naciones Unidas, 1980, p. 6, 2011, pp. 27-28; Oficina de las Naciones Unidas en Ginebra, 2015, 2016). Los servicios de traducción prestados al ACNUDH representaron el 60,8 % del volumen de traducción de la ONUG en 2010. Entre 2008 y 2010 el volumen de traducción en el ACNUDH aumentó el 40 % según datos de la DCI.

Ante ello, una recomendación principal de la DCI es que

cuando se creen nuevos órganos institucionales que requieran servicios de conferencias, los órganos legislativos de las organizaciones del sistema de las Naciones Unidas deberían prever los recursos presupuestarios vinculados con el consiguiente aumento de la carga de trabajo, en particular de traducción e interpretación. (2011, p. 28)

Sin embargo, la previsión de los recursos no es siempre sencilla. En el ACNUDH, numerosos tipos documentales no tienen una extensión máxima prevista y «por lo tanto, resulta casi imposible elaborar un plan previo sin conocer con anticipación el número previsto de páginas de los futuros documentos» (2011, p. 28). Por otro lado, más del 20 % de los documentos se presentan en la ONUG en una lengua oficial distinta al inglés. Ese hecho, unido a la dificultad para encontrar personal cualificado, obliga a recurrir al relé (Oficina de las Naciones Unidas en Ginebra, 2017, p. 20).

Ese gran número de documentos en lenguas oficiales distintas al inglés se debe al origen externo de los textos. Según informes de 2017, el 68 % de los documentos traducidos en Ginebra fueron externos (Oficina de las Naciones Unidas en Ginebra, 2017, p. 20), frente al 7 % de lo traducido en la Sede (Dependencia Común de Inspección de las Naciones Unidas, 2017, p. 20). La documentación externa puede suponer redacción pobre, extensión excesiva y entrega extemporánea (Dependencia Común de Inspección de las Naciones Unidas, 1980, p. 36). Todo ello afecta negativamente a los servicios de conferencias.

Otro motivo de inquietud para la DCI y la ONUG es la estandarización de la terminología. En el primer informe en la materia emanado de la DCI ya se afirmaba la necesidad de eliminar duplicados y desarrollar un «banco común computadorizado de datos terminológicos» (1980, p. 31 y 40). La recomendación se aplicó en 2011 con la puesta en marcha de UNTERM (Oficina de las Naciones Unidas en Ginebra, 2012, p. 27).

Las nuevas herramientas en línea que describimos a continuación, en particular eLUNA, han contribuido a aumentar la eficiencia de los servicios de traducción y paliar los

problemas conexos (Oficina de las Naciones Unidas en Ginebra, 2016, 2017, 2018). Estas iniciativas en curso están dirigidas a solucionar los problemas de calidad, puntualidad y costo de las traducciones (Dependencia Común de Inspección de las Naciones Unidas, 2020, pp. 17-18).

## **b) Herramientas de apoyo a la traducción para la Organización de las Naciones Unidas**

No disponemos de datos pertinentes sobre la ONU, pero en la UE «up to 40 % of the “new” text can be recreated from past versions» (Wagner et al., 2002, p. 89). Las herramientas TAO, por tanto, ahorran una gran cantidad de tiempo; además, promueven la homogeneidad terminológica, en particular cuando equipos amplios de traductores afrontan proyectos de gran volumen (Vasconcellos, Avey, Gdaniec, Gerber, León, Mitamura, Wright y Budin 2001, p. 697). Considerando los problemas señalados en el apartado anterior, es indiscutible su utilidad para la Organización (Dependencia Común de Inspección de las Naciones Unidas, 2020, pp. 17-18; Fall & Zhang, 2011; Nóbrega, 2008).

A continuación se presentan dos herramientas fundamentales para la traducción en la ONU: UNTERM y eLUNa. La información que sigue se extrajo de los informes institucionales y las guías de usuario de la propia Organización, ya que la literatura académica las ha ignorado hasta el momento.

### **i. United Nations Multilingual Terminological Database (UNTERM)**

Hasta 2011, cada oficina de la ONU disponía de sus propias colecciones terminológicas y glosarios por países. Ese año se fusionaron en UNTERM, aunque con numerosos duplicados que se siguen eliminando en la actualidad. La herramienta es de acceso público y está gestionada por las oficinas de la Organización (Oficina de las Naciones Unidas en Ginebra, 2016, p. 49; Servicio de Traducción al Español de las Naciones Unidas en Nueva York, 2015).

La ONUG afirma que UNTERM pretende evitar la duplicación de esfuerzos, compartir datos de valor añadido y asegurar la interoperatividad (2012, p. 27). La herramienta cuenta con entradas en las seis lenguas oficiales (también algunas en alemán y portugués) que versan, en particular, sobre derechos humanos, medio ambiente, mantenimiento y consolidación de la paz y cuestiones de género y desarrollo (Servicio de Traducción al Español de las Naciones Unidas en Nueva York, 2015). Comprende asimismo nombres de «organismos, departamentos y divisiones e información específica sobre los países, como el nombre oficial, la moneda y su símbolo, la capital, el gentilicio y el código ISO del país y la moneda»<sup>37</sup>. Aunque no se declara en ningún informe oficial, UNTERM también comprende nombres de leyes.

En cada entrada se especifica el *anchor language* (visible también en la versión pública, al contrario que en IATE), normalmente «country-specific», información de la fuente, sinónimos, acrónimos y otras aclaraciones. Se recomienda el uso de los filtros disponibles para hacer búsquedas más ajustadas y obtener un menor número de resultados (Naciones Unidas, s. f.-d).

Con *anchor language* árabe, la categoría de textos normativos es prácticamente inexistente en UNTERM y las entidades corporativas nacionales tienen una presencia muy discreta. Además, el reconocimiento automático de entidades nombradas en árabe es muy poco efectivo (II.3.c). Debido a la escasez de recursos, la falta de pasantes y terminólogos que dominen el árabe y la exigente productividad exigida a los traductores, UNTERM se ha desarrollado en la combinación árabe a español o inglés a un ritmo inferior al ideal<sup>38</sup>.

## ii. Electronic Languages United Nations (eLUNa)

eLUNa es una herramienta de TAO desarrollada para los traductores de la ONU (no es de uso público) que empezó a implantarse en 2014. Está dotada de funcionalidades típicas: segmenta el texto, proporciona un espacio de edición, detecta segmentos (completos o parciales) ya traducidos, los resalta y los presenta en forma de bitexto,

---

<sup>37</sup> <https://onutraduccion.wordpress.com/2015/06/09/nuevo-portal-unterm/> [Última consulta: 17 de febrero de 2017]

<sup>38</sup> Jaime Sánchez-Ratía, comunicación personal, 16 de noviembre de 2018.



localiza coincidencias terminológicas en UNTERM (aunque solo las comunes a todo el sistema y las de la oficina de adscripción del usuario), proporciona acceso a herramientas de TA y permite exportar a formato “.docx”. (Departamento de la Asamblea General y de Gestión de Conferencias de las Naciones Unidas, 2015, pp. 4-7; Dependencia Común de Inspección de las Naciones Unidas, 2020, p. 18).

La herramienta ha integrado varios motores de TA:

Machine translation systems accessible through for eLUNa are: (1) Tapta4UN, a statistical machine translation engine [...] fed with UN secretariat documents, before being handed over to DGACM in 2011; (2) Bing, which is Microsoft’s statistical machine translation engine; (3) UN Bing, which is a special section of Bing, fed exclusively with UN documents. (Departamento de la Asamblea General y de Gestión de Conferencias de las Naciones Unidas, 2015, p. 4)

En noviembre de 2018 integró DeepL<sup>39</sup>.

eLUNa, tras un tiempo en que solo funcionó cuando el inglés formaba parte de la combinación lingüística (Wroblewski, 2014, p. 20), comenzó a incluir otras combinaciones, inclusive árabe a español o inglés. Hasta 2020, para traducir del árabe solo fue posible utilizar Tapta4UN como motor de traducción, cuyos resultados eran desalentadores. Desde enero de ese año, la introducción de MS Translator supuso cierta mejora.

Además de eLUNa, el DGACM cuenta con eLUNa Editorial (aplicación autónoma de búsqueda de texto completo con operadores booleanos y de proximidad en todos los idiomas oficiales que muestra los resultados en formato monolingüe, bilingüe y trilingüe), eLUNa Search (una interfaz para editar documentos de las Naciones Unidas que señala las repeticiones y la terminología localizada en UNTERM) y eLUNa VRS (una interfaz para la redacción de actas literales) (Dependencia Común de Inspección de las Naciones Unidas, 2020, p. 91).

El nivel de recuperación terminológica de eLUNa en árabe es también muy inferior al de otras lenguas. Las razones son varias: carencia de entradas en UNTERM, errores o simples variedades ortográficas (solo reconoce el término cuando se escribe la *hamza* o

---

<sup>39</sup> Manuel Feria García, comunicación personal, 29 de enero de 2020.

la *shadda*, por ejemplo) y, sobre todo, fenómenos de aglutinación gráfica (la mera presencia de la conjunción, el artículo definido o un clítico impide la recuperación). Todo ello se agrava por el origen externo de los textos y la falta de edición en la Organización<sup>40</sup>.

Los informes anuales de la ONUG subrayan que gracias a esas herramientas «operators don't spend time duplicating past efforts, and their translations are more consistent across documents» (2017, 2018, p. 120), y valoran positivamente eLUNa y UNTERM, ya que aseguran una mayor coherencia en los documentos traducidos (2017, p. 19). Todo ello resulta «vital» para la Organización (2016, p. 49). Sin embargo, la DCI reconoce que el apoyo a las distintas lenguas oficiales de la ONU es desigual en esta esfera y se recomienda un mayor control de la calidad para evitar que las malas traducciones se perpetúen (2011, p. 30 y 31). La situación en la traducción del árabe refleja ese apoyo desigual.

Pese al interés de la Organización en fomentar herramientas que permitan producir sin mermar la calidad y hacer efectivo el multilingüismo en un contexto de incremento documental incesante y recursos financieros en mengua, no todas las lenguas se han beneficiado en igual medida de ese desarrollo. El árabe es una de esas lenguas y las posibilidades de que ello afecte a la homogeneidad, en particular si se ha de pasar por el relé, son notables incluso para las EN.

Otra cuestión es si el incremento del volumen documental afecta en igual, menor o mayor medida al árabe que a otras lenguas de la Organización. Hemos podido analizar por vez primera el asunto gracias a la información compilada por Feria y Roldán<sup>41</sup> y adelantamos las conclusiones en el gráfico siguiente, en el que se muestra la evolución del volumen documental presentado en árabe al ACNUDH entre 2000 y 2016, con datos desglosados por número de palabras y expresión de la variación porcentual interanual. En esos 16 años, el aumento promediado fue del 120 %, aunque desde 2013 la situación se estabiliza en torno a una media de 400 000 palabras anuales, frente a las 200 000 de 2000 a 2008. En definitiva, ello significa que el incremento en el volumen de palabras árabes triplicó las cifras generales que la DCI ofrecía para el intervalo 2008 a 2010 (40 %).

---

<sup>40</sup> Manuel Feria García, comunicación personal, 29 de enero de 2020.

<sup>41</sup> Manuel Feria García y Juan José Roldán Verdejo, comunicación personal, 24 de febrero de 2019.

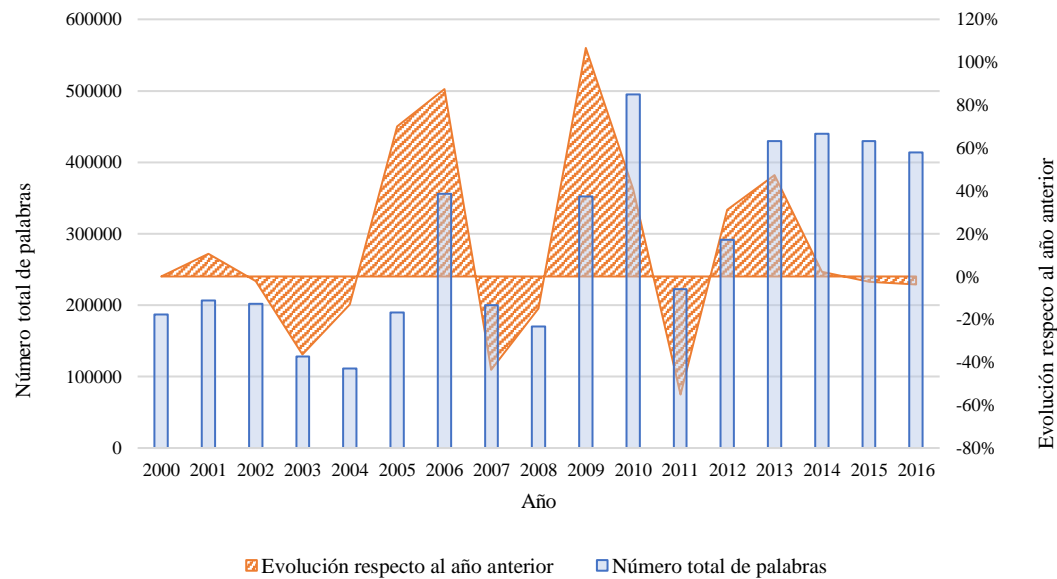


Gráfico 1: Evolución anual del volumen de palabras traducidas del árabe en el ACNUDH entre 2000 y 2016, con datos desglosados por número de palabras y porcentaje de variación interanual

Fuente: Datos proporcionados por M. Feria García y J.J Roldán Verdejo. Elaboración propia

La práctica ausencia de entradas sobre textos normativos árabes en UNTERM, unido a este incremento del volumen de traducción y la falta de personal, podría distanciar la traducción del árabe de los márgenes aceptables en la materia para otras combinaciones lingüísticas, incrementar el tiempo de revisión y dificultar, como veremos más adelante, la EI y la RI.

### c) Orientaciones para la traducción en la ONU

Las secciones de español proporcionan a sus traductores las orientaciones específicas que, por su importancia para nuestra investigación, exponemos a continuación.

#### i. Servicio de Traducción al Español

Las secciones de traducción al español de la Sede y la ONUG disponen de sus propias páginas web independientes. Nos centraremos en las orientaciones dirigidas a los traductores de la ONUG.

El sitio web de la Sección de Traducción al Español en Ginebra, la *Paginilla* (<https://lists.unog.ch>), remite a multitud de recursos, entre otras cosas a bases documentales y terminológicas, manuales, directrices, orientaciones sobre terminología jurídica y de otra índole, información sobre búsquedas de documentos y referencias y jerarquía de las fuentes, enlaces de interés sobre eLUNa, una guía resumida de UNTERM e información de carácter práctico o administrativo. *Paginilla* también dirige al *Manual del traductor del Servicio de Traducción al Español de las Naciones Unidas* (en adelante, *Manual*) (Servicio de Traducción al Español de las Naciones Unidas en Nueva York, s. f.), que se encuentra alojado en el sitio web de la Sección en la Sede y es «de consulta y lectura obligatoria» (Sección de Traducción al Español - *Paginilla*, s. f.-c).

En el *Manual* se formulan recomendaciones sobre el uso de mayúsculas y cursivas y sobre la traducción al español de nombres de partidos políticos, cargos, entidades y actos normativos. Las recomendaciones sobre los textos normativos se amplían en las *Orientaciones para la Traducción de Textos Jurídicos* (en adelante, *Orientaciones*) (Sección de Traducción al Español de las Naciones Unidas - *Paginilla*, 2015).

El *Manual* exige emplear mayúscula inicial en todas las palabras significativas en las denominaciones de leyes, declaraciones, instrumentos normativos y otros documentos oficiales o históricos, organismos, instituciones, departamentos, asociaciones, grupos de trabajo, comités y órdenes religiosas. Las palabras sin carga significativa portarán minúscula, «a menos que en la denominación en inglés figure con mayúscula la palabra correspondiente». Los nombres de estatutos y reglamentos, sin embargo, portan mayúsculas iniciales con independencia de cómo figuren en inglés (Servicio de Traducción al Español de las Naciones Unidas en Nueva York, s. f.). El *Manual* cita los ejemplos siguientes: «Rules of the International Court of Justice» = «Reglamento de la Corte Internacional de Justicia» y «provisional rules of procedure of the Security Council» = «Reglamento Provisional del Consejo de Seguridad».

Es importante subrayar que esas recomendaciones comenzaron a aplicarse en la ONUG en 2014<sup>42</sup>. Antes solo portaba mayúscula inicial la primera palabra de la denominación. El propio *Manual* advierte de que en UNTERM se alojan entradas antiguas que no siguen estas normas y que cualquier discrepancia se debe poner en conocimiento de los

---

<sup>42</sup> Manuel Fera, comunicación personal, 12 de mayo de 2015.

terminólogos (Servicio de Traducción al Español de las Naciones Unidas en Nueva York, s. f.). Este cambio en la norma relativa al uso de mayúsculas iniciales es un elemento importante en nuestra investigación, puesto que los documentos analizados se tradujeron entre 2000 y 2017. Así pues, la falta de homogeneidad puede deberse a la evolución terminológica de la Organización o a la evolución de sus normas editoriales, y nada de ello es necesariamente calificable como negativo.

Las *Orientaciones* formulan las recomendaciones siguientes en la línea del *Manual*:

- Se escriben con mayúscula inicial todas las palabras significativas (normalmente sustantivos y adjetivos) que forman parte del título de una ley o cualquier otro instrumento normativo interno o internacional, como tratados, convenciones, convenios, pactos y demás instrumentos internacionales;
- Se escribe con mayúscula inicial cada una de las palabras que integran el título de los órganos jurisdiccionales nacionales y los Ministerios, las Secretarías de Estado y demás órganos políticos, jurídicos o administrativos internos;
- Por regla general, el nombre de los órganos jurisdiccionales nacionales se traducirá al español;
- Por regla general, se traducirá al español el nombre de los órganos políticos y administrativos;
- Para la traducción de normativa nacional se establece que

Law on...: El uso jurídico prefiere el “de” al “sobre” para traducir la preposición *on* que introduce el título de muchas leyes. *Law on...* se traducirá pues como “Ley de...”

Gerundio: En cuanto al gerundio inglés *creating, regulating, establishing*, etc., que suele aparecer en numerosos actos jurídicos, la formulación más extendida es Ley “por la que se crea”, Decreto “por el que se regula”, Reglamento “por el que se establece”, etc.

Act/Law amending...: Suele traducirse como “Ley por la que se modifica...” o “Ley de modificación...”. Salvo en caso de reformas o revisiones constitucionales, debe evitarse el término “enmendar”, que transmite la idea de “arreglar o quitar defectos”, cuando simplemente se trata de introducir cambios en la norma preexistente. (Sección de Traducción al Español de las Naciones Unidas - Paginilla, 2015)

*Paginilla* también formula directrices y recomendaciones para la traducción del árabe al español suscritas por Sánchez-Ratia, principalmente aclaraciones sobre la transcripción

de nombres propios, expresiones religiosas, fechas y cifras. Aunque el autor no alude a los elementos analizados en esta investigación, subraya las dificultades que presenta la traducción del árabe, en particular la ausencia de mayúsculas, la escasa estandarización de la terminología (en consonancia con Feria 2013), la falta de traductores y la necesidad de recurrir al relé (Sección de Traducción al Español - Paginilla, s. f.-a).

## ii. Servicio de Traducción al Inglés y Edición

A diferencia del Servicio de Traducción al Español, no hemos localizado para el inglés un sitio web con recursos exclusivos de la ONU. Por tanto, recopilamos de manera general las directrices pertinentes de la ONU.

En el *United Nations Editorial Manual Online* figuran reglas ortográficas y recomendaciones de estilo. Respecto al uso de mayúsculas en los elementos que se analizan en esta investigación, afirma:

Initial capital letters are used for: [...] The official titles of persons, councils, commissions, committees, Secretariat units, organizations, institutions, political parties and organized movements, [...] when citing the title of a specific decree, [...] in specific references: the Department of Political Affairs of the Secretariat, the Department of Commerce of the United States Government. (*Capitalization in English*, s. f.)

Asimismo, al citar textos normativos en textos de naturaleza no jurídica,

the reference may include the following elements:

1. Country (not "Government of ...") or other jurisdiction
2. Title of law or body of laws, if any (in standard font without quotation marks)
3. Number of law, if any
4. Date of adoption, if not part of title (in parentheses)
5. Article, part, chapter or section, if relevant

[...] When the title of a law is given in another language, it should be translated where possible, or a description provided. When it is not possible to provide a translation or description, the title in the original may be retained but is not italicized. (*Footnotes and other references. IV. Outside sources. E. Legal references*, s. f.)

Así pues, la norma en inglés para las denominaciones de textos normativos y entidades corporativas nacionales es la misma que en español, salvo que en inglés se ha aplicado siempre y en español, en la ONUG, comenzó a aplicarse en 2014.

### 3. Lingüística computacional

A continuación introduciremos nociones relativas a la recuperación y extracción de información y describiremos las categorías de EN analizadas por la literatura previa y sus herramientas de extracción automática, en particular las disponibles en árabe.

#### a) Recuperación y Extracción de Información

Los sistemas de Recuperación de Información (en adelante, RI, *Information Retrieval*) y de Extracción de Información (en adelante, EI, *Information Extraction*) facilitan el acceso a grandes volúmenes de información y reducen de manera drástica el tiempo de búsqueda y consulta. Ambos proporcionan un *output* con formato específico y estructurado, lo que permite alimentar bases de datos (Piskorski & Yangarber, 2013). De este modo, datos no estructurados de conjuntos documentales pueden reutilizarse en sistemas de información (Sáez Guerrero, 2009, p. 5).

La diferencia entre esos sistemas es que los de RI localizan textos relevantes en formato electrónico mediante palabras clave y los de EI localizan y extraen el contenido relevante en esos documentos (Cunningham, 2005; Turmo, Ageno y Català 2006; Sáez Guerrero, 2009; Shaalan, 2014). La eficiencia es una ventaja fundamental de los EI, aunque son computacionalmente más complejos (Cunningham, 2005).

En los sistemas de EI debe definirse *a priori* el tipo de contenido que se desea extraer. Por ello, «when dealing with new domains, new specific knowledge is needed and has to be acquired by such systems» (Turmo et al., 2006, p. 2). Hasta ahora, ello ha implicado en la mayoría de los casos, y así ha sido también en el nuestro, un análisis lingüístico manual o semi automático del dominio en cuestión. Sin embargo, «se están desarrollando métodos de aprendizaje automático para adquirir de forma automática los patrones de extracción útiles» (Oyarzún Hernández & Vergara Salas, 2014).

Los sistemas de EI realizan cinco tareas fundamentales:

- Reconocimiento de Entidades Nombradas (REN, *Named Entity Recognition*): Localización y clasificación de nombres, lugares, fechas, etc.;
- Resolución de correferencia (*Coreference Resolution*): Localización de entidades o referencias que hacen alusión a lo mismo;
- Reconocimiento de función semántica (*Template element construction*): Asignación de funciones semánticas a componentes de una oración;
- Reconocimiento de relación entre entidades (*Template relation construction*): Localización de las relaciones entre dos o más entidades y clasificación con una función semántica;
- Reconocimiento de expresiones temporales (*Scenario template reproduction*): Detección de eventos en el texto, es decir, de expresiones temporales, duraciones, expresiones ancladas a eventos o repeticiones.

En nuestra investigación interesa el Reconocimiento de Entidades Nombradas.

## **b) Reconocimiento de Entidades Nombradas**

El desarrollo de sistemas de EI comenzó a finales de 1980 con las *Message Understanding Conferences* (en adelante, MUC). En 1995, la MUC-6 planteó como reto el reconocimiento de nombres de personas, organismos o lugares («ENAMEX»), expresiones numéricas («NUMEX») y expresiones temporales («TIMEX») en inglés. Se acuñaron así los términos «Entidad Nombrada» (en adelante, EN, *Named Entity*) y «Reconocimiento de Entidades Nombradas» (en adelante, REN, *Named Entity Recognition*) (Grishman & Sundheim, 1996). Desde entonces, el REN es una tarea fundamental de la EI (Cunningham, 2005; Nadeau & Sekine, 2007; Sekine, 2004).

En 2002 comenzaron las *Conference on Computational Natural Language Learning* (en adelante, CoNLL) y las *Automatic Content Extraction* (en adelante, ACE), en las que se ampliaron las lenguas de investigación al alemán, árabe, chino, español y holandés<sup>43</sup>.

---

<sup>43</sup> En las CoNLL y las ACE se han testado numerosas técnicas para el REN, como el principio de máxima entropía (*Maximum Entropy*), máquinas de vectores de soporte (*Support Vector Machine*) y campos aleatorios condicionales (*Conditional Random Fields*) (Benajiba, 2009, p. 46, 48 y 59). Por motivos de espacio y relevancia no abordamos esas cuestiones.



El REN consiste en la localización en textos no estructurados, la extracción y la clasificación automática de categorías predefinidas de EN (Shaalán, 2014, p. 470). La mayoría de herramientas diseñadas para esta tarea se basan en corpus de prensa, ya que, como quedó dicho en el apartado anterior, abordar géneros o dominios diferentes resulta complejo (Nadeau y Sekine, 2007). Esta limitación explica que, pese a que a principios de la década de 2000 Sekine & Nobata (2004) ya clasificaron más de 200 categorías de EN, en la práctica pocas de ellas han sido testadas. Obviamente, como afirma Mohit (2014, pp. 222 y 223), categorías como «persona», «lugar» y «organización» no abarcan las EN relevantes en todo contexto: el de la ONU, por ejemplo, exigió a Samy (2005) emplear categorías como «acrónimo», «evento» o «empleo». Los sistemas siguientes muestran la importancia de crear categorías adecuadas a nuevos dominios y testarlas:

- EXIT extrae antropónimos y nombres de empresas e inmuebles en escrituras de compraventa (Llopis et al., 1998);
- El ámbito de la biomedicina es un dominio destacable (Leaman & Lu, 2016; Li et al., 2015; Tang et al., 2015). Dingare, Finkel, Nissim, Manning y Grover (2004), por ejemplo, definen cinco categorías de EN: proteína, ADN, ARN, líneas celulares y tipos celulares, y subrayan la dificultad del etiquetado y el limitado rendimiento del sistema debido a la presencia de un léxico más complejo que en la prensa.

### c) Reconocimiento de Entidades Nombradas Árabes

El árabe, en general, es una lengua de gran interés para el Procesamiento del Lenguaje Natural (en adelante, PLN). La investigación sobre el Reconocimiento de Entidades Nombradas Árabes (en adelante, RENA) comenzó más tarde que para otras lenguas, sobre 2005 (Darwish & Gao, 2014), y se desarrolló con lentitud (Mohammed & Omar, 2012, p. 1286), entre otras cosas por las características de la lengua (Farghaly & Shaalan, 2009; Habash, 2010; Shaalan, 2014). Entre las dificultades para el tratamiento computacional del árabe conviene señalar las siguientes (Darwish, 2014; Florian et al., 2004; Herrero y Nebot 1994; Mohit, 2014, p. 231; Trabelsi, Zribi, & Kouki, 2015):

- Dirección de escritura de izquierda a derecha e incompatibilidad con muchos *softwares*;

- Diferente escritura de la mayoría de las letras según su posición (inicial, intermedia, final o aislada);
- Uso opcional de las vocales cortas y consiguiente incremento de la ambigüedad;
- Morfología: con más de 500 patrones de derivación, aunque pocas raíces acepten más de 20, el reconocimiento automático de la raíz resulta muy complejo;
- Aglutinación: conjunciones, preposiciones y pronombres personales se aglutinan gráficamente a la palabra;
- Diversidad de opciones de latinización, y
- Ausencia de mayúsculas, indicador fundamental de la presencia de EN en otras lenguas (Samy, 2005, p. 460; Samy et al., 2005, p. 5 y 6).

Pese a esas dificultades, el desarrollo de herramientas de PLN en árabe comienza a despegar en los primeros años del nuevo milenio con Buckwalter (2004), con el desarrollo de BAMA (Buckwalter Arabic Morphological Analyzer), que ya cuenta con un amplio lexicón árabe-inglés y también tiene en cuenta las mayúsculas en inglés para identificar las EN árabes (Shaalán, 2014, pp. 482-483).

Pocos años más tarde, se desarrolló el sistema MADA+TOKAN (Habash et al., 2009), compuesto por un analizador morfológico y desambiguador (MADA) combinado con un *tokenizador* (TOKAN), ambos para la lengua árabe. El analizador morfológico de este sistema crea una lista de todas las posibles interpretaciones morfológicas de cada palabra y aplica una serie de modelos predictivos a efectos de desambiguación. Dicho análisis puede usarse posteriormente para obtener mediante TOKAN la *tokenización* de la palabra en cuestión. La combinación de estos dos programas conforma una herramienta de preprocesado interesante para aplicaciones como la traducción automática o el REN.

Más adelante se desarrolló MADAMIRA (Pasha et al., 2014)<sup>44</sup>, sistema que combina las utilidades de MADA y AMIRA, un *tokenizador* y etiquetador gramatical que, en sus versiones más desarrolladas, también contenía un analizador morfológico y un módulo de REN. El objetivo era mejorar el rendimiento y eficacia al combinar esos sistemas. Su lenguaje de programación es ágil y fácil de integrar en otro *software* existente.

MADAMIRA supuso un salto cualitativo de enorme trascendencia e inspiró el desarrollo de muchos otros sistemas (Guellil et al., s. f.), como QATARA (Darwish et al., 2014),

---

<sup>44</sup> Disponible en <https://camel.abudhabi.nyu.edu/madamira/> [Última consulta: 27 de octubre de 2019]

Standford Arabic Segmenter (Monroe et al., 2014), FARASA (Abdelali et al., 2016), YAMAMA (Khalifa et al., 2016) y CamelParser (Shahrour et al., 2016).

De todos ellos, MADAMIRA y FARASA son los únicos que, hasta donde sabemos, tienen una versión DEMO disponible en línea de manera gratuita. La mayoría de estos programas, al tener fines comerciales (Shaalán, 2014, p. 491), resultan inaccesibles o inadecuados para la investigación, o simplemente no contienen el módulo que aquí interesa (reconocimiento de entidades nombradas).

Entre los investigadores sobre el RENA en corpus monolingües destacan Benajiba, Shaalan, Darwish y sus respectivos equipos (Abdallah et al., 2012; Abdelali et al., 2016; Benajiba, 2009; Benajiba et al., 2007; Benajiba & Rosso, 2008; Darwish, 2014; Darwish & Gao, 2014; Farghaly & Shaalan, 2009; Shaalan, 2014; Shaalan & Raza, 2009; Zitouni & Benajiba, 2014). La mayoría de estas investigaciones propone mejoras en los métodos de RENA mediante sistemas basados en reglas, de aprendizaje automático o híbridos. Shaalan (2014) revisa de manera exhaustiva el estado de la cuestión hasta la fecha de su publicación (definición de RENA, dificultades, sistemas de localización y disponibilidad de corpus anotados y herramientas).

La literatura sobre el REN en corpus multilingües en árabe, inglés y español se limita a Samy (véase [II.1.c](#)), quien estudió patrones y estrategias en la traducción de siete categorías de ENAMEX: antropónimos, topónimos, acrónimos, eventos, instituciones y entidades políticas y administrativas, y empleos (Samy, 2005). Etiquetó primero las EN en inglés y español, por la disponibilidad de herramientas, y seguidamente en árabe de forma manual, aunque

*i*[I]n spite of the laborious task of manual tagging, the fact that the corpus we are using is parallel made it possible to develop a prototype for an Arabic NE Tagger. As a starting point, this tagger takes as input the Spanish tagged Named Entities and the aligned sentence-pairs, and then it proceeds to find candidates in the corresponding Arabic sentence. (2005, p. 6)

Un ejemplo de desarrollo de sistemas de REN para dominios no periodísticos en árabe, en concreto para microblogs, es Darwish y Gao (2014), quienes señalan, entre otros problemas, el uso incorrecto de mayúsculas en inglés y de abreviaturas.

En 2015 se celebró el *First International Conference on Arabic Computational Linguistics (1st International Conference on Arabic Computational Linguistics, 2015)*, en el que se abordaron las interfaces de lenguaje natural, la detección de plagio, RI y EI, la traducción automática, recursos léxicos y aspectos semánticos, sintácticos y morfológicos. También se describieron las dificultades y las herramientas de RENA (Elsayed & Elghazaly, 2015; Mesmia et al., 2015; Trabelsi et al., 2015), con resultados muy positivos en general y no tan positivos en la categoría «organización» (Trabelsi et al., 2015), en la misma línea de nuestra experiencia con FARASA (véase [III.2.a.iv](#)).

Un avance más reciente en la materia es CasANER, un sistema de reconocimiento y etiquetado de EN árabes basado en un corpus de textos de la Wikipedia en árabe. Este sistema demostró resultados positivos en el reconocimiento de nombres de personas y de organismos (Mesmia et al., 2018).

De esta revisión podemos concluir que los sistemas de REN serían muy útiles en el entorno de las Naciones Unidas al permitir analizar con perspectiva diacrónica el enorme volumen documental presentado por los Estados Miembros. La identificación y recuperación de los nombres de textos normativos y entidades corporativas nacionales sería tan importante para esos documentos como la recuperación de la terminología para los documentos científicos y técnicos (Jacquemin & Bourigault, 2003) y permitiría dar seguimiento a las modificaciones o la ejecución de una ley determinada o a la actuación de una entidad particular. Esto explica el interés de esta investigación por tales sistemas.

Sin embargo, los sistemas de REN apenas son todavía útiles para textos en árabe, con excepciones como FARASA o MADAMIRA, cuya eficiencia real es en realidad cuestionable hasta el momento. Las dificultades señaladas para el PLN en árabe no son ajenas a esta situación de precariedad.

Además, ninguna de las clasificaciones de EN consultadas considera la categoría «nombre de texto normativo». Como consecuencia, ninguna de las herramientas existentes permite localizar esa categoría, con independencia de la lengua.

## Capítulo III: Corpus y metodología

### 1. Corpus

En el capítulo 1 adelantábamos que las unidades de análisis se extrajeron de un corpus paralelo (en adelante, «el corpus paralelo») compuesto por documentos originales árabes y sus traducciones al español y el inglés. Los documentos originales del corpus paralelo proceden de un corpus previo (en adelante, el «corpus original») compuesto por los informes (en árabe, francés e inglés) presentados entre 2000 y 2018 al Grupo de Trabajo sobre el EPU y a los comités de tratados por los Estados Miembros pertenecientes a la Liga de los Estados Árabes. Puesto que las unidades de análisis en esta investigación son EN árabes, procedimos a descartar del corpus original todos los informes presentados en francés e inglés. El resultado fue un subcorpus (en adelante, «el subcorpus») de documentos originales árabes. Posteriormente aplicamos al subcorpus diversos criterios de representatividad y seleccionamos una muestra. Esa muestra, y las traducciones al español e inglés de los documentos que la conforman, constituye el corpus paralelo empleado en esta tesis y del que se extrajeron las unidades de análisis.

La primera parte de este capítulo describe en detalle las características del corpus original ([III.1.a](#)), los criterios de selección del subcorpus ([III.1.b](#)), la muestra ([III.1.c](#)) y, por último, la recopilación del corpus paralelo ([III.1.d](#)).

#### a) El corpus original

##### i. Características generales

El corpus original está compuesto por los informes iniciales y periódicos de Estado (en árabe, francés e inglés) presentados entre 2000 y 2018 por los Estados Miembros pertenecientes a la Liga de los Estados Árabes a los comités de tratados y al Grupo de

Trabajo sobre el EPU. La elección de estos informes, con exclusión de otros documentos (por ejemplo, relativos a los procedimientos de denuncia), se explica en [I.3.c](#). El corpus fue compilado por Roldán y en él se basan Roldán y Feria (en prensa). En adelante, toda la información sobre el «corpus original» se basa en comunicaciones personales de Roldán y Feria.

Roldán tomó el año 2000 como fecha de inicio de la recopilación, como hacen también otros estudios basados en córpora de la ONU, debido a que las versiones electrónicas previas a ese año solo están disponibles en formato PDF imagen (Guzmán & Prieto Ramos, 2021; Rafalovitch & Dale, 2009, p. 294). La última actualización del corpus original se realizó el 14 de marzo de 2019.

El corpus no está sujeto a derechos de propiedad industrial e intelectual y cumple con los requisitos fundamentales de fiabilidad (Bowker y Pearson 2002, p. 9): es auténtico, los documentos están disponibles en formato electrónico, es extenso y fue seleccionado mediante criterios específicos.

Los informes se descargaron de los sitios web del ACNUDH<sup>45</sup> y del Consejo de Derechos Humanos<sup>46</sup>. Cuando en esas páginas no se halló versión en formato texto (Microsoft Word) se acudió a [undocs.org](http://undocs.org), que permite acceder con toda facilidad al Sistema de Archivo de Documentos (*Document Storage System*) y descargar los informes en Word. En el 2 % de los informes del corpus original tampoco ahí pudo localizarse la versión Word (acaso porque los Estados implicados nunca la remitieron) y solo se tuvo acceso al informe en cuestión en formato PDF imagen.

Los informes presentados a los comités de tratados y al Grupo de Trabajo sobre el EPU comprenden una portada con información fundamental sobre el documento: el tratado internacional de referencia, la signatura asignada, la fecha de recepción y publicación, la lengua original, el órgano receptor, el tipo de informe y aclaraciones adicionales (véanse los ejemplos presentados en el [Anexo 3](#)). Esa información permitió identificar y clasificar los informes.

---

<sup>45</sup> <https://tbinternet.ohchr.org/SitePages/Home.aspx> [Última consulta: 12 de enero de 2019].

<sup>46</sup> <https://www.ohchr.org/SP/HRBodies/UPR/Pages/Documentation.aspx> [Última consulta: 13 de febrero de 2019].

En ocasiones, las portadas de un mismo conjunto documental (original y sus traducciones) incluyen erratas o información contradictoria, en particular en relación con la lengua original. Esto supuso a Roldán ciertos problemas de clasificación. Cuando en el original se atribuye el inglés como lengua original y todas sus traducciones le atribuyen el árabe, esta segunda lengua se consideró la original. Cuando del informe solo existían una versión en inglés y otra en árabe, y ambas afirmaban ser traducciones, se comprobó cuál era la lengua de redacción más utilizada por el Estado y se le asignó como lengua original.

Por otro lado, los Estados pueden presentar sus informes en dos lenguas. Roldán consideró bilingüe el informe cuya portada especifica que se presentó en dos lenguas, siempre que en efecto se encontraran disponibles ambas lenguas; en caso contrario, incluso si la portada los califica de bilingües, se consideraron monolingües. El corpus original incluye los dos originales de todos los informes bilingües.

Se excluyeron del corpus original los documentos siguientes:

- Informes sin portada;
- Informes sin signatura (por ejemplo, los anexos);
- Informes incompletos (por ejemplo, CRPD/C/QAT/CO/1/Add.2), e
- Informes en formato PDF con mala calidad de imagen (por ejemplo, CRC/C/OPSC/SYR/Q/1/Add.1).

Calcular el tamaño del corpus original generó más problemas de lo previsto. En general, el tamaño de un corpus se mide en palabras gráficas. Este concepto se definió por primera vez en 1967 como «a string of contiguous alphanumeric characters with white spaces on either side; [it] may include hyphens and apostrophes, but no other punctuation marks» (Kučera & Francis, 1967, citado en Manning & Schütze, 1999, p. 125). Las palabras gráficas del corpus original se calcularon de manera automática mediante Microsoft Word 16.22 cuando pudo descargarse el informe en formato texto.

En los documentos en formato PDF, Roldán calculó el número de palabras gráficas por aproximación debido al elevado precio de los programas de reconocimiento óptico de caracteres árabes y su limitada efectividad (Alghamdi & Teahan, 2017). Según el criterio de la ONU, una página en el formato usual de la Organización comprende 330 palabras gráficas. Sin embargo, el corpus original demostró que esa norma no se ajusta a la realidad

de los informes nacionales de seguimiento, que tienen una media 288,5 palabras gráficas por página en árabe, 409,6 en inglés y 420,7 en francés.

El problema es que el concepto de palabra gráfica resulta inapropiado para realizar comparaciones y cálculos interlingüísticos debido al carácter arbitrario de los sistemas de escritura (Grzybek, 2007, p. 18). La aglutinación gráfica de los clíticos en árabe, en particular, hace imposible cualquier comparación interlingüística basada en la palabra gráfica, ya que los clíticos constituyen por su frecuencia «a basic characteristic of the Arabic language» (Neifar et al., 2018, p. 247). Era necesario, por tanto, realizar un incremento porcentual en el número de palabras árabes.

Para calcular el tamaño del corpus original, Roldán siguió el criterio de incremento porcentual en el número de palabras gráficas del árabe aplicado primeramente por García-Barrero, Feria y Turell (2013) en el ámbito de la lingüística forense. Esos autores se basaron para ello en las conclusiones de Alotaiby, Alkharashi y Foda (2014, 2010, 2009), quienes, tras tokenizar un corpus de 600 millones de palabras árabes de género periodístico (Arabic Gigaword) utilizando el tokenizador AMIRA 2.0, determinaron que el número de palabras gráficas en árabe se incrementa el 41,3 % al desaglutinar los clíticos. En esta tesis también aplicamos ese porcentaje de incremento. Una vez aplicado ese criterio, la media de palabras árabes por página en los documentos del corpus original pasó de 288,5 a 407,7, lo que se aproxima en gran medida a las medias del inglés y el francés (409,6 y 420,7, respectivamente). A nuestro juicio, ello prueba la validez del criterio adoptado.

Aplicando ese incremento porcentual, inclusive a los informes cuya extensión se calculó por aproximación, el corpus original comprende 9 201 260 palabras: 7 002 227 en árabe, 1 871 236 en francés y 327 797 en inglés (76 %, 20 % y 4 %, respectivamente).

A continuación, describimos el corpus original atendiendo a varios parámetros: Estado Miembro, órgano receptor en la ONU y tipo de informe. Esa información será fundamental para demostrar la posibilidad de seleccionar un subcorpus suficiente ([III.1.b](#)) y para la selección de la muestra ([III.1.c](#)).



**ii. Datos desglosados según Estado, lengua, número de informes y número de palabras**

La participación de cada Estado en el corpus puede medirse en número de informes y en número de palabras. El primer criterio revela en particular el grado de cumplimiento del Estado con su obligación de informar, aunque resulta de menor interés para las investigaciones lingüísticas.

El gráfico 2 describe el corpus original con datos desglosados por Estado, número de informes y lengua.

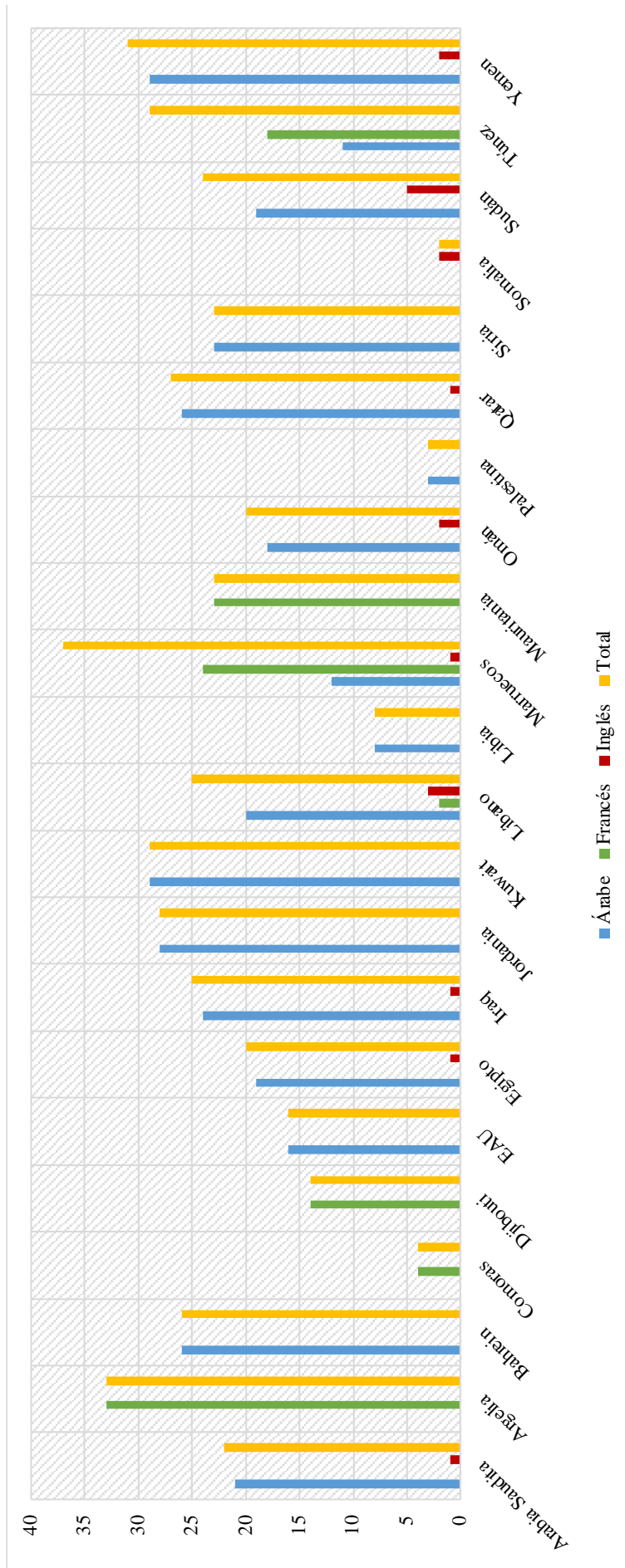


Gráfico 2: Descripción del corpus con datos desglosados por Estado, número de informes y lengua

Fuente: Datos proporcionados por Roldán y Feria. Elaboración propia



Marruecos, Argelia y el Yemen fueron los Estados que mayor número de informes presentaron (37, 33 y 31, respectivamente).

En relación con las lenguas de redacción se observa lo siguiente:

- 2 Estados (Líbano y Marruecos) utilizaron las 3 lenguas en la redacción de sus informes;
- 7 Estados (Bahrein, Emiratos Árabes Unidos, Jordania, Kuwait, Libia, Palestina y Siria) presentaron sus informes exclusivamente en árabe;
- 4 Estados (Argelia, Comoras, Djibouti y Mauritania) presentaron sus informes exclusivamente en francés;
- 1 Estado (Somalia) empleó exclusivamente el inglés.

En relación con el número de informes presentados en cada lengua se observa lo siguiente:

- Kuwait, el Yemen y Jordania fueron los Estados que mayor número de informes presentaron en árabe (29, 29 y 28, respectivamente);
- Argelia, Marruecos y Mauritania fueron los que mayor número de informes presentaron en francés (33, 24 y 23, respectivamente), y
- Sudán y Líbano fueron los Estados que mayor número de informes presentaron en inglés (5 y 3, respectivamente).

La extensión de los distintos tipos de informes es heterogénea (véase [I.3.b-d](#)), por lo que un mayor número de informes no implica necesariamente una mayor cantidad de palabras, que es lo que a nosotros realmente nos interesa para conformar nuestro subcorpus. El gráfico 3 describe el corpus original con datos desglosados por Estado, número de palabras presentadas y lengua.

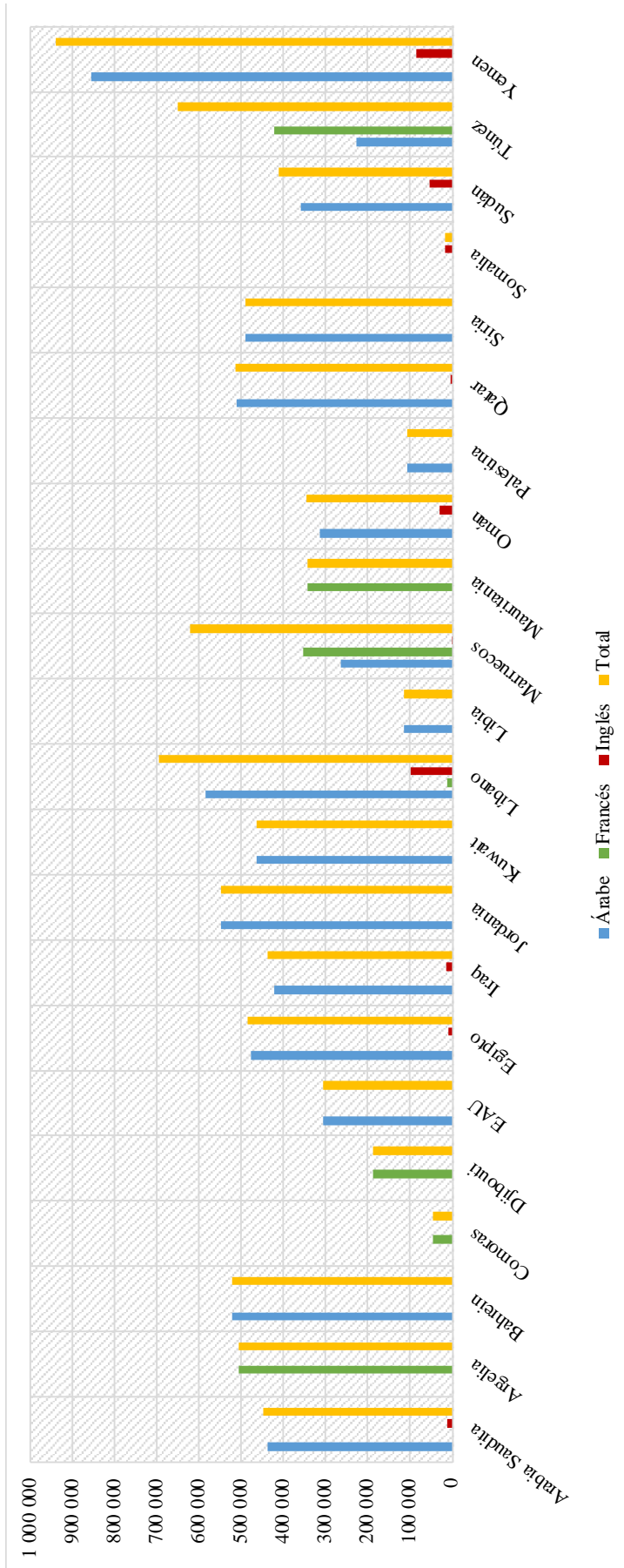


Gráfico 3: Descripción del corpus con datos desglosados por Estado, número de palabras y lengua

Fuente: Datos proporcionados por Roldán y Feria. Elaboración propia

El Estado que mayor número de palabras presentó en el periodo analizado fue el Yemen (939 160). Lo siguen el Líbano (694 543) y Túnez (649 675). Los Estados que mayor volumen de palabras presentaron en cada lengua de redacción fueron los siguientes:

- En árabe, el Yemen presentó 854 577, seguido del Líbano (584 277) y Jordania (548 339).
- En francés, Argelia presentó 506 233, seguido de Túnez (423 071) y Marruecos (353 715).
- En inglés, el Líbano presentó 98 973, seguido del Yemen (84 583) y Sudán (53 770).

Por tanto, en efecto, un mayor número de informes no implica necesariamente un mayor número de palabras presentadas. A la vista de los datos anteriores se comprueba que:

- Sin considerar la lengua de redacción, los 6 informes de diferencia entre Marruecos y el Yemen suponen un contraste de más de 300 000 palabras a favor del último;
- En el caso del árabe, con el mismo número de informes presentados, los documentos del Yemen casi doblan el número de palabras de Kuwait;
- En el caso del francés, Marruecos presentó 6 informes más que Túnez, pero Túnez presentó 70 000 palabras más y
- En el caso del inglés, las cifras son demasiado reducidas para establecer cualquier tipo de comparación relevante.

Estos datos, particularmente los relativos a la lengua árabe, son fundamentales para la selección de nuestra muestra.

La tendencia al empleo de una u otra lengua ha variado a lo largo del periodo analizado (véase gráfico 4). El uso de la lengua árabe se incrementa con el paso del tiempo y esa tendencia se agudiza. Contrasta, en paralelo, el empleo del inglés, que tiende a cero. El número de palabras presentadas en francés experimenta un leve aumento, pero muy lejos del incremento que se registra en árabe.

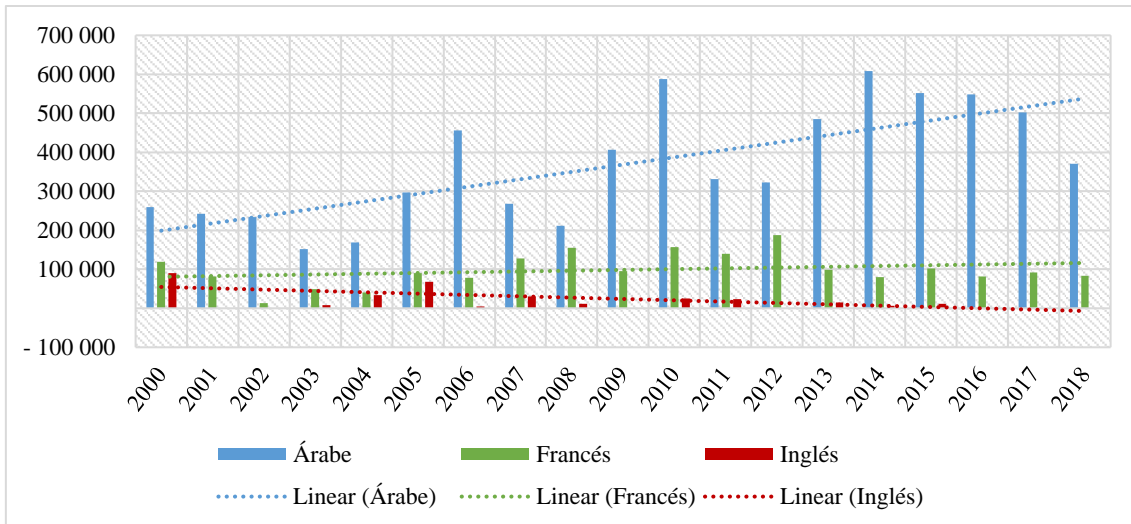


Gráfico 4: Volumen de palabras por año y por lengua (barras verticales) y tendencia del uso de las lenguas (línea de puntos)

Fuente: Datos proporcionados por Roldán y Feria. Elaboración propia

### iii. Datos desglosados según el órgano receptor en la ONUG

Del corpus original también se pueden extraer datos relativos al número de informes y palabras dirigidos a cada mecanismo de seguimiento. El gráfico 5 ilustra el número de informes incluidos en el corpus y dirigidos a cada comité y al Grupo de Trabajo sobre el EPU.

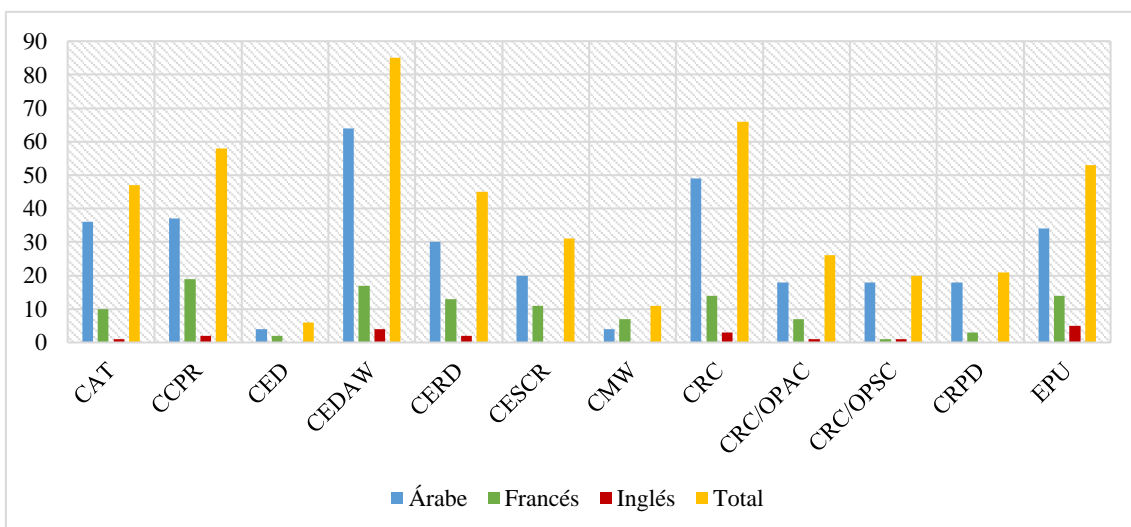


Gráfico 5: Número de informes totales y por lengua presentados a cada órgano

Fuente: Datos proporcionados por Roldán y Feria. Elaboración propia

CEDAW y CRC fueron los mayores receptores con 85 y 66 informes respectivamente. En el otro extremo se encuentran el CMW y el CED, con 11 y 6 informes respectivamente.

En el caso de los informes presentados en árabe, el patrón se repite: CEDAW y CRC fueron los que más recibieron (64 y 49, respectivamente) y CMW y CED los que menos (4 en los dos casos). Para el francés, en cambio, observamos ligeras variaciones, sin que ningún comité destaque sobre los demás. En el caso del inglés, el número de informes presentados es muy reducido, y en algunos comités es cero (CED, CMW y CRPD).

Las grandes diferencias señaladas se deben a los tres factores siguientes:

- Tiempo de vigencia del tratado (véase [tabla 1](#)): a mayor número de años de vigencia, mayor número de informes presentados;
- Tiempo transcurrido desde la adhesión o ratificación: cuanto menor tiempo transcurrido desde la ratificación por un Estado, menor número de informes por él presentados;
- Duración de los ciclos: cuanto menor el tiempo entre los ciclos, lo que se determina en cada tratado, mayor el número total de informes.

Esas tendencias generales pueden no cumplirse en determinados casos por demoras de los Estados o solicitudes de informes consolidados de los comités<sup>47</sup>. Por otra parte, como quedó demostrado, la extensión variable de los informes no permite establecer una relación directa entre el número de informes y el número de palabras dirigidos a un órgano, dato que nos resulta de mayor interés para compilar el subcorpus. El gráfico 6 muestra la distribución del número total de palabras presentadas a cada órgano con datos desglosados por lengua.

---

<sup>47</sup> Comunicación personal de Manuel Feria García (23/12/2021).



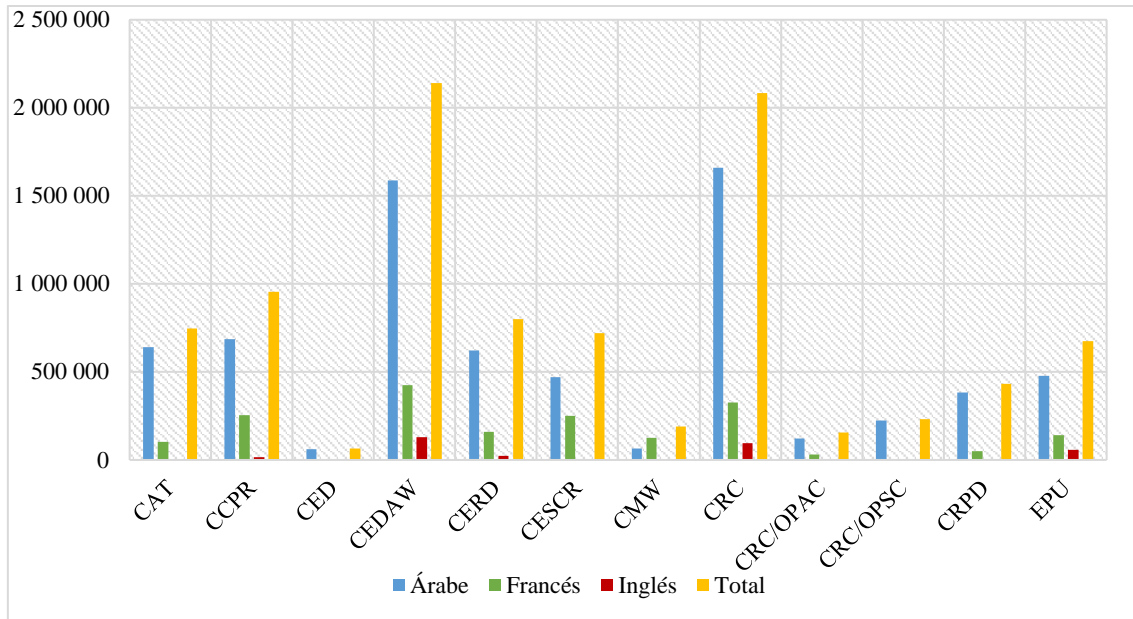


Gráfico 6: Distribución de palabras totales y por lengua dirigidas a cada órgano

Fuente: Datos proporcionados por Roldán y Feria. Elaboración propia

Los comités que mayor volumen de palabras recibieron fueron CEDAW y CRC. En total (suma de todas las lenguas) cada uno de estos comités recibió más de 2 millones de palabras, y solo en árabe cada uno recibió más de 1,5 millones. Las cifras para el francés y el inglés en estos comités son muy bajas en comparación con el árabe.

A la vista de los gráficos 5 y 6 conviene resaltar la gran diferencia entre número de informes y de palabras: si bien el CEDAW recibió 19 informes más que el CRC, este recibió 70 000 palabras más que el primero.

#### iv. Datos desglosados según el tipo de informe

El gráfico siguiente muestra la distribución del número de palabras totales, con datos desglosados por lengua, según el tipo de informe incluido en el corpus original (véase [I.3.b-d](#)).

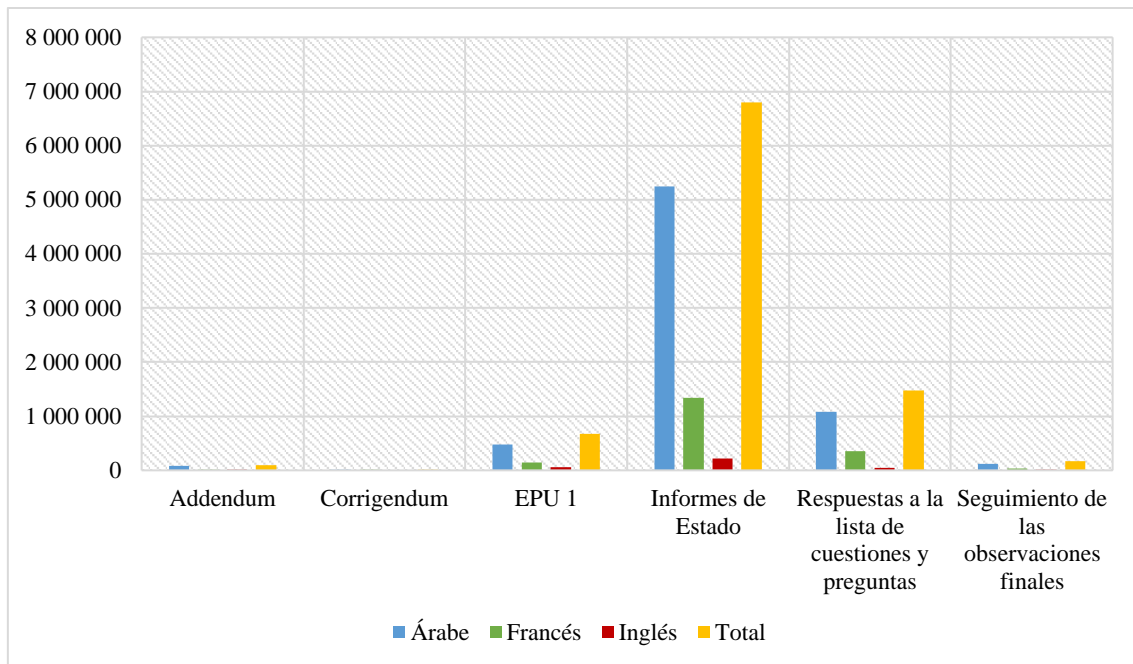


Gráfico 7: Número de palabras por tipo de informe, con datos desglosados por lengua, en el corpus original

Fuente: Datos proporcionados por Roldán y Feria. Elaboración propia

Los Estados presentaron más palabras en sus informes de Estado (74 % del total del corpus original) que mediante todos los demás tipos de informes juntos. Esto se puede explicar por varios motivos:

- Se recomienda que los informes de Estado no sean excesivamente largos y que la información se redacte de manera «concisa y estructurada». En particular, se recomienda una extensión máxima de 60 páginas para los informes iniciales de Estado y de 40 para los documentos periódicos ulteriores (Naciones Unidas, 2009, pp. 7, párr. 19). El número de páginas recomendado para los informes dirigidos al Grupo de Trabajo sobre el EPU es menor: de 10 a 20 (Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos, s. f., p. 14; Villán Durán, 2016, p. 137). A pesar de estas recomendaciones, el corpus original revela que en la mayoría de los casos la extensión de los informes excede a la recomendada, especialmente en los informes de Estado. En el corpus original, el documento más extenso en árabe dirigido a los comités derivados de tratados tiene 217 páginas y el más extenso en árabe dirigido al Grupo de Trabajo sobre el EPU, 42 páginas;
- Los informes dirigidos al Grupo de Trabajo sobre el EPU son más escasos debido a la corta vida activa de este mecanismo de seguimiento;

- Si bien los informes de Estado son obligatorios, algunos tipos de informe, como el *Addendum* o el *Corrigendum*, son infrecuentes.

Se deduce del gráfico 7 que los informes iniciales y periódicos de Estado redactados en árabe son el grueso del corpus original (5 245 096 palabras, es decir, el 57 %). Este dato, como veremos en los siguientes apartados, es crucial para la selección del subcorpus.

#### **v. Recapitulación**

El corpus original cumple los criterios de fiabilidad, puesto que los textos son auténticos, de libre acceso y seleccionados en función de criterios específicos (informes de derechos humanos presentados entre 2000 y 2018 por los Estados Miembros pertenecientes a la Liga de los Estados Árabes a los comités de tratados y al Grupo de Trabajo sobre el EPU), están disponibles en formato electrónico y el volumen de palabras es suficiente (9 millones).

El 76 % del corpus original son palabras árabes y el 57 % informes iniciales o periódicos de Estado originalmente en árabe. El número de palabras en francés e inglés es mucho menor y la tendencia es a que los Estados árabes dejen de emplear el inglés como lengua de comunicación (Gráfico 4). Los órganos receptores más destacables son el CRC y el CEDAW y, lógicamente, se registra una relación proporcional directa entre el volumen de palabras recibido y la antigüedad del comité.

La siguiente fase de nuestra investigación consistió en, a la vista de los datos presentados en este apartado, seleccionar un subcorpus que se ajustara a los requerimientos impuestos por nuestras hipótesis ([I.4.b](#)). La selección se describe en el apartado siguiente.

#### **b) Selección del subcorpus**

El corpus original no presenta las características adecuadas para corroborar o refutar nuestras hipótesis nulas por incluir informes originalmente escritos en inglés y francés, cuando nuestras unidades de análisis son EN árabes. Fue preciso, por tanto, disponer de un subcorpus con originales exclusivamente en árabe. Las características del corpus original, antes descritas, en particular su volumen, nos permitieron excluir los originales

en francés e inglés (1 871 236 y 327 797 palabras, respectivamente) y el corpus solo se redujo de 9 201 260 a 7 002 227 palabras.

### **c) Selección de la muestra**

Resultó materialmente imposible analizar el subcorpus en su totalidad debido a sus dimensiones y a las limitaciones de la gestión automática de los datos. Por ello, se seleccionó una muestra suficiente. A continuación se describen los criterios de selección.

#### **i. Criterio 1: Estado**

Debía decidirse en primer lugar si emplear unidades de análisis de uno o más Estados y, en ese caso, de cuántos Estados. Si la muestra comprendiera documentos de más de un Estado el análisis podría ser interestatal (analizando la homogeneidad traductora en el conjunto de los Estados seleccionados) o intraestatal (analizando la homogeneidad traductora en cada Estado). Sin embargo, incluso si dos EN nacionales coincidieran en dos Estados, su traducción no tendría necesariamente que ser homogénea en los conjuntos documentales de ambos. Por tanto, no era posible abordar la homogeneidad traductora interestatal con los elementos seleccionados: de emplear documentos de más de un Estado las conclusiones solo podrían ser intraestatales.

La ventaja del análisis intraestatal de varios Estados es que la coincidencia de los resultados proporcionaría solidez estadística al conjunto. Sin embargo, los factores que podrían afectar a la homogeneidad traductora en el producto no dependen del Estado, sino del proceso de traducción y de los agentes implicados, que por definición son los mismos para todos los Estados (véase [I.3.f](#)). Por tanto, con igual número de palabras, la diferencia entre realizar un análisis intraestatal de dos Estados o de un Estado estribaría en que el segundo proporcionaría conclusiones más sólidas, ya que implicaría un mayor número de palabras presentadas por un mismo Estado y, en consecuencia, un mayor número de unidades de análisis. Dicho de otro modo, con igual número de palabras, el análisis intraestatal de dos Estados permitiría realizar una comparación de datos individuales menos sólidos. Por tanto, el análisis intraestatal de dos Estados solo sería ventajoso si se multiplicara por dos la muestra mínima representativa para un Estado, lo que resultaba

imposible debido a las dificultades que justamente nos obligaron a seleccionar una muestra.

Por todo ello se decidió seleccionar la muestra de un solo Estado. El siguiente paso fue elegir el Estado con mayor representación en el corpus y seleccionar de entre sus documentos la muestra que mejor diera respuesta a nuestras hipótesis. Puesto que la República del Yemen es el Estado con mayor representación en el subcorpus (854 577 palabras, en torno al 12 % del total), y los Estados que le siguen, el Líbano y Jordania, están representados, respectivamente, con el 68 % y el 64 % menos de palabras, la elección era obvia.

Antes de continuar con los criterios de selección conviene hacer algunas aclaraciones respecto al Estado elegido. La República Árabe de Yemen se convirtió en Estado Miembro de las Naciones Unidas en 1947 y la República Democrática Popular de Yemen, en 1967. Ambos Estados se unieron en 1990 para constituir la República del Yemen<sup>48</sup>. Por tanto, teniendo en cuenta que el corpus contiene documentos a partir de 2000, el Yemen solo participa en esta investigación como Estado unificado. En la actualidad, la República del Yemen es Estado parte en siete tratados y tres protocolos facultativos<sup>49</sup>.

Los gráficos 8 y 9 presentan el número de palabras y el número de informes de la República del Yemen que integran el subcorpus.

---

<sup>48</sup> <http://www.un.org/es/member-states/> [Última consulta: 12 de diciembre de 2018]

<sup>49</sup> Estos son: Convención Internacional sobre la Eliminación de Todas las Formas de Discriminación Racial; Convención sobre la Eliminación de Todas las Formas de Discriminación contra la Mujer; Pacto Internacional sobre Derechos Civiles y Políticos; Pacto Internacional de Derechos Económicos, Sociales y Culturales; Convención sobre los Derechos del Niño; Convención contra la Tortura y Otros Tratos o Penas Cruelles, Inhumanos o Degradantes; Protocolo Facultativo de la Convención sobre los Derechos del Niño relativo a la venta de niños, la prostitución infantil y la utilización de niños en la pornografía; Protocolo Facultativo de la Convención sobre los Derechos del Niño relativo a la participación de niños en los conflictos armados; Convención Internacional sobre los Derechos de las Personas con Discapacidad; y Protocolo Facultativo de la Convención sobre los Derechos de las Personas con Discapacidad. Consultado en: <https://indicators.ohchr.org> [Última consulta: 09 de octubre de 2021]

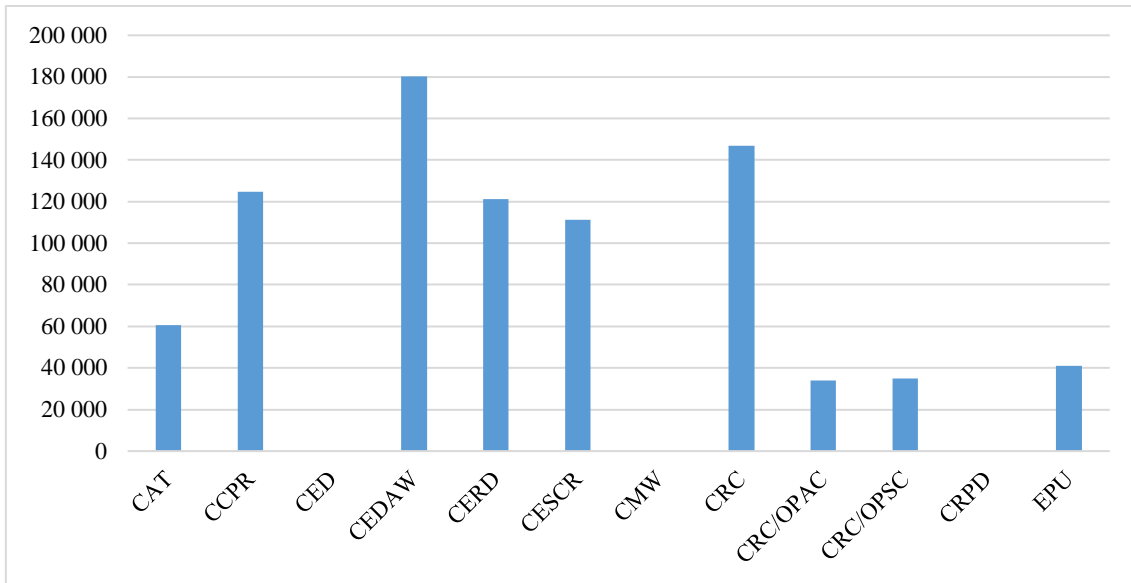


Gráfico 8: Número de palabras presentadas por la República del Yemen que integran el subcorpus, con datos desglosados por órgano receptor

Fuente: Datos proporcionados por Roldán y Feria. Elaboración propia

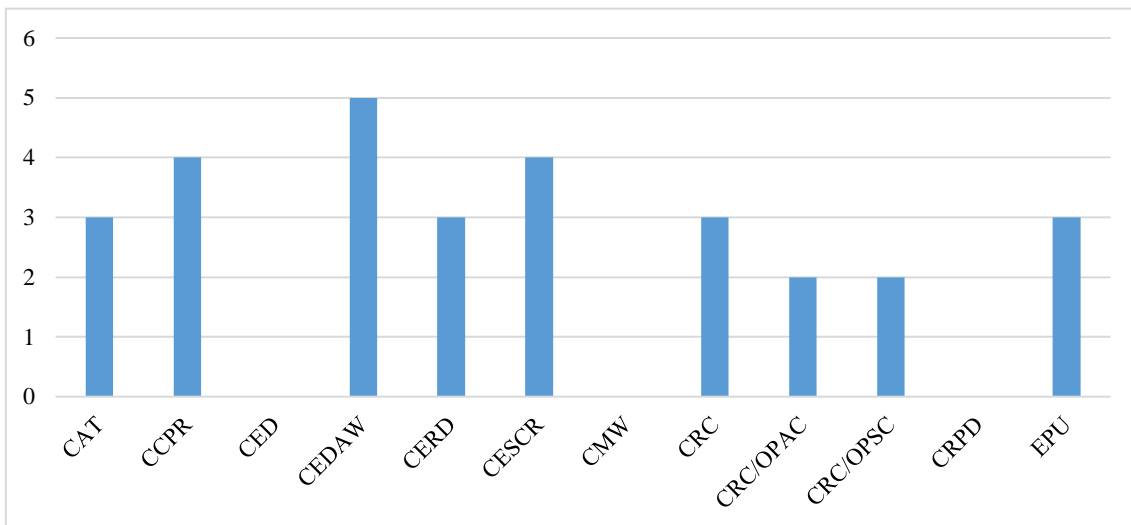


Gráfico 9: Número de informes presentados por la República del Yemen que integran el subcorpus, con datos desglosados por órgano receptor

Fuente: Datos proporcionados por Roldán y Feria. Elaboración propia

De nuevo los protagonistas son CEDAW y CRC. No hay informes presentados al CED y al CMW porque el Yemen no ha ratificado los tratados pertinentes.

**ii. Criterio 2: Tipo de informe**

Recordemos que para testar las hipótesis secundarias (véase [I.4.b](#)) es tan importante disponer de un número suficiente de informes como que estos contengan un número suficiente de palabras. El subcorpus contiene 29 documentos del Yemen: 19 informes iniciales o periódicos de Estado, 5 respuestas a las listas de cuestiones, 3 informes EPU 1 y 2 informes de seguimiento a las observaciones finales (para las características de cada tipo documental, véase [I.3.b-d](#) y [I.4.a](#)).

Los informes de Estado, además de ser los documentos más numerosos (cuatro veces más que el siguiente tipo documental), presentan las ventajas siguientes:

- Pertinencia del contenido: esos informes detallan las medidas adoptadas para dar efecto a los derechos reconocidos en los tratados internacionales e informar sobre el progreso realizado (Asamblea General de las Naciones Unidas, 1989, art. 44). Las directrices pertinentes de la Organización especifican que los informes «no deberán limitarse a presentar listas o descripciones de los instrumentos jurídicos aprobados en el país en los últimos años, sino que deberán indicar cómo se reflejan esos instrumentos jurídicos en la realidad económica, política, social y cultural y las condiciones existentes en el país» (Naciones Unidas, 2009, párr. 25). La presencia de un elevado número de EN, por tanto, podía darse por segura;
- Extensión: hemos comprobado que la extensión de estos documentos es por lo general mayor (y así lo afirman también Guzmán y Prieto Ramos (2021, p. 18)). De hecho, los 19 informes de Estado representan el 85 % de las palabras del Yemen incluidas en el subcorpus. La mayor extensión, unida a la pertinencia del contenido, multiplica el número previsible de unidades de análisis concurrentes.

Los otros documentos, por su parte, presentan las desventajas siguientes:

- Informes EPU 1: comprenden un número de palabras muy reducido (40 911) y su carácter generalista, frente al detalle del informe nacional, genera pocas expectativas de hallar numerosas unidades de análisis;
- Listas de cuestiones e informes de seguimiento de las observaciones finales: su extensión es muy reducida (<10 000 palabras) y su número muy escaso.

Por todo ello, los candidatos a formar parte de la muestra eran los 19 informes de Estado. De ellos, descartamos 3 por encontrarse en formato PDF. La tabla siguiente presenta los 16 documentos restantes.

Comité	Signatura	Número de palabras	Ciclo y año
CAT	CAT/C/16/Add.10	19 019	I – 2003
CAT	CAT/C/YEM/2	37 819	II – 2009
CCPR	CCPR/C/YEM/2001/3	18 673	III – 2001
CCPR	CCPR/C/YEM/2004/4	58 118	IV – 2005
CCPR	CCPR/C/YEM/5	43 616	V – 2010
CEDAW	CEDAW/C/YEM/7-8	47 449	VII-VIII - 2014
CERD	CERD/C/362/Add.8	23 635	XI-XIV - 2002
CERD	CERD/C/YEM/16	50 975	XV-XVI - 2006
CERD	CERD/C/YEM/17-18	46 540	XVII-XVIII - 2010
CESCR	E/1990/5/Add.54	24 647	I – 2002
CESCR	E/C.12/YEM/2	30 493	II – 2009
CESCR	E/C.12/YEM/3	21 805	III - 2014
CRC	CRC/C/129/Add.2	38 271	III - 2004
CRC	CRC/C/YEM/4	83 483	IV - 2012
CRCOPAC	CRC/C/OPAC/YEM/1	29 988	I - 2013
CRCOPSC	CRC/C/OPSC/YEM/1	23 847	I - 2009
		<b>598 379</b>	

Tabla 2: Informes iniciales y periódicos de Estado del subcorpus candidatos a formar parte de la muestra, con datos desglosados por órgano receptor, número de palabras, ciclo y año de presentación

Fuente: Datos proporcionados por Roldán y Feria. Elaboración propia

Finalmente, de entre esos documentos se eligió la muestra considerando el tercer y último criterio de inclusión.

### iii. Criterio 3: Ciclos y comités

Para testar la hipótesis secundaria 1 se seleccionaron los comités con al menos dos informes disponibles. Cuando un comité recibió más de dos informes se eligieron el de mayor antigüedad y el más reciente. En aplicación de lo anterior, se excluyeron de la



muestra los dos protocolos facultativos y el informe del CEDAW, por disponerse solo de un informe de cada tipo, y los informes intermedios de los comités CCPR, CESCR y CERD.

Como resultado, la muestra quedó compuesta por los 10 documentos que figuran en la tabla siguiente.

Comité	Signatura	Número de palabras	Ciclo y año
CAT	CAT/C/16/Add.10	19 019	I – 2003
	CAT/C/YEM/2	37 819	II – 2009
CCPR	CCPR/C/YEM/2001/3	18 673	III – 2001
	CCPR/C/YEM/5	43 616	V – 2010
CERD	CERD/C/362/Add.8	23 635	XI-XIV – 2002
	CERD/C/YEM/17-18	46 540	XVII-XVIII – 2010
CESCR	E/1990/5/Add.54	24 647	I – 2002
	E/C.12/YEM/3	21 805	III – 2014
CRC	CRC/C/129/Add.2	38 271	III – 2004
	CRC/C/YEM/4	83 483	IV – 2012
		<b>357 509</b>	

Tabla 3: Informes incluidos en la muestra

Fuente: Datos proporcionados por Roldán y Feria. Elaboración propia

A fin de facilitar la lectura, en adelante nos referiremos a los informes y comités de la muestra, de manera abreviada y simplificada, como se muestra en la tabla siguiente.

Acrónimo oficial del comité	Denominación simplificada del comité	Signatura oficial del informe	Denominación simplificada del informe
CAT	Comité A	CAT/C/16/Add.10	A1
		CAT/C/YEM/2	A2
CCPR	Comité B	CCPR/C/YEM/2001/3	B1
		CCPR/C/YEM/5	B2
CERD	Comité C	CERD/C/362/Add.8	C1
		CERD/C/YEM/17-18	C2
CESCR	Comité D	E/1990/5/Add.54	D1
		E/C.12/YEM/3	D2
CRC	Comité E	CRC/C/129/Add.2	E1
		CRC/C/YEM/4	E2

Tabla 4: Denominación simplificada de los comités e informes de la muestra

La muestra, por tanto, comprende 357 509 palabras, lo que supone un 42 % del total de palabras presentadas por el Estado en lengua árabe y un 5 % del total de palabras árabes del corpus original.

#### d) **Compilación del corpus paralelo**

Los motivos que nos movieron a seleccionar para el análisis el español y el inglés entre las lenguas oficiales de la Organización fueron los siguientes:

- Inglés: La importancia del inglés en la Organización quedó demostrada en los capítulos I y II. Además, el inglés sirve en numerosas ocasiones de lengua puente para realizar relés a otras lenguas oficiales (Departamento de la Asamblea General y de Gestión de Conferencias de las Naciones Unidas, s. f.-c; Fernández-Vítóres, 2014, p. 10; Sección de Traducción al Español - Paginilla, s. f.-a);
- Español: Es la lengua oficial de 21 Estados del mundo (Moreno Fernández & Otero Roth, 2016, p. 32) y la tercera lengua con mayor presencia en la ONU (Fernández-Vítóres, 2014, p. 21). Según Guzmán y Prieto Ramos, el Servicio de Traducción al español lidia con más de 7 000 documentos al año (unos 44 millones de palabras) y, frente a lo que ocurre en inglés, lengua en la que se redactan nueve de cada diez

documentos de la Organización (2021, p. 4), en la ONU raramente se redactan originales en español, por lo que el español se ha descrito como «lengua de traducción» (Fernández-Vítóres, 2014, p. 10). Su inclusión en el análisis, además, podría permitir analizar el posible impacto del relé.

Una vez definida la muestra y determinadas las lenguas de análisis localizamos las traducciones correspondientes al inglés y al español para formar el corpus paralelo, que se describe en la tabla 5.

Informe	Palabras en árabe	Palabras en español	Palabras en inglés
A1	19 019	18 760	18 267
A2	37 819	38 869	34 511
B1	18 673	18 973	18 501
B2	43 616	45 584	42 263
C1	23 635	24 798	22 379
C2	46 540	46 841	42 705
D1	24 647	24 505	22 488
D2	21 805	23 888	19 783
E1	38 271	40 065	35 530
E2	83 483	86 665	71 685
Total	357 509	368 948	328 112
<b>TOTAL</b>	<b>1 054 569</b>		

Tabla 5: Corpus paralelo, con datos desglosados por número de palabras

Fuente: Datos proporcionados por Roldán y Feria. Elaboración propia

El corpus paralelo, por tanto, está compuesto por 1 054 569 palabras, comprende los informes seleccionados para la muestra y sus traducciones al español y el inglés y, a nuestro juicio, permite testar las hipótesis de investigación.

## 2. Metodología

En la segunda parte de este capítulo se definen las unidades de análisis y se describen en detalle su localización, selección y extracción, así como los análisis aplicados a los datos.

### a) Unidades de análisis

#### i. Definición de texto normativo y entidad corporativa

En esta investigación se analizan las dos categorías de EN definidas a continuación:

- Texto normativo: «Cualquier acto o documento que, independientemente del órgano del cual emana y del régimen jurídico que lo caracteriza, exprese (o “contenga”) normas generales y abstractas» (Carbonell y Pedroza de la Llave, 2000, p. 15);
- Entidad corporativa: «Una organización o grupo de personas y/o de organizaciones que se identifica por un nombre específico y que actúa, o puede actuar, como una unidad» (International Federation of Library Associations and Institutions, 2016).

#### ii. Variantes completas e incompletas

La determinación de nuestro concepto de unidad de análisis resultó más compleja de lo que cabía esperar a la vista de los enunciados anteriores. Aunque, por definición, una EN no pueda denominarse de maneras diferentes, en la práctica puede aludirse a ella de forma completa o incompleta. Veamos un ejemplo.

Código Civil (Ley N° 19 de 1992)	Civil Code (Act No. 19 of 1992)	<p>القانون المدني الصادر برقم 19 لسنة 1992</p> <p><i>lq`nwn`lmdny`lš`dr brqm 19 lsnī 1992</i></p> <p><i>[Ley civil promulgada con número 19 del año 1992]</i></p>
----------------------------------	---------------------------------	---

Código Civil (Ley N° 19, de 1992)	Civil Code (Act No. 19 of 1992)	القانون المدني رقم 19 لسنة 1992 <i>'lq`nwn`lmdny rqm 19 lsnī 1992</i> [ley civil con número 19 del año 1992]
Código Civil	Civil Code	القانون المدني <i>'lq`nwn`lmdny</i> [ley civil]

Ejemplo 1: Variantes completas e incompletas de EN (informe B1, párrs. 68; 11, 99; 69, 70; respectivamente)

En esta investigación, esas formas potenciales de denotar una EN se denominan «variantes». En el ejemplo anterior, las variantes están ordenadas de mayor a menor compleción y la negrita indica la parte omitida en la variante siguiente.

No sería necesario ahondar en esta cuestión si contásemos con un sistema eficaz de REN y de resolución de correferencia diseñado para las entidades nombradas seleccionadas. Sin embargo, este no es el caso (véase [II.3.c](#)). Por ello, consideraremos cada variante una unidad de análisis distinta; dicho de otro modo, aunque las tres variantes que figuran en el ejemplo 1 correspondan a la misma EN, las analizaremos como unidades computacionalmente diferenciadas.

La decisión anterior, si bien responde a necesidades de orden computacional, se ve también apoyada por el protocolo recogido en el *Manual*. A finales de 2019, el *Manual* determinó (apartado «Nombres geográficos») que se utilizará siempre en español la forma que figure en el original, con independencia de que sea el nombre oficial o el corto, y añadió que «p[P]uede ocurrir también que en el original se utilice una forma no recogida en UNTERM, bien porque el documento no se someta a revisión editorial o porque sea una comunicación de un Estado publicada como documento de las Naciones Unidas [...]. En ese caso, en español seguiremos el original» (Servicio de Traducción al Español de las Naciones Unidas en Nueva York, s. f.).

### iii. Otras consideraciones

Una secuencia de palabras gráficas latinas cuyos elementos porten mayúscula inicial puede conformar una EN simple o una EN calificada, generalmente por unos o varios calificadores de lugar. La actual Carta Magna española, por ejemplo, se denomina «Constitución Española», no «Constitución del Reino de España», mientras que la Constitución de 1931 se llamaba «Constitución de la República Española». «Constitución de la República Española», por tanto, es una EN en su totalidad; también lo es «Constitución Española» (donde «Española» no es un calificador de lugar), mientras que «Constitución de España» es un EN acompañado de un calificador de lugar. En «Constitución Española» es la mayúscula inicial del adjetivo la que nos indica que nos hallamos ante una EN compuesta por un sustantivo y un adjetivo.

En grafía árabe, sin embargo, la inexistencia de mayúsculas supone la imposibilidad de diferenciar gráficamente entre «Constitución Española» y «Constitución española». Así pues, en un texto en grafía árabe, no es el sistema gráfico el que define si un elemento es o no parte de una EN y no es posible saber dónde comienza y acaba una EN, con o sin calificador, sin una labor de documentación. Veamos un ejemplo.

Constitución de la República del Yemen Constitución de la República	Constitution of Yemen Constitution	دستور الجمهورية اليمنية  <i>dstwr 'lġmhwryṭ 'lymnyṭ</i>  [ <i>constitución de la república del yemen</i> ]
Constitución yemení Constitución del país	Yemeni Constitution	الدستور اليمني  <i>'ldstwr 'lymny</i>  [ <i>constitución yemení</i> ]

Ejemplo 2: Fenómenos potenciales de calificación de EN (informe B2 (párrs. 7, 145 y 299) y C2 (párrs. 6, 220 y 236))

En el ejemplo anterior, sin información extratextual, es imposible determinar si «*dstwr 'lġmhwryṭ 'lymnyṭ*» [*constitución de la república del yemen*] y «*'ldstwr 'lymny*» [*constitución yemení*] son una EN simple o una EN calificada.

En estos casos, como en los reseñados en ii), cada variante se considera computacionalmente una unidad de análisis diferenciada.

Por otro lado, en nuestro corpus figuran los órganos jurisdiccionales siguientes:

Traducciones en español	Traducciones en inglés	Transliteración	Original árabe
Tribunal Supremo Corte Suprema	Supreme Court supreme court High court	<i>'lmḥkmī 'l'ly'</i> [tribunal más alto]	المحكمة العليا
tribunales penales	criminal courts	<i>mḥkmī 'lḡn'y't</i> [tribunal de los delitos]	محكمة الجنايات
tribunal de menores tribunales	juvenile courts juvenile court	<i>mḥkmī 'l'hd't</i> [tribunal de los menores]	محكمة الأحداث
tribunal penal	criminal court	<i>'lmḥkmī 'lḡz'yī</i> [tribunal penal]	المحكمة الجزائية
tribunal civil	civil court	<i>'lmḥkmī 'lmdnyī</i> [tribunal civil]	المحكمة المدنية
Tribunal de apelación tribunal de apelación tribunales de apelación Tribunal de Apelación	Court of Appeal appeal court court of appeal courts of appeal Court	<i>mḥkmī 'l'st,n'f</i> [tribunal de la apelación]	محكمة الاستئناف

Ejemplo 3: Órganos jurisdiccionales presentes en la muestra (informes A1, A2, B1, B2, C1, C2, E2)

Puesto que en árabe no existen mayúsculas, no resulta fácil dilucidar si la referencia en árabe a un órgano jurisdiccional en singular aludía a él de forma genérica (por ejemplo, «el tribunal penal», con independencia de cuál fuese) o se trataba de una EN («el tribunal penal», es decir, uno en concreto), con independencia de la interpretación de los

traductores. Ante ello, solo hemos considerado EN el Tribunal Supremo, por ser sin duda alguna un órgano único.

#### iv. Localización de las unidades de análisis

La localización de las unidades de análisis se realizó de manera manual por los motivos que se enuncian a continuación.

Como explicamos previamente ([II.3.b](#)), los programas de reconocimiento automático de EN no contemplan la categoría «texto normativo». Además, los que admiten el reconocimiento de unidades en lengua árabe resultan ineficientes, particularmente en búsquedas bajo la categoría «ORG», en las que, recordemos, el valor-F de los resultados rondaba en el mejor de los casos el 65 % (Trabelsi et al., 2015).

Como se expuso en [II.3.c](#), existen en la actualidad dos programas con una versión demo disponible en línea en los que se puede realizar pruebas de REN en texto árabe: MADAMIRA (Pasha et al., 2014) y FARASA (Abdelali et al., 2016). Procedimos a probar esas versiones demo. Sin embargo, resultó imposible hacerlo con MADAMIRA, puesto que su versión demo presenta una limitación en número de caracteres aún mayor que FARASA<sup>50</sup>, que de por sí solo permite entradas de hasta 400 caracteres. Por ello, nos limitamos a introducir tres párrafos de la muestra en FARASA. Los párrafos se seleccionaron atendiendo a la limitación de caracteres y buscando una nutrida presencia de EN. Los párrafos seleccionados fueron 13 y 5, respectivamente, de los informes A2 y E2.

A continuación presentamos los resultados de la localización (ejemplos 4, 5 y 6). Las EN reconocidas se señalan con negrita.

ولقد حلت هذه اللجنة محل اللجنة الفرعية الدائمة المنبثقة عن اللجنة الوطنية العليا لحقوق الإنسان (سابقا) وتتكون من ممثلين لوزارات (الخارجية، التخطيط والتعاون الدولي، الإعلام، الداخلية، الشؤون الاجتماعية والعمل، العدل، والشؤون القانونية) وممثلين (لمكتب رئاسة الجمهورية، مكتب رئاسة الوزراء، مكتب النائب العام، والجهاز المركزي للأمن السياسي).

<sup>50</sup> <http://qatsdemo.cloudapp.net/farasa/demo.html>



*wlqd hlt hđh 'llġnī mhl 'llġnī 'lfr'yi 'ld' mī 'lmbntqī 'n 'llġnī 'lwtnyi 'l'ly' lġwq 'ly'ns'n (s' bq') wtkwn mn mmṭlyn lwz'r't ('lh' rġyt, 'lthyt w'lt' wn 'ldwly, 'ly' 'l'm, 'ld' hlyt, 'lšw'wn 'l'ġtm' 'yt w'l'ml, 'l'dl, w'lšw'wn 'lq'nwnyt) mmmṭlyn (lmktb r, 'sī 'lġmhwyt, mktb r, 'sī 'lwz'r', mktb 'ln', b' 'l'm, w'lġh'z 'l'mrkzy ll'mn 'l'sy'sy*

El Comité vino a reemplazar al subcomité permanente de la extinta Comisión Superior Nacional para los Derechos Humanos e incluye a representantes de los Ministerios de Relaciones Exteriores, Planificación, Cooperación Internacional, Información, Interior, Asuntos Sociales y Trabajo, Justicia y Asuntos Jurídicos y de la Oficina de la Presidencia de la República, del Gabinete del Primer Ministro, de la Oficina del Fiscal General y del **Órgano Central de Seguridad Política**.

The committee replaced the standing sub-committee of the (former) Higher National Committee for Human Rights and comprises representatives of the foreign, planning, international cooperation, information, interior, social affairs and labour, justice and legal affairs ministries and of the Office of the President of the Republic, the Prime Minister's Office, the Prosecutor-General's Office and **the Central Agency for Political Security**.

Ejemplo 4: Resultado de búsqueda con FARASA (informe A2, párrafo 13)

صدر قرار وزاري بتوقيع من وزير الشؤون الاجتماعية والعمل نائب رئيس المجلس الأعلى للأمومة والطفولة تضمن تحديد الجهات المشاركة في جمع المعلومات من الجهات الحكومية ومنظمات المجتمع المدني وقد تضمن ذلك الجهات التالية: (وزارة الصحة العامة والسكان، ووزارة الشؤون الاجتماعية والعمل، ووزارة حقوق الإنسان، ووزارة العدل، ووزارة التربية والتعليم، وزارة الأوقاف، ووزارة المالية، ووزارة التخطيط والتعاون الدولي

*šdr qr'r wz'ry btwqy' mn wzyr 'lšw'wn 'l'ġtm' 'yt w'l'ml n', b r, ys 'lmġls 'l' 'l'à ll'mwmī w'ltfwlī tđmn thdyd 'lġh' t 'lms' rktī fy ġm' 'lm' lwm' t mn 'lġh' t 'lhkwmyt wmnz m' t 'lmġtm' 'lmdny wqd tđmn dlk 'lġh' t 'lt' lyt: (wz' rī 'lšhī 'l' mī w'lsk'n, wwz' rī 'lšw'wn 'l'ġtm' 'yt w'l'ml, wwz' rī hqww 'ly'ns'n, wwz' rī 'l'dl, wwz' rī 'l'trbyī w'lt' lym, wz' rī 'l'wq'f, wwz' rī 'lm' lyt, wwz' rī 'lthyt w'lt' wn 'ldwly*

Los organismos públicos y las organizaciones de la sociedad civil encargados de reunir la información han sido designados por un decreto conjunto del Ministro de Asuntos Sociales y Trabajo y del Vicepresidente del Consejo Superior de Bienestar Materno-infantil; se trata esencialmente de las instituciones siguientes: Ministerios de Salud Pública y Población, de Asuntos Sociales y Trabajo, de Derechos Humanos, de Justicia, de Educación, de los awqaf, de Finanzas, de Planificación y Cooperación Internacional.

Designated pursuant to a decision of the Minister of Social Affairs and Labour, vice-chairperson of the Supreme Council for Motherhood and Childhood, the entities involved in gathering information from governmental bodies and civil society organizations included the following: the Ministry of Public Health and Population; the Ministry of Social Affairs and Labour; the Ministry of Human Rights; the Ministry of Justice; the Ministry of Education; the Ministry of Awqaf; the Ministry of Finance; the Ministry of Planning and International Cooperation.

Ejemplo 5: Resultado de búsqueda con FARASA (informe E2, párrafo 5)

ومجلس النواب، ومجلس الشورى، ومكتب النائب العام، وجمعية الإصلاح، واتحاد نساء اليمن، وهيئة التنسيق للمنظمات غير الحكومية، واللجنة الوطنية للمرأة، مؤسسة الصالح الاجتماعي للتنمية، ومركز المعلومات والتأهيل لحقوق الإنسان، والمرصد اليمني لحقوق الإنسان، ومجلس شوري الشباب، والأمانة الفنية للمجلس الأعلى للأمومة والطفولة.

*wmġls 'lnw'b, wmġls 'lšwrà, wmkth 'ln'b 'l'm, wġm 'yī 'ly'sl'h, w'th'd ns'? 'lymn, why, ī 'ltnsyq llmnm't ġyr 'lhkwmyī, w'llġnī 'lwīnyī llmr'ī, mw'ssī 'lš'lh 'l'ġm 'y llnmyī, wmrkz 'lm'lw'm't w't' hyl lhqwq 'ly'ns'n, w'lmršd 'lymny lhqwq 'ly'ns'n, wmġls šwrà 'lš'b, w'l'm'nī 'lfnīyī llmġls 'l'la ll'mwmī w'ltfwlī*

La Cámara de Representantes, el Consejo Consultivo, la Fiscalía General, la **Asociación de Beneficencia Al-Islah**, la Federación de Mujeres **Yemeníes**, la Comisión Coordinadora de las Organizaciones No Gubernamentales, la Comisión Nacional de la Mujer, la Asociación "Saleh" para el Desarrollo Social, el Centro de Información y Capacitación en materia de Derechos Humanos, el Observatorio Nacional de Derechos Humanos, el **Consejo Consultivo para la Juventud** y la Secretaría Técnica del Consejo Superior de Bienestar Materno-infantil.

the House of Representatives; the Shura Council; the Office of the Public Prosecutor; the **Reform Association**; the **Yemen Women's Union**; the Coordinating Authority for Child Rights Non-governmental Organizations (NGOs); the Women's National Committee; the Saleh Foundation for Social Development; the Human Rights Information and Training Centre; the Yemeni Observatory for Human Rights; the **Youth Advisory Council**; and the Technical Secretariat of the Supreme Council for Motherhood and Childhood.

Ejemplo 6: Resultado de búsqueda con FARASA (informe E2, párrafo 5)

En el ejemplo 2, FARASA muestra una sensibilidad del 8 % (1/13). Si bien sería de comprender que en ese fragmento no reconociera las EN incompletas («ministerios de X, Y, Z», en lugar de «Ministerio de X, Ministerio de Y y Ministerio de Z»), en el ejemplo 3, pese a que todas las EN estaban completas, la sensibilidad del programa fue del 0 %. En el ejemplo 4, la sensibilidad fue del 25 % (3/12) con una precisión del 67 % (2/3). Por tanto, FARASA no cumple en absoluto con nuestras expectativas y descartamos utilizar el programa.

Por cuanto precede, la localización de las unidades de análisis hubo de realizarse de manera manual. Sin embargo, como quedó demostrado en las fases 1 y 2 de la investigación, era posible semiautomatizar el proceso sustituyendo por minúsculas las mayúsculas iniciales precedidas de signo de puntuación<sup>51</sup> y seguidamente localizando y resaltando las mayúsculas restantes para localizarlas a simple vista<sup>52</sup>. Ante la inexistencia

<sup>51</sup> Empleando Microsoft Office 365, versión 16.22 de Microsoft Word, en «Buscar» insertamos ([.;\_?)]\_([A-Z]) (los guiones bajos equivalen a un espacio). A continuación, marcamos la casilla «Resaltar todos los elementos encontrados en» y seleccionamos la opción «Documento principal». En el desplegable, seleccionamos la opción «Todo» y a continuación marcamos la casilla «Usar caracteres comodín». Finalmente hicimos clic en «Buscar todos». De esta manera, se seleccionaron todas las mayúsculas precedidas de los signos de puntuación establecidos en la fórmula de búsqueda. A continuación, convertimos todas esas mayúsculas seleccionadas en minúsculas con el comando Mayúscula+Fn+F3.

<sup>52</sup> Para ello, en la opción de «Buscar y reemplazar avanzado» seleccionamos la casilla «Reemplazar». En la celda «Buscar» insertamos ([A-Z]) e hicimos clic en la casilla «Reemplazar» sin escribir nada. En el

de mayúsculas en árabe, solo podía operarse de ese modo en los documentos con grafía latina.

La localización resultó ser eficiente, pese a que no todos los resaltados eran resultados positivos. Sin embargo, los textos normativos en español solo portaban mayúscula inicial en su primera palabra por mandato del *Manual del Traductor* (véase [II.2.c.i](#)), por lo que no se resaltaban los precedidos por signo de puntuación. Así pues, aunque partir de 2014 las denominaciones de textos normativos comenzaron a portar mayúscula inicial en todas sus palabras, y pese a que se perdían muy pocas unidades de análisis en los textos anteriores a esa fecha (ya que el artículo determinado suele preceder a la denominación del texto normativo), convenía aplicar el procedimiento a los documentos en inglés.

El sistema presenta dos potenciales errores: 1) algunas mayúsculas resaltadas correspondientes a unidades de análisis podían habernos pasado inadvertidas; 2) los traductores al inglés podían no haber identificado algunos EN y, en consecuencia, podían haberlas traducido por secuencias de palabras sin mayúscula inicial. Respecto a 1) conviene señalar que hemos revisado seis veces la totalidad de los documentos, por lo que resulta difícil que hubiéramos incurrido en ese error. Respecto a 2), no hemos encontrado una sola ocasión en que eso hubiera ocurrido.

Como conclusión podemos afirmar que, pese a no haberse podido automatizar la localización de las unidades de análisis, y pese a los posibles errores, el procedimiento semiautomatizado de localización de las unidades de análisis fue eficaz.

## **b) Extracción y organización de los datos**

Gracias a que los informes de Estado se redactan en párrafos numerados (Naciones Unidas, 2009, p. 50), una vez localizadas las EN en inglés bastó con acudir al párrafo correspondiente en árabe para encontrar el original. Tras determinar los párrafos donde aparecía dicha unidad de análisis en árabe, extrajimos las «soluciones de traducción» en español e inglés y comprobamos y anotamos, una a una, las traducciones de todas las ocurrencias de una misma EN.

---

menú correspondiente seleccionamos «Usar caracteres comodín» y en el menú desplegable debajo de «Reemplazar» seleccionamos la opción «Resaltar». Finalmente hicimos clic en «Reemplazar todo».

### i. Soluciones de traducción

En esta investigación denominamos «solución de traducción» a cada versión diferente en una misma lengua de destino de una EN original. En el ejemplo 7 se muestran dos soluciones intratextuales de traducción para una única unidad de análisis.

Ministerio de Relaciones Exteriores	وزارة الخارجية
Ministerio de Asuntos Exteriores	wz 'rī 'lh 'rġyī [ <i>ministerio de exteriores</i> ]

Ejemplo 7: Dos soluciones intratextuales de traducción para una única unidad de análisis (informe A1, párrafos 93 y 126)

Para nuestros propósitos de investigación, el grado de corrección de las soluciones de traducción no es relevante y no lo valoramos: en todo caso afirmaremos, como en el ejemplo 8, que la solución de traducción en cuestión carece de motivación aparente.

Penal Code	قانون الإجراءات الجزائية q 'nwn 'ly 'ġr 'p 't 'lġz 'yī [ <i>ley de procedimiento penal</i> ]
------------	--

Ejemplo 8: Solución de traducción sin motivación aparente (informe A1, párrafo 96)

Por otro lado, no todas las soluciones de traducción registradas constituyen datos válidos para nuestros propósitos. A continuación presentamos una breve taxonomía de las soluciones de traducción registradas y determinamos qué tipos son analizables.

Las soluciones de traducción localizadas responden a técnicas o estrategias de traducción bien conocidas (Gil Bardají, 2008, 2009). Sin entrar en disquisiciones traductológicas que no atañen a una cuantificación computacional, nos conviene distinguir entre:

- Soluciones equivalentes: cuando en el original hay una EN y se traduce por su equivalente, y

- Soluciones implicitadoras o explicitadoras: cuando en el original hay una EN explícita y en la traducción se implícita, o viceversa.

Las soluciones equivalentes pueden ser completas o incompletas. Entendemos por «solución completa» la que refleja, como en el ejemplo 7, todos los elementos léxicos de la unidad de análisis, con independencia de que esta reflejara o no la denominación completa de la EN. El ejemplo 9 ilustra una solución incompleta.

Código	Code	قانون الإجراءات الجزائية  <i>q'nwn 'ly'gr'p't 'lǧz',yī</i>  [ley de procedimiento penal]
--------	------	--

Ejemplo 9: Solución de traducción incompleta (informe A1, párrafo 59)

Las soluciones implicitadoras o explicitadoras pueden igualmente ser completas o incompletas. Los ejemplos 10 y 11 muestran sendos casos de implicitación completa y parcial, el primero en la traducción al inglés, y el segundo en las traducciones al inglés y el español por transformación de la EN en un adjetivo (en negrita la EN y el elemento transformado).

la primera conferencia de diálogo sobre la justicia penal en la legislación yemení, organizada por el <b>Ministerio de derechos humanos</b> en cooperación con el Instituto danés de derechos humanos	First National Dialogue Conference on Criminal Justice and Yemeni Legislation	مؤتمر الحوار الوطني الأول الموسوم بالعدالة الجنائية في التشريع اليمني، الذي نظمته وزارة حقوق الإنسان بالتعاون مع المعهد الدانمركي لحقوق الإنسان  <i>mw'tmr 'lhw'r 'lwṭny 'l'wl 'lmwswm b'l'd'li 'lǧn',yī fy 'ltšry' 'lymny, 'lǧy nzmth wz'rī ḥqwq 'ly'ns'n b'lt'wn m' 'lm'hd 'ld'nmrky ḥqwq 'ly'ns'n</i>  [conferencia de diálogo nacional primera sobre la justicia penal en la legislación yemení, la cual la organizó el <b>ministerio de derechos humanos</b> con la cooperación del instituto danés de derechos humanos]
---	---	---

Ejemplo 10: Implicitación completa en inglés, sin motivación aparente, de una unidad de análisis (en negrita; informe B2, párrafo 73)

<p>independientemente de que sean candidatos a las elecciones presidenciales, <b>parlamentarias</b> o de los consejos locales</p>	<p><i>regardless of whether they are candidates in presidential, <b>parliamentary</b> or local council elections</i></p>	<p>سواء لمرشحي انتخابات الرئاسة أو البرلمان أو المجالس   <i>sw'ʔ lmršhy 'nth 'b't 'lr, 'sī 'w 'lbrlm 'n 'w 'lmğ 'ls 'lmħlyī</i>                   [para los candidatos de las elecciones para la presidencia, <b>el parlamento</b> o los consejos locales]</p>
---	--	--

Ejemplo 11: Implicación parcial, en español e inglés (informe C1, párrafo 116)

Algunas implicaciones responden a razones desconocidas. En esos casos, a veces se omite la unidad de análisis únicamente (ejemplo 10) y otras, la unidad de análisis desaparece por haberse suprimido partes mayores del documento que la incluían: una sección (ejemplo 12, pese a que en otros documentos se tradujeron apéndices similares), uno o varios párrafos (ejemplo 13), un epígrafe (como el d), párr. 119, del informe D1 en inglés y en español) o una oración o parte de ella.

<p>∅</p>	<p>قانون الجرائم والعقوبات رقم 13 لسنة 1994   <i>q'nwn 'lgr',m w'l'qwb't rqm 13 lsnī 1994</i>                   [ley de los delitos y las penas número 13 del año 1994]                   قانون تنظيم السجون رقم 48 لسنة 1991   <i>q'nwn tnyzm 'lsğwn rqm 48 lsnī 1991</i>                   [ley de la organización de las prisiones número 48 del año 1991]</p>
----------	---

Ejemplo 12: Implicación de dos unidades de análisis por supresión de una sección completa de un documento por razones desconocidas (informe A1, apéndice de fuentes principales)

<p>∅</p>	<p>قانون العمل رقم 5 لسنة 1995   <i>q'nwn 'l'ml rqm 5 lsnī 1995</i>                   [ley del trabajo número 5 del año 1995]</p>
----------	---

Ejemplo 13: Implicación de una unidad de análisis en la versión inglesa por supresión de dos párrafos por razones desconocidas (informe E2, párrafos 55 y 56)

En ocasiones (ejemplo 12), la implicación supone la desaparición de la única ocurrencia de la unidad de análisis en el documento, lo que imposibilitó su análisis.

De todos los fenómenos observados, aquellos que afectan directamente a los análisis son los equivalentes completos, los equivalentes incompletos y las implicaciones (fueran omisiones -implicaciones completas- o, por ejemplo, reformulaciones -implicaciones parciales). No se enumeran en este listado las explicitaciones ya que resulta imposible saber cuáles son las distintas (o mismas) traducciones de un EN que no figuran de manera explícita en el original (una o repetidas veces). Así pues, las explicitaciones quedaron excluidas del análisis.

Por otra parte, los equivalentes incompletos y las implicaciones (tanto completas como parciales) responden en general a motivaciones estilísticas o constituyen ajustes de gramática comparada que no reflejan a nuestro juicio falta de coherencia. Además, detectamos muy pocas unidades de análisis en el corpus paralelo que presentaran este tipo de soluciones de traducción en ambas lenguas, por lo que decidimos analizar en esta tesis exclusivamente las EN cuyas soluciones de traducción son equivalentes completos.

Las EN correspondientes a las soluciones incompletas y las implicaciones con más de una ocurrencia intratextual pudieron analizarse gracias a las demás ocurrencias en los textos. No ocurrió lo mismo, sin embargo, con las implicaciones de unidades con solo una ocurrencia intratextual. Al no existir traducción, el análisis en los términos que aquí se plantea no es posible y por tanto tuvimos que eliminar ese elemento del análisis.

En definitiva, del número total de EN identificadas, solamente hubo que eliminar aquellas que solo presentaban una ocurrencia en el texto y que, además, se habían implicado en la traducción.

## **ii. Organización de los datos**

Los datos se volcaron en archivos de formato Excel. Cada fila correspondía a una EN, acompañada de la siguiente información: signatura del informe en cuestión, año de publicación del informe, párrafos en los que figura la EN original, categoría, EN original

y soluciones de traducción al inglés y al español. En los casos en los que se extrajeron varias soluciones de traducción para una misma EN, se añadieron las columnas correspondientes, como ilustra el siguiente ejemplo.

Informe	Año	Párrafos	Categoría de EN	Original	Inglés – S1	Inglés - S2
D1	2002	49x2, Ref.31	Entidad corporativa nacional	اللجنة الوطنية للمرأة	National Committee for Women	National Women's Committee

Ejemplo 14: Organización de los datos

### c) Análisis de los datos

A continuación se presentan las medidas aplicadas para cuantificar la homogeneidad en relación con la hipótesis principal, por un lado, y por otro, y de manera independiente, con las hipótesis secundarias.

#### i. Homogeneidad y disimilitud: el algoritmo de Jaro

En esta investigación se cuantifica la homogeneidad mediante dos medidas:

- Cuantificación del porcentaje de unidades de análisis heterogéneas en la traducción al español y al inglés, y
- Cuantificación del grado de disimilitud de las soluciones de traducción heterogéneas al español y al inglés.

Para cuantificar las diferencias entre las soluciones de traducción realizamos un análisis de tipo computacional. En consecuencia, abordamos las soluciones de traducción en cuanto que cadenas de caracteres y medimos sus diferencias, consecuentemente, mediante un sistema de medición de diferencias entre cadenas de caracteres: el algoritmo de la distancia de Jaro.

Los sistemas de medición de diferencias entre cadenas de caracteres se emplean en el desarrollo de sistemas de extracción de información para, por ejemplo, la resolución de correferencia (es decir, para determinar si dos cadenas son referentes de una misma



identidad). En esta investigación empleamos la distancia de Jaro para cuantificar el grado de homogeneidad o disimilitud entre las distintas soluciones de traducción aplicadas a una misma EN. No nos consta que se hayan utilizado esta metodología con anterioridad para medir la homogeneidad traductora, por lo que debemos subrayar la originalidad de la propuesta.

El *string matching* (emparejamiento de cadenas) se define como «the process of determining whether two name strings are instances of the same name» (Patman & Thompson, 2003, p. 28). Por ejemplo, «MSFT» y «Microsoft Corporation» son dos referentes de la misma entidad, al igual que un nombre y su hipocorístico («José» y «Pepe»).

El emparejamiento correcto es crítico para la EI, en particular de los nombres propios (Christen, 2006, pp. 290 y 292). Según Amón y Jiménez (2010, p. 8), las variaciones o errores en las cadenas de caracteres más adversas para la EI son los errores ortográficos y tipográficos, las abreviaturas, los *tokens* omitidos, los prefijos y sufijos sin valor semántico, los *tokens* en desorden y los espacios en blanco.

Para solventar el problema se emplean medidas de similitud y de distancia. De superarse un umbral predeterminado en la medida, las cadenas se consideran representaciones de la misma entidad (Amón & Jiménez, 2010, p. 14). Existen diversos métodos de medición:

- Técnicas fonéticas: «establecen comparaciones de palabras por similitud en la percepción de sonidos, cuando hay más de una forma escrita para representar el mismo nombre. Estas técnicas asignan códigos a cada secuencia de caracteres basados en el sonido que estos producen» (Delgado Loyola, 2015, p. 98). Se desarrollaron principalmente para la lengua inglesa (Christen, 2006, pp. 293-294);
- *Pattern matching techniques*: emparejan cadenas mediante algoritmos de similitud o distancia;
- Técnicas mixtas: combinan las dos anteriores para mejorar el resultado.

Los algoritmos más utilizados para medir la similitud y la distancia son los de Levenshtein, Damerau-Levenshtein, Jaro y Winkler. El algoritmo de Levenshtein calcula el mínimo número de cambios requeridos (inserción, eliminación, sustitución o transposición) para igualar una cadena de caracteres a otra (Levenshtein, 1966). El

algoritmo de Damerau-Levenshtein, una variante del anterior, considera la trasposición un solo cambio (Christen, 2006) (Levenshtein lo considera dos: una inserción y una eliminación). Por su parte, Jaro y Winkler determinan la similitud o coincidencia, el primero en base a las inserciones, eliminaciones y transposiciones requeridas para igualar una cadena a otra considerando el número y orden de los caracteres comunes y la longitud de la cadena.

The Winkler [...] algorithm improves upon the Jaro algorithm by applying ideas based on empirical studies [...] which found that fewer errors typically occur at the beginning of names. The Winkler algorithm therefore increases the Jaro similarity measure for agreeing initial characters (up to four). (Christen, 2006, p. 296)

Otros algoritmos discriminan por tipo de cambio. Fernández Orquín et al. (2009), por ejemplo, penalizan menos los errores ortográficos.

Puesto que la variación en cualquier posición de la cadena de caracteres afecta de igual modo a la coherencia traductora, no es relevante aplicar en esta investigación algoritmos cuyos resultados varíen en función de la posición en la que se hallen diferencias o la función del tipo de diferencia.

A continuación describiremos con mayor detalle el algoritmo de Jaro.

Dadas dos cadenas  $A$  y  $B$ , siendo  $|A|$  y  $|B|$  el número de caracteres dentro de cada cadena, y  $m$  el número de caracteres coincidentes en ambas cadenas, la distancia de Jaro entre ambas queda definida matemáticamente como:

$$j_d = \frac{1}{3} \cdot \left( \frac{m}{|A|} + \frac{m}{|B|} + \frac{m-t}{m} \right) \quad (1)$$

El valor resultante varía de 0 a 1: 1 indica identidad y 0 máxima disimilitud. Para que caracteres comunes a ambas cadenas se consideren coincidentes, su diferencia de posición  $x$  dentro de cada cadena debe cumplir la siguiente condición:

$$x \leq \left\lceil \frac{\max(|A|, |B|)}{2} - 1 \right\rceil \quad (2)$$

El parámetro  $t$  de la ecuación (1), conocido como número de transposiciones, se define como la mitad del número de caracteres coincidentes desordenados, esto es, los comunes a ambas cadenas y ubicados en distinta posición en ellas.

Supongamos que deseamos analizar la similitud entre las cadenas LEMAS y LEAMOS. En primer lugar, el algoritmo calcula la diferencia de posición máxima o rango ( $x$ ) para la cual podemos considerar dos caracteres coincidentes. En este caso, la longitud de ambas cadenas es  $|A| = 5$  y  $|B| = 6$ . Mediante la fórmula (2) se obtiene un valor de distancia máxima  $x = 2$ . Por lo tanto, caracteres idénticos no coincidentes en posición dentro de las dos cadenas serán considerados coincidentes siempre que la diferencia de posición entre ambos sea menor que 2.

En la ilustración 3 se representa como ejemplo el cálculo de la distancia de Jaro entre las cadenas LEMAS y LEAMOS. Cada iteración del algoritmo analiza un carácter de la cadena de referencia LEMAS (marcado en rojo intenso en la figura) y busca posibles caracteres coincidentes en la cadena LEAMOS. Para considerar dos caracteres coincidentes, su diferencia de posición dentro de las dos cadenas debe encontrarse dentro de un rango determinado por  $x$  (elementos en rojo tenue en la imagen) y, además, no haber sido emparejados con anterioridad (elementos en azul). Los resultados serían equivalentes si usásemos la cadena «LEAMOS» como referencia.

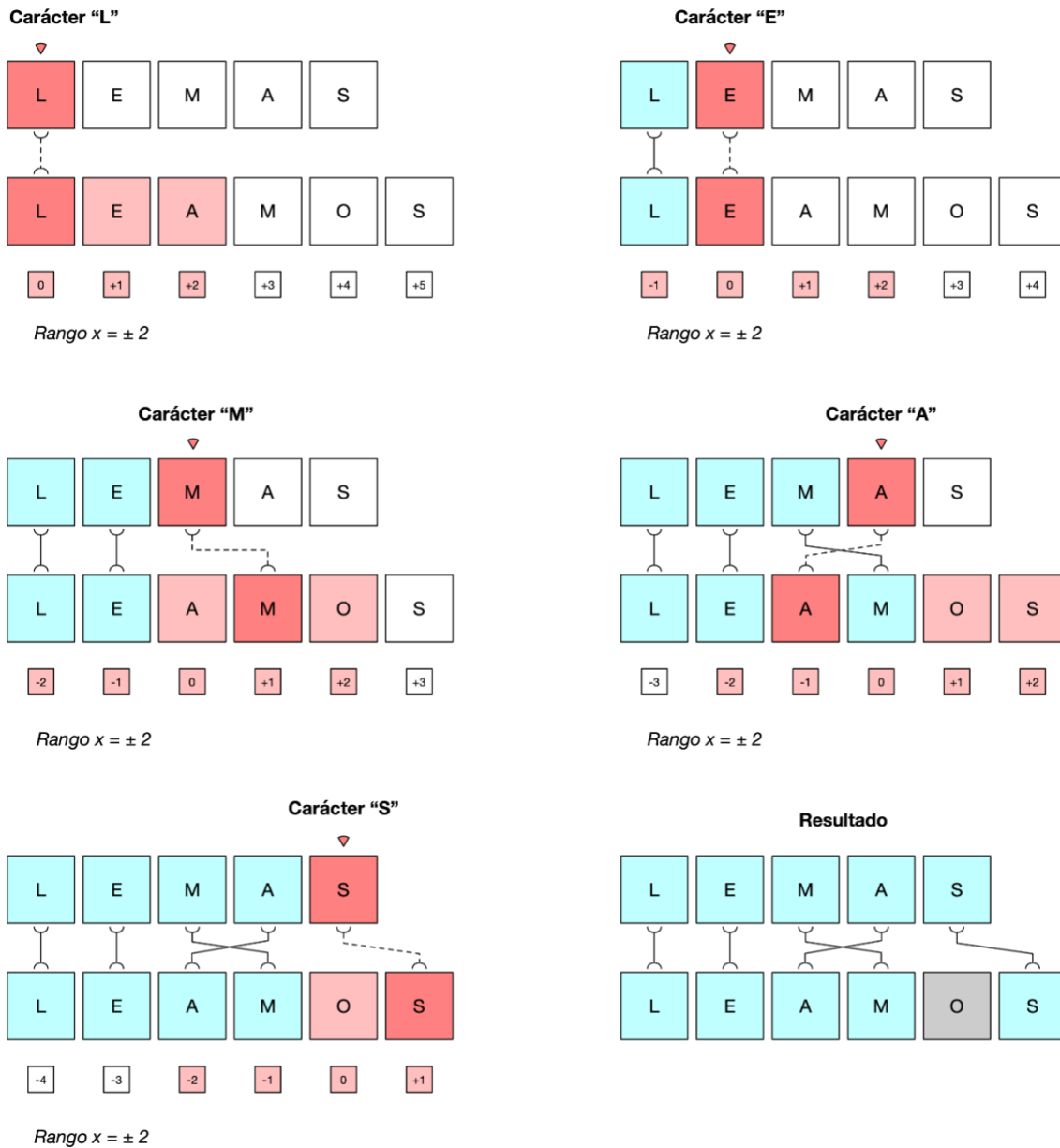


Ilustración 3: Representación gráfica del cómputo de caracteres coincidentes para el cálculo de la distancia de Jaro entre las cadenas LEMAS y LEAMOS.

Por lo tanto, el número de caracteres coincidentes que cumplen la condición indicada en la formula (2) es  $m = 5$  (L, E, A, M, S). El número de transposiciones  $t$  se calcula como la mitad del número de caracteres coincidentes desordenados, que en este caso son dos, A y M, por lo tanto  $t = 1$ . Aplicando la formula (1) a estas dos cadenas obtendríamos la distancia de Jaro:

$$j_d = \frac{1}{3} \cdot \left( \frac{5}{5} + \frac{5}{6} + \frac{5-1}{5} \right) \approx 0.88 \quad (3)$$

Puesto que esta investigación se centra en el estudio de la variabilidad y no de la similitud, definimos el índice de disimilitud ( $d_i$ ) como sigue:

$$d_i = 1 - j_d \quad (4)$$

Esta medida de disimilitud, que es el índice complementario a la distancia de Jaro, es la que seguiremos en nuestro análisis. De esta forma, finalmente, el valor de disimilitud entre las cadenas «LEMAS» y «LEAMOS» sería:

$$d_i = 1 - 0.88 = 0.12 \quad (5)$$

Para la automatización del análisis de los datos se programó una implementación propia gracias a una colaboración de ese algoritmo en MATLAB (Mathworks) en su versión R2017a (9.2).

Por otro lado, se programó un *script* para la automatización de los análisis. El programa recibe como entrada los datos en archivos en formato Excel creados para el análisis de cada plano. En estos archivos, cada fila corresponde a una EN original y las columnas contienen las soluciones de traducción adoptadas. El programa calcula de forma automática las combinaciones únicas de las soluciones de traducción para una misma EN y aplica el algoritmo de cálculo de disimilitud para cada una de estas combinaciones. Finalmente, el *script* genera un archivo Excel de salida para cada archivo de entrada con los resultados de las operaciones individuales y otros datos estadísticos, como el promedio de disimilitud para las traducciones de cada EN, un promedio global para cada archivo y su desviación estándar.

La eficacia de este método se contrastó con los resultados obtenidos en las fases previas de esta investigación.

A continuación describimos los análisis realizados.

## **ii. Análisis intratextual (hipótesis principal)**

Para este análisis se extrajeron todas las unidades de análisis presentes en los informes de nuestra muestra y se localizaron sus soluciones de traducción. La información extraída se organizó en archivos Excel con datos desglosados por informe y lengua. En cada archivo se crearon sendas columnas para la unidad y su categoría, así como otras tantas para las soluciones de traducción.

A continuación, se calculó el grado de heterogeneidad para cada informe, definido como el número de unidades de análisis heterogéneas respecto al total analizado.

Posteriormente, se aplicó el algoritmo de cálculo de disimilitud a cada unidad de análisis heterogénea y se calculó la media de las disimilitudes de todas las combinaciones posibles no repetidas de sus distintas soluciones de traducción. Por último, se determinó la disimilitud media para cada informe, así como su desviación estándar.

## **iii. Análisis intertextual 1 (hipótesis secundaria 1)**

Recordemos que la muestra se compone de dos informes por comité, que fueron seleccionados con el criterio de mayor distancia temporal (véase [III.1.c.iii](#)).

En primer lugar, determinamos las unidades de análisis comunes a cada par de esos informes, ya que solo ellas nos permitirían testar la hipótesis secundaria 1. Esas unidades se organizaron en archivos Excel con datos desglosados por informe, comité y lengua. En cada archivo se crearon sendas columnas para la unidad y su categoría, y otras tantas para cada solución de traducción heterogénea. A continuación, se determinaron los porcentajes de heterogeneidad en dos planos:

- Plano intracomité: en él, se considera heterogénea la unidad de análisis que presente variación en cualquiera de los informes presentados al mismo comité, y
- Plano intradocumental: en él, se considera heterogénea la unidad de análisis que presente variación en un solo informe.

Con el algoritmo correspondiente se calculó el grado de disimilitud de cada unidad de análisis heterogénea, el promedio de disimilitud de todas las combinaciones posibles no repetidas de esas soluciones de traducción en cada plano, la media de las disimilitudes anteriores y su desviación estándar. Tras ello se compararon los datos y se extrajeron conclusiones considerando que, a falta de otras posibles explicaciones, un grado de heterogeneidad mayor en el plano intracomité que en el intradocumental podría implicar un efecto negativo de la variable diacrónica sobre la homogeneidad.

#### **i. Análisis intertextual 2 (hipótesis secundaria 2)**

Para este análisis se determinaron las unidades comunes a al menos dos comités, ya que solo ellas nos permitirían testar la hipótesis secundaria 2. La información se organizó en un archivo Excel en el que se incluyeron datos desglosados por informe, comité y lengua, así como las soluciones de traducción de cada unidad de análisis y su categoría.

A continuación, se determinaron los porcentajes de heterogeneidad en dos planos:

- Plano intercomité: en él, se considera heterogénea la unidad de análisis que presente variación en cualquiera de los informes presentados a alguno de los comités, y
- Plano intracomité: en él, se considera heterogénea la unidad de análisis que presente variación en cualquiera de los informes presentados al mismo comité.

Con el algoritmo se calculó el grado de disimilitud de cada unidad de análisis heterogénea, el promedio de disimilitud de todas las combinaciones posibles no repetidas de esas soluciones de traducción de cada plano, la media de las disimilitudes anteriores y su desviación estándar. Tras ello, se compararon los datos y se extrajeron conclusiones considerando que, a falta de otros posibles motivos, un grado de heterogeneidad mayor en el plano intercomité que en el intracomité podría implicar un efecto negativo de la variable temática sobre la homogeneidad.





## Capítulo IV: Datos y análisis de los resultados

El capítulo 4 consta de dos partes. En la primera figuran los datos extraídos mediante las mediciones descritas. En [IV.1.a](#), [IV.1.b](#) y [IV.1.c](#) se presentan, respectivamente, los datos extraídos mediante los análisis intratextual e intertextuales. En la segunda parte de este capítulo se analizan esos datos.

### 1. Datos

#### a) Análisis intratextual

En la tabla 6 se presenta el número de unidades de análisis, con datos desglosados por informe y categoría de EN.

Informes	Textos normativos	Entidades corporativas	Total
A1	29	9	38
A2	54	74	128
B1	57	15	72
B2	66	65	131
C1	53	26	79
C2	55	51	106
D1	73	37	110
D2	16	32	48
E1	61	80	141
E2	43	143	186
<b>TOTAL</b>	<b>507</b>	<b>532</b>	<b>1039</b>

Tabla 6: Número de unidades de análisis, con datos desglosados por informe y categoría de EN

Los datos anteriores comprenden las unidades de análisis extraídas de cada informe, con independencia de que se repitan en otros informes. Recordemos que, en esta investigación, cada formulación de una EN constituye una unidad de análisis diferenciada. Por tanto, el número de unidades de análisis no se corresponde con el número de EN.

Las tablas 7 y 8 presentan los resultados de los análisis descritos en [III.2.c.ii](#). En la tabla 7 figuran las unidades de análisis traducidas de manera no homogénea, con datos desglosados por informe, categoría de EN y lengua, y expresados en términos totales y porcentuales. Adviértase que «n» es el número de unidades de análisis heterogéneas, en total o en particular para su categoría, inclusive las unidades de análisis repetidas en otros informes. El porcentaje que figura debajo de «n» corresponde al porcentaje de unidades de análisis heterogéneas respecto al total, absoluto o de su categoría, de unidades de análisis en el documento en cuestión.

Informe	Español			Inglés		
	Total	Textos normativos	Entidades corporativas	Total	Textos normativos	Entidades corporativas
A1	n=6	n=3	n=3	n=3	n=2	n=1
	15,8 %	10,3 %	33,3 %	7,9 %	6,9 %	11,1 %
A2	n=19	n=11	n=8	n=8	n=2	n=6
	14,8 %	20,4 %	10,8 %	6,3 %	3,7 %	8,1 %
B1	n=4	n=3	n=1	n=4	n=2	n=2
	5,6 %	5,3 %	6,7 %	5,6 %	3,5 %	13,3 %
B2	n=36	n=18	n=18	n=12	n=8	n=4
	27,5 %	27,3 %	27,7 %	9,2 %	12,1 %	6,2 %
C1	n=8	n=7	n=1	n=9	n=9	n=0
	10,1 %	13,2 %	3,8 %	11,4 %	17 %	0 %
C2	n=14	n=6	n=8	n=14	n=5	n=9
	13,2 %	10,9 %	15,7 %	13,2 %	9,1 %	17,6 %
D1	n=8	n=5	n=3	n=5	n=3	n=2
	7,3 %	6,8 %	8,1 %	4,5 %	4,1 %	5,4 %
D2	n=5	n=0	n=5	n=2	n=0	n=2
	10,4 %	0 %	15,6 %	4,2 %	0 %	6,3 %
E1	n=11	n=5	n=6	n=11	n=4	n=7
	7,8 %	8,2 %	7,5 %	7,8 %	6,6 %	8,8 %
E2	n=21	n=3	n=18	n=8	n=1	n=7
	11,3 %	7 %	12,6 %	4,3 %	2,3 %	4,9 %
TOTAL	<b>n=132</b>	<b>n=61</b>	<b>n=71</b>	<b>n=76</b>	<b>n=36</b>	<b>n=40</b>
	<b>12,7 %</b>	<b>12,0 %</b>	<b>13,3 %</b>	<b>7,3 %</b>	<b>7,1 %</b>	<b>7,5 %</b>

Tabla 7: Unidades de análisis traducidas de manera no homogénea en el análisis intratextual, con datos desglosados por informe, categoría de EN y lengua, y expresados en términos totales y porcentuales

El número de unidades de análisis de una y otra categoría es muy similar en este plano: 507 y 532. De ellas, se tradujeron de manera no homogénea 61 textos normativos y 71 entidades corporativas en español, y 36 y 40, respectivamente, en inglés. El número de unidades de análisis heterogéneas en español prácticamente dobla al del inglés (tabla 7).

En la mayoría de los casos, el número de soluciones de traducción para cada unidad de análisis se mantuvo en el mínimo, es decir, 2; solo en algún caso se registraron 3

soluciones para una unidad y, como excepción, «قانون الجرائم والعقوبات» [ley de los delitos y las penas] presenta en el informe B2 un total de 7 soluciones de traducción en español<sup>53</sup>.

El porcentaje de heterogeneidad es igual o superior en español, al punto de que, en B2, triplica la del inglés (27,5 % y 9,2 %, respectivamente).

Respecto a los textos normativos, solo D2 presenta una homogeneidad del 100 % en ambas lenguas. En la categoría de entidades corporativas nacionales, solo C1 en inglés presenta una homogeneidad del 100 %.

En la tabla 8 figuran los resultados del análisis de disimilitud, con datos desglosados por lengua e informe. La desviación estándar correspondiente figura entre paréntesis.

Informe	Disimilitud – español			Disimilitud – inglés		
	Total	Textos normativos	Entidades corporativas nacionales	Total	Textos normativos	Entidades corporativas nacionales
A1	0,104 (SD= 0,073)	0,092 (SD= 0,071)	0,115 (SD=0,089)	0,164 (SD = 0,156)	0,241 (SD= 0,113)	0,01 (SD= 0)
A2	0,247 (SD= 0,114)	0,245 (SD= 0,081)	0,251 (SD= 0,156)	0,171 (SD = 0,099)	0,181 (SD= 0,021)	0,168 (SD= 0,117)
B1	0,233 (SD= 0,148)	0,177 (SD= 0,118)	0,402 (SD= 0)	0,215 (SD = 0,183)	0,179 (SD= 0,226)	0,252 (SD= 0,21)
B2	0,199 (SD= 0,13)	0,208 (SD= 0,117)	0,191 (SD= 0,145)	0,179 (SD = 0,081)	0,178 (SD= 0,062)	0,181 (SD= 0,12)
C1	0,152 (SD= 0,085)	0,149 (SD= 0,091)	0,169 (SD= 0)	0,113 (SD = 0,083)	0,113 (SD= 0,083)	-
C2	0,217 (SD = 0,154)	0,157 (SD= 0,105)	0,262 (SD= 0,175)	0,307 (SD = 0,154)	0,224 (SD= 0,104)	0,354 (SD= 0,162)
D1	0,213 (SD = 0,153)	0,163 (SD= 0,149)	0,297 (SD= 0,144)	0,217 (SD = 0,123)	0,173 (SD= 0,142)	0,283 (SD= 0,073)
D2	0,148 (SD = 0,053)	-	0,148 (SD= 0,053)	0,104 (SD = 0,069)	-	0,104 (SD = 0,069)
E1	0,169 (SD = 0,105)	0,135 (SD= 0,084)	0,197 (SD= 0,12)	0,172 (SD = 0,102)	0,227 (SD= 0,142)	0,14 (SD= 0,063)
E2	0,158 (SD = 0,116)	0,184 (SD= 0,171)	0,154 (SD= 0,111)	0,168 (SD = 0,048)	0,159 (SD= 0)	0,169 (SD= 0,051)
<b>Medias totales</b>	<b>0,18 (SD= 0,11)</b>	<b>0,17 (SD= 0,11)</b>	<b>0,22 (SD= 0,10)</b>	<b>0,18 (SD= 0,11)</b>	<b>0,19 (SD= 0,11)</b>	<b>0,18 (SD= 0,10)</b>

Tabla 8: Resultados del análisis de disimilitud intratextual, con datos desglosados por lengua e informe. Entre paréntesis, desviación estándar (SD)

La disimilitud varía del 0,104 (A1) al 0,247 de (B2) en español. En inglés, oscila entre el 0,104 (D2) y el 0,307 (C2). No se observa un mayor promedio de disimilitud en una

<sup>53</sup> Código de los Delitos y las Penas; Código Penal; Código penal; Código de delitos y penas; Código de Delitos y Penas; Código de delitos y sanciones; Código de Delitos y Sanciones.

lengua que en otra. Las medias también se asemejan mucho en las dos lenguas y en las dos categorías analizadas, aunque los valores relativos a las entidades corporativas nacionales en español son ligeramente superiores al resto.

### b) Análisis intertextual 1

En la tabla 9 figuran las unidades de análisis localizadas, con datos desglosados por categoría de EN.

Comité	Textos normativos	Entidades corporativas nacionales	Total
Comité A	8	4	12
Comité B	16	5	21
Comité C	12	9	21
Comité D	3	7	10
Comité E	13*	20	33*
<b>TOTAL</b>	<b>52*</b>	<b>45</b>	<b>97*</b>

Tabla 9: Unidades de análisis localizadas en el análisis intertextual 1, con datos desglosados por categoría de EN

\* Estas cifras deben incrementarse en +1 en español por la presencia de una unidad de análisis en los dos informes originales del comité E cuya única repetición en E2 fue omitida en la versión inglesa, pero no en la española. Como resultado, el análisis en español cuenta con un elemento más que el inglés. Tendremos en cuenta esta puntualización a la hora de interpretar las tablas 10, 11 y 12.

En la tabla 10 se presentan los datos relativos a las unidades de análisis heterogéneas en el plano intracomité, con datos desglosados por categoría y lengua, expresados en términos totales y porcentuales.

Comités	Falta de homogeneidad					
	Español			Inglés		
	Total	Textos normativos	Entidades corporativas nacionales	Total	Textos normativos	Entidades corporativas nacionales
Comité A	n=9	n=7	n=2	n=8	n=7	n=1
	75 %	87,5 %	50 %	66,7 %	87,5 %	25 %
Comité B	n=17	n=14	n=3	n=15	n=12	n=3
	81 %	87,5 %	60 %	71,4 %	75 %	60 %
Comité C	n=13	n=11	n=2	n=11	n=7	n=4
	61,9 %	91,7 %	22,2 %	52,4 %	58,3 %	44,4 %
Comité D	n=9	n=3	n=6	n=7	n=2	n=5
	90 %	100 %	85,7 %	70 %	66,7 %	71,4 %
Comité E	n=22	n=10	n=12	n=22	n=10	n=12
	64,7 %	71,4 %	60 %	66,7 %	76,9 %	60 %
Total general	<b>n=70</b>	<b>n=45</b>	<b>n=25</b>	<b>n=63</b>	<b>n=38</b>	<b>n=25</b>
	<b>71,4 %</b>	<b>84,9 %</b>	<b>55,6 %</b>	<b>64,9 %</b>	<b>73,1 %</b>	<b>55,6 %</b>

Tabla 10: Unidades de análisis traducidas de manera no homogénea en el plano intracomité, con datos desglosados por categoría, lengua y comité, expresados en términos totales y porcentuales

El número de unidades de análisis es muy similar en ambas categorías: 52 y 45 para textos normativos y entidades corporativas nacionales.

El número de soluciones de traducción varía del mínimo (2) hasta 8 para el español («قانون الجرائم والعقوبات» [ley de los delitos y las penas] en los informes B1 y B2) y 5 para el inglés («قانون الإجراءات الجزائية» [ley de los procedimientos penales] en los informes A1 y A2).

En general, la falta de homogeneidad medida en términos porcentuales es alta en este plano. La heterogeneidad siempre es mayor en español que en inglés (más del 70 % como promedio), a excepción del comité E. El porcentaje de unidades de análisis no homogéneas varía considerablemente de unos comités a otros: D es el menos homogéneo en español (90 %) y B, en inglés (71,4 %).

Los textos normativos fueron traducidos de manera menos homogénea que las entidades corporativas nacionales (84,9 % y 73,1 %, respectivamente, para español e inglés, frente

a 55,6 % para español e inglés). En D se registra una heterogeneidad del 100 % en los textos normativos al español, aunque la cifra no es representativa debido al escaso número de unidades de análisis (n=3). En el extremo opuesto, C es el comité que registra mayor homogeneidad, tanto en inglés (52,4 %) como en español (61,9 %).

En la tabla 11 se presentan los datos relativos a las unidades de análisis traducidas de manera no homogénea en el plano intradocumental, con datos desglosados por categoría y lengua, expresados en términos totales y porcentuales.

Informe	Español			Inglés		
	Total	Textos normativos	Entidades corporativas	Total	Textos normativos	Entidades corporativas
A1	n=4	n=3	n=1	n=2	n=2	n=0
	33,3 %	37,5 %	25 %	16,7 %	25 %	0 %
A2	n=4	n=4	n=0	n=1	n=1	n=0
	33,3 %	50 %	0 %	8,3 %	12,5 %	0 %
B1	n=2	n=2	n=0	n=3	n=2	n=1
	8,5 %	12,5 %	0 %	14,3 %	12,5 %	20 %
B2	n=9	n=6	n=3	n=2	n=2	n=0
	42,9 %	37,5 %	60 %	9,5 %	12,5 %	0 %
C1	n=3	n=2	n=1	n=3	n=3	n=0
	14,3 %	16,7 %	11,1 %	14,3 %	25 %	0 %
C2	n=5	n=3	n=2	n=6	n=2	n=4
	23,8 %	25 %	22,2 %	28,6 %	16,7 %	44,4 %
D1	n=4	n=2	n=2	n=3	n=1	n=2
	40 %	66,7 %	28,6 %	30 %	33,3 %	28,6 %
D2	n=1	n=0	n=1	n=1	n=0	n=1
	10 %	0 %	14,3 %	10 %	0 %	14,3 %
E1	n=6	n=3	n=3	n=7	n=3	n=4
	17,6 %	21,4 %	15 %	21,2 %	23,1 %	20 %
E2	n=9	n=1	n=8	n=6	n=4	n=2
	26,5 %	7,1 %	40 %	18,2 %	30,8 %	10 %
TOTAL	n=47	n=26	n=21	n=34	n=20	n=14
	48 %	49,1 %*	46,7 %	35 %	38,5 %	26,9 %

Tabla 11: Unidades de análisis traducidas de manera no homogénea en el plano intradocumental, con datos desglosados por categoría, lengua y por informe, expresados en términos totales y porcentuales

La tabla 12 compara los resultados obtenidos tras el análisis de los dos planos. Estos datos confirman la diferente afectación en el grado de homogeneidad en cada plano. Todos los porcentajes de heterogeneidad son mayores en el plano intracomité que en el intradocumental en cada caso (entre un 8,9% y un 35,8% superiores). En todos los casos el porcentaje de heterogeneidad en español es superior al mismo en inglés.

	Español			Inglés		
	Total	Textos normativos	Entidades corporativas	Total	Textos normativos	Entidades corporativas
Plano intracomité	n=70	n=45	n=25	n=63	n=38	n=25
	<b>71,4 %</b>	84,9 %	55,6 %	<b>64,9 %</b>	73,1 %	55,6 %
Plano intradocumental	n=47	n=26	n=21	n=34	n=20	n=14
	<b>48 %</b>	49,1 %	46,7 %	<b>35 %</b>	38,5 %	26,9 %

Tabla 12: Heterogeneidad comparada de los planos intracomité e intradocumental, con datos desglosados por lengua y categoría, y expresados en términos porcentuales y totales

En la tabla 13 figuran los resultados del análisis de la disimilitud intracomité, con datos desglosados por lengua y comité.

Comité	Disimilitud – español			Disimilitud – inglés		
	Total	Textos normativos	Entidades corporativas nacionales	Total	Textos normativos	Entidades corporativas nacionales
<b>Comité A</b>	0,215 (SD = 0,141)	0,229 (SD= 0,13)	0,169 (SD= 0,226)	0,308 (SD = 0,126)	0,308 (SD= 0,136)	0,31 (SD= 0)
<b>Comité B</b>	0,221 (SD = 0,091)	0,216 (SD= 0,083)	0,243 (SD= 0,143)	0,2 (SD = 0,13)	0,207 (SD= 0,141)	0,17 (SD= 0,088)
<b>Comité C</b>	0,216 (SD = 0,148)	0,213 (SD= 0,158)	0,229 (SD= 0,112)	0,265 (SD = 0,08)	0,241 (SD= 0,08)	0,306 (SD= 0,07)
<b>Comité D</b>	0,22 (SD = 0,138)	0,295 (SD= 0,206)	0,183 (SD= 0,093)	0,175 (SD = 0,096)	0,203 (SD= 0,025)	0,163 (SD= 0,144)
<b>Comité E</b>	0,193 (SD = 0,105)	0,233 (SD= 0,093)	0,159 (SD= 0,105)	0,155 (SD = 0,128)	0,135 (SD= 0,127)	0,173 (SD= 0,132)
<b>Medias totales</b>	<b>0,213 (SD= 0,125)</b>	<b>0,223 (SD= 0,134)</b>	<b>0,197 (SD= 0,136)</b>	<b>0,221 (SD= 0,112)</b>	<b>0,219 (SD= 0,102)</b>	<b>0,224 (SD= 0,087)</b>

Tabla 13: Resultados del análisis de la disimilitud intracomité, con datos desglosados por lengua y comité. Entre paréntesis, desviación estándar (SD)

El grado de disimilitud en el plano intracomité oscila de 0,193 (E) a 0,221 (B) en español, con una media de 0,213. En inglés, de nuevo, la variación es mayor: de 0,155 (E) a 0,308



(A), con una media de 0,221. No se observan grandes diferencias entre lenguas o entre categorías.

A continuación se presentan los resultados de disimilitud en el plano intradocumental.

Informe	Disimilitud – español			Disimilitud – inglés		
	Total	Textos normativos	Entidades corporativas nacionales	Total	Textos normativos	Entidades corporativas nacionales
<b>A1</b>	0,073 (SD= 0,07)	0,092 (SD= 0,071)	0,014 (SD=0)	0,241 (SD= 0,113)	0,241 (SD= 0,113)	-
<b>A2</b>	0,213 (SD= 0,09)	0,213 (SD= 0,09)	-	0,166 (SD=0)	0,166 (SD=0)	-
<b>B1</b>	0,216 (SD= 0,138)	0,216 (SD= 0,138)	-	0,153 (SD= 0,165)	0,179 (SD= 0,226)	0,103 (SD= 0)
<b>B2</b>	0,228 (SD= 0,103)	0,205 (SD= 0,089)	0,274 (SD= 0,132)	0,14 (SD= 0,005)	0,14 (SD= 0,005)	-
<b>C1</b>	0,211 (SD= 0,036)	0,231 (SD= 0,005)	0,169 (SD= 0)	0,144 (SD= 0,115)	0,144 (SD= 0,115)	-
<b>C2</b>	0,286 (SD= 0,174)	0,175 (SD= 0,111)	0,452 (SD= 0,071)	0,337 (SD= 0,144)	0,28 (SD= 0,143)	0,365 (SD= 0,157)
<b>D1</b>	0,322 (SD= 0,137)	0,281 (SD= 0,182)	0,362 (SD= 0,127)	0,275 (SD= 0,053)	0,259 (SD= 0)	0,283 (SD= 0,073)
<b>D2</b>	0,137 (SD= 0)	-	0,137 (SD= 0)	0,152 (SD= 0)	-	0,152 (SD= 0)
<b>E1</b>	0,108 (SD= 0,059)	0,087 (SD= 0,073)	0,128 (SD= 0,046)	0,169 (SD= 0,113)	0,212 (SD= 0,169)	0,137 (SD= 0,057)
<b>E2</b>	0,192 (SD= 0,156)	0,325 (SD= 0)	0,172 (SD= 0,154)	0,157 (SD= 0,051)	0,159 (SD= 0)	0,157 (SD= 0,057)
<b>Medias totales</b>	<b>0,199 (SD= 0,177)</b>	<b>0,202 (SD= 0,084)</b>	<b>0,214 (SD= 0,066)</b>	<b>0,196 (SD= 0,076)</b>	<b>0,198 (SD= 0,0856)</b>	<b>0,199 (SD= 0,057)</b>

Tabla 14: Resultados de disimilitud en el plano intradocumental, con datos desglosados por lengua y categoría. Entre paréntesis, desviación estándar (SD)

El grado de disimilitud en este análisis oscila entre 0,073 (A1) y 0,322 (D1) en el caso del español y de 0,14 (B2) a 0,337 (C2) para el inglés. Observamos que las medias son similares entre lenguas (0,199 y 0,196, respectivamente, para el español e inglés) y entre categorías.

En definitiva, los datos son solo ligeramente superiores en el plano intercomité y no se constatan grandes diferencias entre lenguas o categorías.

### c) Análisis intertextual 2

En la tabla 15 figura el número de unidades de análisis repetidos en al menos dos comités.

Lengua	Textos normativos	Entidades corporativas nacionales	Total
Español	64	71	135
Inglés	65	73	138

Tabla 15: Número de unidades de análisis identificadas para el estudio del efecto de la variable temática, con datos desglosados por categoría y por lengua

Resulta llamativo el elevado número de esas unidades, lo que confirma que numerosas EN tienen una presencia transversal.

Para el análisis de las soluciones de traducción al español, de las 138 unidades de análisis eliminamos 2 entidades corporativas y 1 texto normativo, pues solo presentaban una ocurrencia en dos informes de dos comités distintos y en uno de ellos no aparecía la traducción. Por ello, para el análisis en español consideramos el número indicado en la tabla 15.

En la tabla 16 se presentan y comparan los resultados obtenidos tras los análisis en los planos intracomité e intercomité. Todos los datos se expresan en términos totales y porcentuales.

	Español			Inglés		
	Total	Textos normativos	Entidades corporativas	Total	Textos normativos	Entidades corporativas
Totales coincidentes en el plano intracomité	n=56	n=26	n=30	n=35	n=15	n=20
	41,48 %	40 %	41,09 %	25,36 %	23,08 %	27,40 %
Totales coincidentes en el plano intercomité	n=121	n=61	n=60	n=88	n=45	n=43
	89,63 %	95,31 %	82,2 %	63,77 %	69,23 %	58,90 %

Tabla 16: Heterogeneidad comparada de los planos intracomité e intercomité, con datos desglosados por lengua y categoría, y expresados en términos porcentuales y totales

En inglés, el 25,36 % de las unidades analizadas presentan heterogeneidad y, en español, el 41,48 %. No se observan grandes diferencias entre categorías de EN en ninguna de las dos lenguas.

El 30 % de las unidades de análisis heterogéneas en inglés presentan 2 soluciones de traducción y el 84 %, entre 2 y 6 soluciones. La media de soluciones de traducción al inglés es de 4,45. El número máximo de soluciones de traducción en esta lengua es 17 para «قانون الإجراءات الجزائية» [ley de los procedimientos penales].

En español, el 32,8 % de las unidades de análisis heterogéneas presentan 2 soluciones de traducción y el 80 %, entre 2 y 6 soluciones de traducción. La media de soluciones de traducción al español es 4,52. El número máximo de soluciones de traducción halladas es de 17 para un solo elemento «قانون الجرائم والعقوبات» [ley de los delitos y las penas]. Las cifras, por tanto, son muy similares a las del inglés.

Al igual que en los análisis anteriores, el número de unidades de análisis heterogéneas es mayor en español que en inglés (89,6 % frente a 63,77 %) y, en ambas lenguas, la homogeneidad es menor en los textos normativos. De forma consecuente, el porcentaje de falta de homogeneidad mayor se registra en la traducción al español de textos normativos (95,31 %) y el menor, en la traducción al inglés de las entidades corporativas nacionales (58,9 %).

Los datos comparados entre los dos planos confirman un mayor grado de homogeneidad en los datos de un solo comité que en los de varios comités.

La tabla 17 presenta los resultados del análisis de la disimilitud en el plano intracomité e intercomité.

Disimilitud						
Plano	Español			Inglés		
	Total	Textos normativos	Entidades corporativas nacionales	Total	Textos normativos	Entidades corporativas nacionales
Plano intracomité	0,213 (SD= 0,119)	0,232 (SD= 0,118)	0,177 (SD= 0,115)	0,226 (SD= 0,119)	0,234 (SD= 0,12)	0,213 (SD= 0,118)
Plano intercomité	0,2 (SD= 0,098)	0,224 (SD= 0,094)	0,176 (SD= 0,097)	0,164 (DST= 0,102)	0,162 (SD= 0,109)	0,167 (SD= 0,096)

Tabla 17: Resultados del análisis de la disimilitud en el plano intracomité e intercomité, desglosados por lengua. Entre paréntesis, desviación estándar (SD, según sus siglas en inglés)

El grado de disimilitud no se ve afectado por la variable temática. Es más, para las dos lenguas, la media de las disimilitudes es menor en términos totales en el plano intercomité que en el intracomité. En relación con las categorías de EN, el grado de disimilitud en los textos normativos es también mayor que en las entidades corporativas.

## 2. Análisis de los resultados

### a) Relevancia de los resultados

En base a los datos obtenidos, podemos afirmar que:

- El elevado número de EN confirma la importancia de estos elementos en los documentos de la ONU analizados;
- La falta de homogeneidad intratextual en la traducción de las unidades de análisis es notable;
- En la traducción de los informes dirigidos a un mismo comité, la falta de homogeneidad es mayor en la traducción de varios documentos que en la de un solo documento, por tanto, la falta de homogeneidad intertextual es mayor que la intratextual;
- La falta de homogeneidad intertextual es mayor entre los informes enviados a varios comités que entre los informes enviados a un solo comité;
- El grado de heterogeneidad es significativamente mayor en español que en inglés, con independencia del tipo de análisis realizado y de la categoría de EN;

- El grado de heterogeneidad es mayor en la traducción de textos normativos que en la traducción de entidades corporativas nacionales, con independencia de la lengua y del tipo de análisis realizado;
- La disimilitud entre las cadenas de caracteres que componen las soluciones de traducción heterogéneas no revela diferencias notables por la lengua, la categoría de EN o el tipo de análisis;
- El grado de disimilitud se sitúa en medias cercanas o superiores al 20 %.

Por tanto, han quedado refutadas la hipótesis nula principal y las hipótesis secundarias, tanto en la traducción al inglés como al español. Sin embargo, se observan importantes diferencias en el grado de heterogeneidad entre categorías de EN y lenguas. En el grado de disimilitud entre las cadenas de caracteres que componen las soluciones de traducción heterogéneas, por el contrario, no se observan diferencias notables entre lenguas, categorías de EN o tipo de análisis.

A continuación analizamos con mayor detalle estos resultados.

La falta de homogeneidad en el análisis intratextual es llamativa, pese a ser notablemente menor que la intertextual. Ningún informe presenta una homogeneidad del 100 %. Solo D2 presenta una homogeneidad del 100 % en las traducciones de los textos normativos al español y el inglés. Cabe esperar que la falta de homogeneidad en este plano tenderá a multiplicar la heterogeneidad intertextual, y en efecto así lo demuestran nuestros datos.

A la vista de la literatura previa, la falta de homogeneidad intratextual puede atribuirse a la escasa presencia en UNTERM de las categorías de EN árabes analizadas y a la estrechez de los plazos de entrega, con la división del trabajo y la dificultad de revisión que ello acarrea. Sin embargo, la diferencia notable en el grado de heterogeneidad intratextual entre lenguas y categorías de EN no puede explicarse atendiendo solo a los dos factores señalados, puesto que esos factores son igualmente aplicables al inglés y a cualquier tipo de unidad de análisis.

Por ello, la interpretación de nuestros datos exige tomar en consideración la concurrencia de factores adicionales. Recordemos que la literatura previa, académica e institucional, subrayaba que la incorporación del árabe como lengua oficial en la ONU había acarreado complicaciones de reclutamiento y capacitación y obligado en ocasiones a recurrir al relé. Por ello, debíamos considerar esta circunstancia como un potencial factor que debíamos

tener en cuenta, a diferencia de Guzmán y Prieto Ramos (2021).

La interpretación de nuestros datos exigía, por tanto, determinar si en la muestra hubo relé. Para ello se realizó un análisis cualitativo complementario y, en efecto, los paralelismos respecto al inglés observados en las traducciones al español (tabla 18) y las discrepancias respecto al original repetidas en la traducción a ambas lenguas (tabla 19) apoyan la idea de que al menos algunas de las traducciones analizadas estaban intermediadas por el inglés. La posibilidad de intermediación del español es descartable.

A1 (inglés)	A1 (español)	A2 (inglés)	A2 (español)
Consultative Council	Consejo Consultivo	Shura (Advisory) Council	Consejo de la Shura (asesor)
Penal Code	Código Penal	Code of Offences and Penalties	Código de Delitos y Penas
Code of Conduct for Lawyers	Código de conducta de los abogados	Law Profession Regulation Act	Ley de ejercicio de la profesión jurídica

Tabla 18: Ejemplos de paralelismos entre el español y el inglés en los informes del comité A

Informe	Unidad de análisis	Traducción al español	Traducción al inglés
A1	القانون رقم 47 لسنة 1991 بشأن عملية دخول وخروج الأجانب [ley número 47 del año 1991 de la entrada y la salida de los extranjeros]	Ley N° 48 de 1999 relativa a la entrada y salida de extranjeros	Act No. 48 of 1999, concerning the entry and exit of aliens
B1	قانون العمال رقم 5 لسنة 1995 [ley de los trabajadores número 5 del año 1995]	Labour Act No. 5 of 1970	Ley del trabajo N° 5 de 1970
C1	القانون رقم 52 لسنة 1990 بشأن الصحافة والمطبوعات [ley número 52 del año 1990 de la prensa y las publicaciones]	Ley de prensa y publicaciones N° 25 de 1990	Press and Publications Act No. 25 of 1990
D1	قانون الخدمة المدنية والإصلاح الإداري رقم 19 [ley de la administración pública y la reforma administrativa número 19]	Ley de la administración pública y la reforma administrativa N° 199	Civil Service and Administrative Reform Act No. 199
D1	قانون السلطة القضائية رقم 1 لسنة 1990 [ley del poder judicial número 1 del año 1990]	Ley de la autoridad judicial N° 13 de 1990	Judicial Authority Act No. 13 of 1990

Tabla 19: Ejemplos de discrepancias con el original repetidas en la traducción al español y al inglés (resaltadas en negrita)

El relé, por tanto, podría estar afectando negativamente a la calidad del producto entendida en términos de homogeneidad.

Por otro lado, Guzmán y Prieto Ramos (2021), tras analizar la variación en textos de diversas extensiones, sugieren que parece existir una correlación entre la extensión del texto analizado y el grado de heterogeneidad intratextual. Ese fenómeno podría explicarse considerando el incremento en el número de traductores humanos intervinientes en el proceso, lo que sería igualmente aplicable al relé. En definitiva, nuestros datos, unidos a los de esos autores, indicarían que cuanto mayor es el número de traductores y/o revisores implicados, ya sea por el volumen de texto o por cualquier otra circunstancia del encargo, inclusive la necesidad de recurrir al relé, mayor será la tendencia a la entropía en el producto.

Esa misma explicación podría aplicarse al interpretar la mayor heterogeneidad en los análisis intertextuales que en el intratextual. La implicación de un mayor número de traductores humanos se sumaría entonces al posible efecto de las variables diacrónica y temática.

Otro elemento, en este caso relativo al proceso de documentación, también pudiera estar reflejándose negativamente en el producto en español: el hecho de que, aunque la presencia de las EN analizadas en UNTERM sea igualmente escasa en ambas lenguas, la posibilidad de documentación en Internet es incomparablemente mayor para el inglés. No nos referimos exclusivamente a la posibilidad general de documentación en la red: la realidad es que, en la práctica, los Estados árabes son bilingües árabe/inglés (en Medio Oriente) o árabe/francés (en el Norte de África), con independencia de que constitucionalmente se declaren como tales.

Así pues, siendo el Yemen un país de Medio Oriente, será el propio Estado el que en sus páginas web oficiales, proponga, si es que lo hace, denominaciones en inglés para sus EN. Por tanto, con independencia de que la EN en cuestión se recoja en UNTERM, los traductores al inglés y al español (traduciendo estos desde el árabe o desde el inglés) se verán enfrentados a procesos de documentación de naturaleza muy distinta. En el caso de la traducción al español, y ante la falta de respuesta de UNTERM, el elemento principal de referencia no será la propuesta del propio Estado, sino las traducciones previas realizadas en o para la ONU y estas, como hemos podido comprobar, son heterogéneas.

Otra circunstancia explicaría la mayor heterogeneidad en nuestros datos que en los datos de Guzmán y Prieto Ramos (2021). Nos referimos a la escasa o nula presencia en la red de información administrativa, incluso en árabe, sobre Estados poco desarrollados, como el Yemen, en comparación con los Estados originarios de la muestra analizada por los autores citados. De hecho, la página web del Ministerio de Justicia del Yemen (<https://moj.gov.ye/>) se encuentra inhabilitada desde fecha que no podemos determinar.

Por otro lado, los resultados obtenidos del cálculo de la disimilitud, frente a los datos relativos al grado de heterogeneidad, no revelan tendencias diferenciadas claras entre lenguas, y solo una pequeña diferencia entre categorías. No podemos comparar estos datos con los de la literatura previa en traducción, cuya existencia no nos consta. Sin embargo, la literatura relativa a la eficacia de diversos sistemas de medición de cadenas de caracteres, inclusive Jaro-Winkler y Levenshtein, determinan que esos sistema solo son eficaces a partir de un 80 % de similitud (Wang et al., 2017), e incluso en ese umbral es frecuente la detección de falsos positivos (Vathy-Fogarassy et al., 2020).

A la vista de la literatura previa especializada, por tanto, y a falta de confirmación con una herramienta suficientemente desarrollada para la lengua árabe, podemos afirmar que las posibilidades de reconocimiento automático eficaz de entidades nombradas con el grado de disimilitud obtenido en nuestros análisis son escasas.

#### **b) Aportaciones al estado de la cuestión**

Hemos constatado que la literatura previa sobre la traducción en los organismos supranacionales se caracterizaba por:

- i. Ser escasa, en particular la relativa a la ONU;
- ii. Centrarse principalmente en la traducción jurídica en la UE;
- iii. Basarse en la experiencia profesional de los autores o entrevistados, lo que explica que la mayoría se centrara en divulgar los problemas del traductor, es decir, en el proceso;
- iv. Carecer de análisis empíricos del producto y, en caso de proporcionar datos empíricos, recopilarlos mediante encuestas.



En lo que a la literatura en lingüística computacional respecta, no hemos hallado estudios en los que se emplee una herramienta eficaz de REN para la lengua árabe. Solo los estudios de Samy y su equipo profundizan en un tipo documental muy similar al nuestro y en nuestra misma combinación lingüística, aunque manifiestan las dificultades que entraña el etiquetado en lengua árabe y se ven obligados a su procesamiento manual.

En cuanto a la homogeneidad en traducción, la literatura previa se caracterizaba por:

- i. Emplear córpora limitados o que difícilmente puede considerarse que cumplen los requisitos metodológicos impuestos por la lingüística de corpus;
- ii. Orientarse a demostrar que las equivalencias correctas en materia de EN se fundan en relaciones de correspondencia unívoca.

Guzmán y Prieto Ramos (2021) supone un hito respecto a toda esa literatura previa. En el marco del proyecto LETRINT analizan el mismo entorno de traducción y tipo documental que nosotros, con objetivos similares (salvo que se orientan a la evaluación de la calidad y nosotros a la recuperación automática de la información) y un corpus de naturaleza y extensión temporal idénticas. La gran diferencia entre el trabajo citado y el nuestro radica en la combinación lingüística y la direccionalidad, cuyos únicos antecedentes son Feria (2013) y Roldán y Feria (en prensa).

Es lógico, pues, que las aportaciones principales de nuestra investigación respecto a Guzmán y Prieto Ramos (2021) deriven de la diferente combinación lingüística abordada. Como hemos podido comprobar, nuestra combinación supone, entre otras, las diferencias siguientes:

1) la intermediación de una tercera lengua en el proceso (relé), con la consiguiente involucración de un número mayor de actores. Este punto exigirá atención en todas las investigaciones posteriores sobre la traducción del árabe en este entorno institucional;

2) las diferentes posibilidades de documentación, en el seno de la institución y fuera de ella, mucho menores en el caso del árabe, y

3) el menor grado de «transparencia» entre lo traducido y su traducción, toda vez que el análisis de Guzmán y Prieto Ramos (2021) versa sobre lenguas más cercanas (por su origen indoeuropeo) y empleadas por Estados con una estrecha relación política, jurídica y cultural, nada de lo cual se cumple en nuestro caso.

Entre nuestra investigación y la de Guzmán y Prieto Ramos (2021) median también diferencias metodológicas notables:

4) Ellos analizan términos de varios Estados y nosotros de uno solo, ya que abordan las unidades de análisis en cuanto que elementos terminológicos y nosotros en cuanto que EN. Adviértase que una EN no tiene necesariamente que traducirse del mismo modo para varios Estados, en particular si se trata de Estados multilingües con sus propias propuestas de traducción oficial y ninguna de ellas particularmente «transparente». A cambio, en nuestra investigación ampliamos el número de unidades de análisis de 30 a varios cientos;

5) Ellos alinearon el corpus de manera automática para la extracción de datos, lo que en nuestro caso no fue posible por la falta de herramientas adecuadas. Esa carencia de motores de alineación efectivos, producto entre otras cosas de las características ortotipográficas y discursivas de la lengua árabe, como demuestran Roldán y Feria (en prensa), tiene un gran impacto negativo en la investigación de la traducción del árabe;

6) Ellos no consideraron las diferencias ortográficas o derivadas de la flexión de las unidades de análisis debido a su enfoque terminológico. En nuestro caso, al no abordar esas unidades en cuanto que EN, era imprescindible considerar esas diferencias.

En el plano profesional e institucional, nuestra investigación demuestra la necesidad de estandarizar en mayor medida las EN en lengua árabe en las bases de datos terminológicas de la Organización, en particular las más ocurrentes. Esta necesidad es aún mayor debido a la necesidad práctica de acudir al relé al traducir del árabe.

Por otro lado, creemos haber demostrado que el diseño de una herramienta de REN eficaz en árabe, tan útil para diversos fines, exige una labor previa de análisis multilingüe que, en definitiva, es la que hemos comenzado en esta investigación. En el futuro esperamos desarrollar en gran medida esa línea de investigación gracias a los datos recopilados.

## Capítulo V: Conclusiones

### 1. Descripción general de la investigación

En esta tesis doctoral se ha analizado la homogeneidad en la traducción del árabe al español e inglés de las entidades nombradas (EN) que figuran en informes presentados por la República del Yemen a los órganos con competencia en la ONU para la supervisión del cumplimiento de los derechos humanos. La homogeneidad en la traducción de esos elementos promueve la efectividad del diálogo internacional y asegura la calidad presente y futura del producto de la traducción.

El proceso de diálogo entre los órganos de las Naciones Unidas y los Estados, con su consecuente intercambio documental, tiene a menudo un carácter multilingüe. En nuestro caso, el órgano pertinente de la ONU se pronuncia en inglés, como es habitual, y la República del Yemen informa en árabe. Además, la política de multilingüismo de la Organización dicta que todos los documentos estén disponibles en todas las lenguas oficiales. El papel que desempeñan los Servicios de Traducción en este contexto es, por tanto, fundamental. El escenario descrito resultaba muy idóneo para nuestros propósitos de investigación.

En esta tesis doctoral se analiza la homogeneidad en la traducción de dos categorías de EN: textos normativos, categoría no incluida en las taxonomías de EN actuales hasta donde se nos alcanza y cuya misma definición constituye un aporte de esta tesis, y entidades corporativas nacionales. La amplia presencia de estas categorías de EN en el corpus analizado demostró que se trata de elementos clave para el diálogo descrito.

Con objeto de medir la potencial falta de homogeneidad en la traducción de las categorías de EN señaladas realizamos dos análisis. Por un lado, cuantificamos el porcentaje de unidades de análisis traducidas de manera heterogénea y, por otro, el grado de disimilitud

en las traducciones de las unidades heterogéneas. En relación con el grado de disimilitud, abordamos las soluciones de traducción como cadenas de caracteres y medimos la diferencia entre ellas con el algoritmo de la distancia de Jaro. En la literatura previa no nos consta que este tipo de medición se haya aplicado en el ámbito de la Traductología. La metodología empleada, por tanto, debe también considerarse una aportación novedosa de esta investigación.

Para realizar los análisis descritos era necesario extraer los datos de un corpus apropiado. Así pues, en primer lugar, procedimos a la compilación de un corpus paralelo de informes de Estado seleccionados de un corpus mayor de textos presentados a las Naciones Unidas en lengua árabe. El tipo documental se determinó mediante criterios de relevancia (los informes de Estado, por su extensión, eran los mejores candidatos al contener un mayor número de unidades de análisis) y disponibilidad. Nuestro corpus incluyó, finalmente, 10 informes presentados por el Yemen ante 5 comités de seguimiento, así como sus versiones en español e inglés. En total, el corpus comprende 1 054 569 palabras.

Durante la fase exploratoria (elaboración del TFM) se previó extraer del corpus automáticamente las unidades de análisis mediante herramientas de Reconocimiento de Entidades Nombradas; sin embargo, ninguna de las herramientas disponibles para la lengua árabe resultó satisfactoria. Por ello, la extracción de las unidades de análisis se realizó de manera semi automática.

A continuación resumimos los capítulos de la tesis doctoral.

## **2. Resumen de los capítulos**

En el capítulo 1 se presenta la tesis doctoral, se justifica la elección del tema, se determinan las hipótesis y objetivos y se contextualiza el entorno de traducción que define el problema de investigación.

En el capítulo 2 se repasa la literatura previa sobre la traducción en los organismos internacionales y la homogeneidad en la traducción, los corpus de textos de la ONU y los sistemas de recuperación y extracción de información en lengua árabe. El capítulo 2 también aporta información relevante extraída de informes institucionales.

En el capítulo 3 se describen el corpus y la metodología empleados.

En el capítulo 4 se presentan y analizan los datos considerando el entorno de traducción y el estado de la cuestión descritos en el capítulo 2.

Finalmente, en el capítulo 5 se presentan las conclusiones de esta tesis doctoral.

A continuación exponemos los objetivos y las hipótesis considerados en la investigación.

### **3. Refutación de las hipótesis y consecución de los objetivos**

La presente investigación planteó la siguiente hipótesis nula principal:

*Hipótesis principal:* En las traducciones del árabe al español y el inglés realizadas en o para la ONU no se registra falta de homogeneidad en las denominaciones de los textos normativos y las entidades corporativas nacionales.

Dicha hipótesis quedó refutada. Hemos logrado demostrar mediante datos estadísticos sólidos y desde una perspectiva computacional que en nuestro corpus se incurre en falta de homogeneidad en la traducción del árabe al español y el inglés de las categorías de EN analizadas.

De refutarse la hipótesis nula principal, en esta tesis se planteaban dos hipótesis nulas secundarias:

*Hipótesis secundaria 1:* En la traducción de los informes dirigidos a un mismo comité, esa falta de homogeneidad no es mayor en la traducción de varios documentos que en la traducción de un solo documento.

*Hipótesis secundaria 2:* Esa falta de homogeneidad no es mayor en la traducción de los informes dirigidos a varios comités que en la traducción de los informes dirigidos a un solo comité.

La hipótesis secundaria 1, que apunta a la variable diacrónica, y la hipótesis secundaria 2, que no necesariamente apunta a esa variable y se centra en el posible impacto del cambio de temática, también quedaron refutadas. Demostramos que, en el corpus analizado, la falta de homogeneidad en la traducción de las categorías de EN pertinentes

se agrava con la distancia temporal y temática, y en particular cuando se constata la recurrencia al relé.

Nuestra investigación también ha logrado sus objetivos. El más general de ellos era **promover el conocimiento sobre la traducción en los organismos internacionales, en particular en la ONU**. Para ello hemos descrito el contexto institucional del producto analizado, en particular los ciclos de diálogos, los procesos de traducción y las herramientas y capacidades de los Servicios de Traducción, inclusive considerando la información contenida en los informes pertinentes de la ONU. Adviértase la importancia de este entorno institucional, ya que ciertas combinaciones lingüísticas, como la contemplada en esta tesis, solo pueden analizarse en él.

En concreto, pretendíamos **analizar empíricamente el producto de la traducción del árabe en la ONU**. La literatura previa sobre la traducción en organismos internacionales era muy escasa y, salvo las últimas aportaciones de Guzmán y Prieto Ramos (2021), Prieto Ramos (2020), Prieto Ramos y Guzmán (2018), y Roldán y Feria (en prensa), se centraba en el análisis del proceso desde la perspectiva de los propios traductores. De esa literatura, solo Roldán & Feria, en prensa, y nuestra investigación analizan empíricamente el producto en la traducción del árabe. La consecución de este objetivo, por tanto, constituye una contribución notable a la disciplina.

Nuestra investigación también pretendía contribuir a **promover el procesamiento automático de la lengua árabe y las herramientas terminológicas en esa lengua**. Este objetivo se ha alcanzado, en primer lugar, al mostrar las graves carencias actuales en la materia. Hemos identificado algunas de esas carencias y confirmado que las herramientas de extracción automática de información para la lengua árabe distan de ser efectivas y que es necesario mejorar las herramientas terminológicas disponibles en el contexto institucional estudiado.

Finalmente, con la consecución de los objetivos anteriores, hemos logrado **tender puentes entre la investigación en Traductología y el mundo profesional de la traducción**.

#### 4. Principales resultados

A continuación presentamos los principales resultados obtenidos en nuestra investigación.

Nuestro análisis, pese al escaso desarrollo del procesamiento automático de la lengua árabe, ha demostrado la viabilidad de la novedosa metodología propuesta. Gracias a ella hemos alcanzado conclusiones estadísticas sólidas y hemos revelado la necesidad de mejorar el rendimiento de las herramientas para el procesamiento automático de la lengua árabe existentes, así como hemos acumulado una gran cantidad de datos que contribuirán en última instancia a mejorar esas herramientas. La mejora de tales herramientas, por otro lado, contribuirá a asegurar la homogeneidad traductora en contextos internacionales multilingües y promoverá la calidad del producto en todas las lenguas oficiales, con independencia de la combinación lingüística y de la necesidad de recurrir al relé.

Esta investigación ha revelado también la amplia presencia de las categorías de EN analizadas y su importancia para el diálogo internacional. Hemos contabilizado un total de 1039 unidades de análisis (507 textos normativos y 532 entidades corporativas nacionales). Los análisis realizados confirman la falta de homogeneidad intratextual e intertextual en la traducción de esas categorías de EN árabes al español y el inglés en la ONU. Si el fenómeno es tan claro en relación con las EN, en las que no cabía esperar heterogeneidad alguna, cabe hipotetizar que esa falta de homogeneidad podría extenderse a otros planos.

Los resultados obtenidos han revelado, además, diferencias notables según la lengua y el tipo de categoría de EN. El porcentaje de unidades heterogéneas es siempre mayor en español que en inglés (5,4 % en el análisis intratextual; 13 % y 6,5 % en el análisis intertextual 1 y 16,12 % y 25,86 % en el análisis intertextual 2). En lo concerniente a la heterogeneidad en función de la categoría analizada, los valores son similares en el análisis intratextual. En cambio, se observa una mayor heterogeneidad en los análisis intertextual 1 y 2 en la categoría «texto normativo» que en la categoría «entidad corporativa nacional».

El cálculo de la disimilitud entre las soluciones de traducción ha revelado que no existen diferencias claras entre lenguas y es mayor en la categoría «texto normativo» que en la

categoría «entidad corporativa nacional». No obstante, los valores obtenidos (en torno al 20 % de disimilitud) son un hándicap para la recuperación de esas entidades nombradas.

La diferencia en el grado de heterogeneidad debida a la variable lingüística no se consideró de antemano y, por tanto, no se contemplaba en el planteamiento de las hipótesis nulas. Este descubrimiento resulta de interés particular desde una perspectiva teórica y práctica. Como señalaba la literatura previa, las lenguas oficiales de la ONU no conviven en paridad. Nuestra investigación, por vez primera, ha comenzado a reunir datos estadísticos que confirman la existencia de relé y el impacto de esa práctica en el producto.

### **5. Líneas futuras de investigación**

La ingente cantidad de información extraída abre la posibilidad de iniciar, entre otras, las líneas de investigación futura siguientes:

- Ampliar el corpus utilizado y las categorías de EN árabes analizadas. La ausencia de herramientas REN suficientemente desarrolladas para la lengua árabe imposibilitó realizar los análisis sobre un corpus de mayores dimensiones. El estudio de categorías de EN adicionales abriría la puerta al análisis del siguiente punto;
- Analizar las posibles causas de la diferencia en el grado de homogeneidad entre diferentes categorías de EN;
- Analizar lingüísticamente las unidades de análisis para, en ausencia de mayúsculas, promover su etiquetado automático y la capacidad de las herramientas de traducción automática para gestionarlas;
- Alimentar las herramientas de REN en lengua árabe para mejorar su efectividad;
- Categorizar en mayor medida las diferentes formas de traducir las EN árabes y aplicar esos conocimientos al ejercicio de la profesión y a la docencia.

En definitiva, la presente tesis doctoral, gracias a la aplicación de una metodología computacional muy novedosa en los estudios de traducción, ha contribuido a un mejor conocimiento de la traducción del árabe en un contexto multilingüe supranacional, en particular en lo relativo a la homogeneidad en la traducción de las categorías de EN «texto normativo» y «entidad corporativa nacional». Por ello, consideramos que esta tesis



doctoral supone el inicio de una línea de investigación novedosa y con un gran potencial para el futuro.



# **Conclusiones en lengua inglesa para la obtención de la mención internacional en el título de Doctora**

## **1. General description of the research**

This PhD thesis analyzed consistency in the translation of Arabic Named Entities (NE) into Spanish and English. These were identified in reports submitted by the Republic of Yemen to the human rights monitoring committees of the United Nations (UN). The consistent translation of these elements promotes the effectiveness of international dialogue, as well as ensuring present and future quality of the product.

The dialogue between the United Nations' committees and the States, along with the document exchange it entails, is often of a multilingual nature. In the case being researched research, the UN committee expresses itself in English, as usually occurs, and the Republic of Yemen reports in Arabic. Besides, the multilingualism policy of the Organization establishes that all documents produced must be made available in all their official languages. In this context, the Translation Services play an essential role. Hence, the context described was ideal for our purposes.

This thesis analyzed the consistency in the translation of two NE categories: regulation names, not included to date in any of the available taxonomies to the best of our knowledge and, therefore, its definition is considered a significant contribution of this research, and national corporate entity names. These NE categories proved to be key elements in the described dialogue given their large presence in the analyzed corpus.

To analyze the potential degree of inconsistency in the translation of the NE categories selected, we performed two quantitative analyses. On the one hand, we quantified the degree of elements inconsistently translated. On the other hand, we calculated the degree of variation in the inconsistent translation of elements. In the second analysis, we

considered the translation solutions as strings of characters and measured their variation with an adapted version of the Jaro distance metric. We have not seen this metric applied in previous literature in the scope of Translation Studies. The methodology, therefore, must also be considered a novel contribution of this research to the state of the art.

Given the described analyses, the compilation of a suitable corpus was necessary to extract the data. Therefore, we compiled a parallel corpus that comprised State party's reports extracted from a larger corpus of documents presented to the UN in the Arabic language. The document type was selected on the basis of availability and relevance criteria (State party's reports are the lengthiest in the dialogue between State and Member, and, consequently, more likely to include a larger number of elements). As a result, our corpus included 10 State party's reports submitted by the Republic of Yemen to five different monitoring committees, as well as their versions in Spanish and English. The corpus comprised 1,054,569 words.

During the exploratory phase of this research (final master's degree dissertation), it had been intended to employ Named Entity Recognition tools for the automatic extraction of the analyzed elements. Unfortunately, none of the available tools proved to be suitable for our purposes. Therefore, the element extraction was conducted semiautomatically.

The following section describes the chapters of the thesis.

## **2. Brief description of the chapters**

Chapter 1 presents the doctoral thesis, justifies the topic addressed, explains the objectives, and presents the hypotheses. Ultimately, it provides a description of the translation context of the question subject to research.

Chapter 2 describes the state of the art on translation at international organizations, translation consistency, text corpora at the United Nations and information retrieval and extraction systems for the Arabic language. In addition, we present relevant information from institutional reports.

Chapter 3 describes the corpus and methodology used for data extraction.

Chapter 4 presents the data obtained and analyses the results considering the translation context and the state of the art presented in chapter 2.

Finally, chapter 5 presents the conclusions of this doctoral thesis.

### **3. Refutation of hypotheses and achievement of objectives**

This research established the following main null hypothesis:

*Main hypothesis:* There is no lack of consistency in the translation of national corporate entity names and regulation names from Arabic into Spanish and English in reports translated at or by the United Nations.

Our analysis refuted this hypothesis. We demonstrated with reliable data and from a computational perspective that our corpus presents a lack of consistency in the translation from Arabic into English and Spanish of the analyzed NE categories.

Had the main null hypothesis been refuted, two secondary null hypotheses would have been considered:

*Secondary hypothesis 1:* In the translation of reports addressed to the same committee, the lack of consistency is no greater in the translation of several documents than in that of a single document.

*Secondary hypothesis 2:* The lack of consistency is no greater in the translation of reports addressed to various committees than in reports addressed to a single committee.

The secondary hypothesis 1, considering a time variable, and the secondary hypothesis 2, not necessarily considering the time variable but the possible impact of change of the subject matter, were also refuted. Our analyses showed that the lack of consistency in the translation of the aforementioned NE categories is negatively affected when considering such variables. This is particularly aggravated when resorting to relay.

This research has also achieved our main objectives. The most general of them all was the **promoting of knowledge on translation at international organizations and particularly at the UN**. To this end, we described the institutional setting of the product

under analysis, particularly the cycle of dialogues, the processes of translation, and the tools and capabilities of the Translation Services. We also considered the information contained in reports published by the UN. The significance of this institutional setting is crucial, as some of its linguistic combinations, such as the one analyzed in this thesis, can only be found at the UN.

Specifically, we intended to **empirically analyze the product of translation from the Arabic language at the UN**. Previous literature on translation at international organizations is scarce (except for the most recent contributions such as Guzmán and Prieto Ramos (2021); Prieto Ramos (2020); Prieto Ramos and Guzmán (2018); Roldán and Feria (forthcoming)) and it focuses on the analysis of the translation process from the viewpoint of the Organization's translators. From the list of authors mentioned, only Roldán and Feria (forthcoming) and the present thesis analyze the product in the translation from the Arabic language from an empirical perspective. Achieving this objective constitutes a significant contribution to this discipline.

The aim of our research was also to contribute to the **promoting of automatic processing of the Arabic language and terminology tools in Arabic**. We attained this objective by showing the current serious shortcomings in this matter. We identified some of these shortcomings and confirmed that the automatic information extraction tools available for the Arabic language are far from efficient. This led us to conclude that there is a need for further development of terminological tools available in the institutional setting under study.

Finally, having achieved the previous objectives, we have contributed to **building bridges between Translation Studies and the profession of translation**.

#### **4. Main results**

This section presents the main results of our research.

Our analyses demonstrate, despite the insufficient development of tools for automatic processing for the Arabic language, the feasibility of the novel methodology proposed. With this methodology, we have drawn statistically sound conclusions and we have

revealed the need to improve the performance of automatic processing of tools for the Arabic language. In addition, the vast amount of data gathered will ultimately contribute to the improvement of such tools, which, on the other hand, will contribute to ensuring consistency in the translation at multilingual international settings. Thus, it will promote the quality of the product in all the official languages, regardless of the use of linguistic combination or the need to resort to relay.

This research has confirmed the wide presence of the NE analyzed categories and their importance for international dialogue. We have accounted for a total of 1,039 elements (507 regulation names and 532 national corporate entity names). The statistical analyses conducted demonstrate the lack of intratextual and intertextual consistency in the translation of these NE categories from Arabic into Spanish and English at the UN. By extension, if this phenomenon is so clear in relation to NEs, where a lack of consistency is not to be expected, we can hypothesize that inconsistency could be present elsewhere.

The results obtained also show that there are significant differences depending on the language and NE category. The percentage of inconsistent elements is always greater in Spanish than in English (5.4 % in the intratextual analysis; 13 % y 6.5 % in the first intertextual analysis and 16.12 % y 25.86 % in the second intertextual analysis). With regard to the lack of consistency by analyzed category, the results are similar in the intratextual analysis. However, greater inconsistency was revealed in the regulation names category in both the first and second intertextual analyses.

The quantification of the degree of dissimilarity between translation solutions showed no relevant differences between languages and it is greater in regulation names than in national corporate entity names. Nevertheless, the results obtained (20 % of dissimilarity approximately) are a true handicap for the retrieval of these NEs.

Differences in the degree of inconsistency caused by the language variable were not considered from a starting point and, therefore, this variable was not included in the establishment of our null hypothesis. This finding is of particular interest from a theoretical and practical perspective. As stated in previous literature, the official languages of the UN do not coexist equally. This research is the first to gather data that confirm the existence of relay and the impact it has on the product.

## **5. Future research lines**

The vast amount of information extracted opens the possibility for further research.

Future work directions include:

- Compiling a larger corpus and analyzing additional Arabic NE categories. The inexistence of Named Entity Recognition sufficiently developed for the Arabic language prevented us from analyzing a greater corpus. The study of additional NE categories would open the door to the following point;
- Analyzing the possible causes of the differences observed in the degree of consistency between different NE categories;
- Analyzing linguistically the NEs to promote, in the absence of capital letters, their automatic annotation and the capacity of automatic translation tools to manage them efficiently;
- Training Named Entity Recognition tools in Arabic to improve their effectiveness;
- Further categorizing the different ways of translating Arabic NEs and applying the knowledge gained to the exercising of the profession and teaching of translation.

Ultimately, the present PhD thesis, alongside the application of a novel computational methodology in Translation Studies, has contributed with an increased understanding of translation from the Arabic language at a supranational multilingual setting, particularly regarding consistency in the translation of two NE categories: regulation names and national corporate entity names. Therefore, this PhD thesis should be considered the beginning of an innovative line of research with great future potential.



## Bibliografia

- 1st International Conference on Arabic Computational Linguistics*. (2015).  
<http://site.cicling.org/2015/acling.html>
- Abdallah, S., Shaalan, K., & Shoaib, M. (2012). Integrating Rule-Based System with Classification for Arabic Named Entity Recognition. *Proceedings of the 13th International Conference on Intelligent Text Processing and Computational Linguistics*, 311-322.  
[https://www.researchgate.net/profile/Khaled\\_Shaalan/publication/233728240\\_Integrating\\_Rule-Based\\_System\\_with\\_Classification\\_for\\_Arabic\\_Named\\_Entity\\_Recognition/links/0912f50ad03c6ece08000000.pdf](https://www.researchgate.net/profile/Khaled_Shaalan/publication/233728240_Integrating_Rule-Based_System_with_Classification_for_Arabic_Named_Entity_Recognition/links/0912f50ad03c6ece08000000.pdf)
- Abdelali, A., Darwish, K., Durrani, N., & Mubarak, H. (2016). Farasa: A Fast and Furious Segmenter for Arabic. En J. DeNero, M. Finlayson, & S. Reddy (Eds.), *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations* (pp. 11-16). Association for Computational Linguistics. <http://alt.qcri.org/~ndurrani/pubs/farasa-fast-furious.pdf>
- Alghamdi, M., & Teahan, W. (2017). Experimental evaluation of Arabic OCR systems. *PSU Research Review*, 1(3), 229-241. <https://doi.org/10.1108/PRR-05-2017-0026>
- Aljanati, L. P. (2017). Promoting Multilingual Consistency for the Quality of EU Law. *International Journal for the Semiotics of Law - Revue Internationale de Sémiotique Juridique*, 30(1), 67-79. <https://doi.org/10.1007/s11196-016-9482-9>
- Alotaiby, F., Alkharashi, I., & Foda, S. (2009). Processing Large Arabic Text Corpora: Preliminary Analysis and Results. *Proceedings of the Second International Conference on Arabic Language Resources and Tools*, 78-82.
- Alotaiby, F., Foda, S., & Alkharashi, I. (2014). Arabic vs. English: Comparative

- Statistical Study. *Arabian Journal for Science and Engineering*, 39(2), 809-820.  
<https://doi.org/10.1007/s13369-013-0665-3>
- Alotaiby, F., Foda, S., & Alkharashi, I. (2010). Clitics in Arabic language: A statistical study. En R. Ootoguro, K. Ishikawa, H. Umemoto, K. Yoshimoto, & Y. Harada (Eds.), *PACLIC 24 - Proceedings of the 24th Pacific Asia Conference on Language, Information and Computation* (pp. 595-601). Institute of Digital Enhancement of Cognitive Processing, Waseda University. <http://www.aclweb.org/anthology/Y10-1068>
- Alston, P., & Goodman, R. (2013). *International Human Rights: The Successor to International Human Rights in Context*. Oxford University Press.
- Alto Comisionado de las Naciones Unidas para los Derechos Humanos. (s. f.). *Los órganos de derechos humanos*. Recuperado 15 de agosto de 2018, de <https://www.ohchr.org/SP/HRBodies/Pages/HumanRightsBodies.aspx>
- Amón, I., & Jiménez, C. (2010). Funciones de Similitud sobre Cadenas de Texto: Una Comparación Basada en la Naturaleza de los Datos. *International Conference on Information Resources Management (CONF-IRM)*, 13.  
[https://s3.amazonaws.com/academia.edu.documents/31732366/71644758.20104.pdf?response-content-disposition=inline%3Bfilename%3DFunciones\\_de\\_Similitud\\_sobre\\_Cadenas\\_de.pdf&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=AKIAIWOWYYGZ2Y53UL3A%2F20190827%2Fu](https://s3.amazonaws.com/academia.edu.documents/31732366/71644758.20104.pdf?response-content-disposition=inline%3Bfilename%3DFunciones_de_Similitud_sobre_Cadenas_de.pdf&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=AKIAIWOWYYGZ2Y53UL3A%2F20190827%2Fu)
- Asamblea General de las Naciones Unidas. (1948). *Declaración Universal de Derechos Humanos. Resolución 217 A (III)*.
- Asamblea General de las Naciones Unidas. (1965). *Convención Internacional sobre la Eliminación de todas las Formas de Discriminación Racial. Resolución 2106 A (XX)*. [https://www.ohchr.org/Documents/ProfessionalInterest/cerd\\_SP.pdf](https://www.ohchr.org/Documents/ProfessionalInterest/cerd_SP.pdf)
- Asamblea General de las Naciones Unidas. (1973). *Inclusión del árabe entre los idiomas oficiales y de trabajo de la Asamblea General y sus Comisiones Principales. Resolución 3190 (XXVIII)*. <https://documents-dds-ny.un.org/doc/RESOLUTION/GEN/NR0/287/15/IMG/NR028715.pdf?OpenElement>

nt

- Asamblea General de las Naciones Unidas. (1974). *Traducción al Alemán de Algunos Documentos Oficiales de la Asamblea General y de las Resoluciones del Consejo de Seguridad y del Consejo Económico y Social. Resolución 3355 (XXIX)*.
- Asamblea General de las Naciones Unidas. (1989). *Convención sobre los Derechos del Niño. Resolución 44/25*.
- Asamblea General de las Naciones Unidas. (1993). *Alto Comisionado para la promoción y la protección de todos los derechos humanos. Resolución A/RES/48/141*.
- Asamblea General de las Naciones Unidas. (2006a). *Consejo de Derechos Humanos. Resolución A/RES/60/251*.
- Asamblea General de las Naciones Unidas. (2006b). *Protocolo Facultativo de la Convención sobre los Derechos de las Personas con Discapacidad*.  
<https://www.ohchr.org/SP/ProfessionalInterest/Pages/OptionalProtocolRightsPersonsWithDisabilities.aspx>
- Asamblea General de las Naciones Unidas. (2017). *Multilingüismo - Informe del Secretario General A/71/757*. <https://doi.org/10.18268/bsgm1908v4n1x1>
- Benajiba, Y. (2009). Universidad polit ecnica de madrid [Universidad Politécnica de Valencia]. En *Arabic Named Entity Recognition*.  
<https://riunet.upv.es/bitstream/handle/10251/8318/tesisUPV3049.pdf?sequence=1>
- Benajiba, Y., & Rosso, P. (2008). Arabic Named Entity Recognition using Conditional Random Fields. *Proceedings of Workshop on HLT & NLP within the Arabic World, LREC*, 143-153.  
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.330.8645&rep=rep1&type=pdf>
- Benajiba, Y., Rosso, P., Miguel, J., & Ruiz, B. (2007). ANERsys: An Arabic Named Entity Recognition System Based on Maximum Entropy. En A. Gelbukh (Ed.), *Proceedings of the 8th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing'07)* (pp. 143-153). Springer Berlin

Heidelberg.

[http://s3.amazonaws.com/academia.edu.documents/43691863/ANERsys\\_An\\_Arabic\\_Named\\_Entity\\_Recogniti20160313-6053-7qvxvgr.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1487001034&Signature=tc7HIoHQ8VObPur0RgKHRZP9co4%3D&response-content-disposition=inline%3B](http://s3.amazonaws.com/academia.edu.documents/43691863/ANERsys_An_Arabic_Named_Entity_Recogniti20160313-6053-7qvxvgr.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1487001034&Signature=tc7HIoHQ8VObPur0RgKHRZP9co4%3D&response-content-disposition=inline%3B)

Bou Franch, V., & Castillo Daudí, M. (2010). *Curso de Derecho Internacional de los Derechos Humanos* (2.<sup>a</sup> ed.). Tirant lo Blanch. <http://biblioteca.nubedelectura.com/cloudLibrary/ebook/show/9788499851259>

Bowker, L., & Pearson, J. (2002). *Working with Specialized Language a Practical Guide to Using Corpora*. Routledge. <https://ebookcentral.proquest.com/lib/ugr/reader.action?docID=178476>

Buckwalter, T. (2004). *Buckwalter Arabic Morphological Analyzer Version 2.0*. Linguistic Data Consortium.

Cao, D., & Zhao, X. (2008). Translation at the United Nations as Specialized Translation. *The Journal of Specialised Translation*, 1(9), 39-54. [http://www.jostrans.org/issue09/art\\_cao.pdf](http://www.jostrans.org/issue09/art_cao.pdf)

*Capitalization in English*. (s. f.). United Nations Editorial Manual Online. Recuperado 16 de noviembre de 2018, de <http://dd.dgacm.org/editorialmanual/ed-guidelines/style/capitalization.htm>

Carbonell, M., & Pedroza de la Llave, S. T. (2000). Elementos de técnica legislativa. En *Serie Doctrina Jurídica* (1.<sup>a</sup> ed., Número 312). Universidad Nacional Autónoma de México. <https://archivos.juridicas.unam.mx/www/bjv/libros/5/2149/2.pdf>

Carpuat, M., & Simard, M. (2012). The Trouble with SMT Consistency. *Proceedings of the Seventh Workshop on Statistical Machine Translation*, 442-449.

Cervilla Garzón, M. D. (2018). La aplicabilidad de las normas del Código de Familia marroquí (la Mudawana) que regulan el divorcio en España: el filtro constitucional. *Cuadernos De Derecho Transnacional*, 10(1), 144-163. <https://doi.org/10.20318/cdt.2018.4119>

- Chen, Y., & Eisele, A. (2012). MultiUN v2: UN Documents with Multilingual Alignments. En N. Calzolari, K. Choukri, T. Declerck, M. U. Doğan, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC, 2012)* (pp. 2500-2504).
- Christen, P. (2006). A Comparison of Personal Name Matching: Techniques and Practical Issues. *Sixth IEEE International Conference on Data Mining - Workshops (ICDMW'06)*, 290-294. <https://doi.org/10.1109/ICDMW.2006.2>
- Comisión Europea. (2012). *Translation and Multilingualism*. <https://publications.europa.eu/en/publication-detail/-/publication/e0770e72-afa1-4971-8824-6190512537dc/language-en/format-PDF/source-77214905>
- Consejo de Derechos Humanos de las Naciones Unidas. (s. f.). *Ciclos del Examen Periódico Universal*. Recuperado 23 de octubre de 2018, de <https://www.ohchr.org/SP/HRBodies/UPR/Pages/CyclesUPR.aspx>
- Consejo de Derechos Humanos de las Naciones Unidas. (2007). *Construcción institucional del Consejo de Derechos Humanos. Resolución A/HRC/RES/5/1*.
- Cunningham, H. (2005). Information Extraction, Automatic. En *Encyclopedia of Language and Linguistics* (2.<sup>a</sup> ed., pp. 665–677). Elsevier. [http://ac.els-cdn.com/B0080448542009603/3-s2.0-B0080448542009603-main.pdf?\\_tid=3a4ca0e4-dd6a-11e6-930d-00000aacb361&acdnat=1484736079\\_3001e461f411a013fef59a6ab59dcc24](http://ac.els-cdn.com/B0080448542009603/3-s2.0-B0080448542009603-main.pdf?_tid=3a4ca0e4-dd6a-11e6-930d-00000aacb361&acdnat=1484736079_3001e461f411a013fef59a6ab59dcc24)
- Darwish, K. (2014). Information Retrieval. En Zitouni (Ed.), *Natural Language Processing of Semitic Languages* (pp. 299–334). Springer. <http://link.springer.com/10.1007/978-3-642-45358-8>
- Darwish, K., Abdelali, A., & Mubarak, H. (2014). Using stem-templates to improve Arabic pos and gender/number tagging. En N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the 9th International Conference on Language Resources and Evaluation, LREC 2014* (pp. 2926-2931). <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.708.9239&rep=rep1&ty>

pe=pdf

Darwish, K., & Gao, W. (2014). Simple Effective Microblog Named Entity Recognition: Arabic as an Example. En N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odiijk, & S. Piperidis (Eds.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. European Languages Resources Association (ELRA). [http://www.lrec-conf.org/proceedings/lrec2014/pdf/186\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2014/pdf/186_Paper.pdf)

Delgado Loyola, J. C. (2015). Hacia un Algoritmo Optimo de Emparejamiento de Nombres. *Analitika: Revista de análisis estadístico*, 9, 95-109. [http://www.ecuadorencifras.gob.ec/documentos/web-inec/Revistas/Analitika/Anexos\\_pdf/Analit\\_09/4.pdf](http://www.ecuadorencifras.gob.ec/documentos/web-inec/Revistas/Analitika/Anexos_pdf/Analit_09/4.pdf)

Departamento de Información Pública de las Naciones Unidas. (s. f.). *Normas mínimas para el multilingüismo de los sitios web de las Naciones Unidas*. Recuperado 31 de octubre de 2018, de <http://www.un.org/es/sections/web-governance/minimum-standards-multilingualism-united-nations-websites/index.html>

Departamento de la Asamblea General y de Gestión de Conferencias de las Naciones Unidas. (s. f.-a). *División de Documentación*. Recuperado 23 de octubre de 2018, de [http://www.un.org/es/hq/dgacm/documentation\\_division.shtml](http://www.un.org/es/hq/dgacm/documentation_division.shtml)

Departamento de la Asamblea General y de Gestión de Conferencias de las Naciones Unidas. (s. f.-b). *Servicio de Interpretación*. Recuperado 23 de octubre de 2018, de <http://www.un.org/es/hq/dgacm/interpretation.shtml>

Departamento de la Asamblea General y de Gestión de Conferencias de las Naciones Unidas. (s. f.-c). *Traducción*. Recuperado 23 de octubre de 2018, de <http://www.un.org/es/hq/dgacm/translation/>

Departamento de la Asamblea General y de Gestión de Conferencias de las Naciones Unidas. (2015). *USER'S GUIDE TO eLUNa*. United Nations.

Dependencia Común de Inspección de las Naciones Unidas. (1980). *JIU/REP/80/7. Evaluación del proceso de traducción en el Sistema de las Naciones Unidas*. [https://www.unjiu.org/sites/www.unjiu.org/files/jiu\\_document\\_files/products/es/re](https://www.unjiu.org/sites/www.unjiu.org/files/jiu_document_files/products/es/re)

ports-notes/JIU Products/JIU\_REP\_1980\_7\_Spanish.pdf

Dependencia Común de Inspección de las Naciones Unidas. (2003). *JIU/REP/2002/11. Aplicación del multilingüismo en el sistema de las Naciones Unidas*. <https://doi.org/10.1111/j.1365-294X.2011.05463.x>

Dependencia Común de Inspección de las Naciones Unidas. (2011). *JIU/REP/2011/4. El multilingüismo en las organizaciones del sistema de las Naciones Unidas: estado de aplicación*.

[https://www.unjiu.org/sites/www.unjiu.org/files/jiu\\_document\\_files/products/es/reports-notes/JIU\\_Products/JIU\\_REP\\_2011\\_4\\_Spanish.pdf](https://www.unjiu.org/sites/www.unjiu.org/files/jiu_document_files/products/es/reports-notes/JIU_Products/JIU_REP_2011_4_Spanish.pdf)

Dependencia Común de Inspección de las Naciones Unidas. (2020). *JIU/REP/2020/6. El multilingüismo en el sistema de las Naciones Unidas*. [https://www.unjiu.org/sites/www.unjiu.org/files/jiu\\_rep\\_2020\\_6\\_spanish\\_0.pdf](https://www.unjiu.org/sites/www.unjiu.org/files/jiu_rep_2020_6_spanish_0.pdf)

Dingare, S., Finkel, J., Nissim, M., Manning, C., & Grover, C. (2004). A System For Identifying Named Entities in Biomedical Text: How Results From Two Evaluations Reflect on Both the System and the Evaluations. *Proceedings of the 9th Annual Meeting of the BioLINK SIG at ISMB: Linking Literature, Information and Knowledge for Biology*. <http://www-nlp.stanford.edu/~manning/papers/ismb2004.pdf>

Eisele, A., & Chen, Y. (2010). MultiUN: A Multilingual Corpus from United Nation Documents. En N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odiijk, S. Piperidis, M. Rosner, & D. Tapias (Eds.), *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)* (pp. 2868-2872). European Language Resources Association (ELRA). <https://pdfs.semanticscholar.org/bcd6/de970fb4b91262f3ced4abc05694a70aa751.pdf>

Elsayed, H., & Elghazaly, T. (2015). A Named Entities Recognition System for Modern Standard Arabic using Rule-Based Approach. *First International Conference on Arabic Computational Linguistics*, 51–54. <https://doi.org/10.1109/ACLing.2015.14>

Fall, P. L., & Zhang, Y. (2011). *Multilingualism in the United Nations System Organizations: Status of Implementation*. Joint Inspection Unit, United Nations.

[https://www.unjiu.org/en/reports-notes/JIU\\_Products/JIU\\_REP\\_2011\\_4\\_English.pdf](https://www.unjiu.org/en/reports-notes/JIU_Products/JIU_REP_2011_4_English.pdf)

- Farghaly, A., & Shaalan, K. (2009). Arabic Natural Language Processing: Challenges and solutions. *ACM Transactions on Asian Language Information Processing (TALIP)*, 8(4), 1–22. <https://doi.org/10.1145/1644879.1644881>
- Feria García, M. (2013). Algunos obstáculos en el proceso traductor de textos de las Naciones Unidas sobre derechos humanos redactados en árabe, o redactados en otras lenguas oficiales si incluyen conceptos islámicos. En *Translating the Law. Theoretical and Methodological Issues* (1.ª ed., pp. 47–55). Comares.
- Fernández-Vitores, D. (2014). El español en el sistema de las Naciones Unidas. *Informes del Observatorio / Observatorio Reports*, 25. <https://doi.org/10.15427/OR004-10/2014SP>
- Fernández Orquín, A., Díaz Blanco, J., Fundora Rolo, A., & Muñoz Guillena, R. (2009). Un algoritmo para la extracción de características lexicográficas en la comparación de palabras An algorithm for the extraction of lexicographical characteristics in the comparison of words. *IV Convención Científica Internacional CIUM*.
- Florian, R., Hassan, H., Ittycheriah, A., Jing, H., Kambhatla, N., Luo, X., Nicolov, N., Roukos, S., & Zhang, T. (2004). A Statistical Model for Multilingual Entity Detection and Tracking. En D. Marcu, S. Dumais, & S. Roukos (Eds.), *Proceedings of the Human Language Technologies Conference (HLT-NAACL2004)* (pp. 1-8). Association for Computational Linguistics. [http://acl.ldc.upenn.edu/hlt-naacl2004/main/pdf/128\\_Paper.pdf%5Cnpapers2://publication/uuid/1F731352-9B63-4D94-A8F3-C2945E5B097C](http://acl.ldc.upenn.edu/hlt-naacl2004/main/pdf/128_Paper.pdf%5Cnpapers2://publication/uuid/1F731352-9B63-4D94-A8F3-C2945E5B097C)
- Footnotes and other references. IV. Outside sources. E. Legal references.* (s. f.). United Nations Editorial Manual Online. Recuperado 16 de noviembre de 2018, de [http://dd.dgacm.org/editorialmanual/ed-guidelines/footnotes/footnotes\\_chap\\_04.htm#E](http://dd.dgacm.org/editorialmanual/ed-guidelines/footnotes/footnotes_chap_04.htm#E)
- García-Barrero, D., Feria, M., & Turell, M. T. (2013). Using function words and punctuation marks in Arabic forensic authorship attribution. En R. Sousa-Silva, R. Faria, N. Gavaldà, & B. Maia (Eds.), *Proceedings of the 3rd European Conference*



- of the International Association of Forensic Linguists* (pp. 42-56). Faculdade de Letras da Universidade do Porto. <https://ler.letras.up.pt/uploads/ficheiros/13605.pdf>
- García, R. (2010). El español, lengua de traducción en el sistema de las Naciones Unidas y en el de la Comunidad Europea. *Actas del IV Congreso «El Español, Lengua de Traducción»*, 395-403. [https://cvc.cervantes.es/lengua/esletra/pdf/04/043\\_garcia.pdf](https://cvc.cervantes.es/lengua/esletra/pdf/04/043_garcia.pdf)
- Gil Bardají, A. (2008). *Procedimientos, técnicas, estrategias: operadores del proceso traductor*. <http://www.recercat.cat/handle/2072/8998>
- Gil Bardají, A. (2009). Procedures, techniques, strategies: Translation process operators. *Perspectives: Studies in Translatology*, 17(3), 161-173. <https://doi.org/10.1080/09076760903249372>
- Grishman, R., & Sundheim, B. (1996). Message Understanding Conference -6: A Brief History. *Coling*, 96, 466-471. [http://www.alt.aasn.au/events/altss\\_w2003\\_proc/altss/courses/molla/C96-1079.pdf](http://www.alt.aasn.au/events/altss_w2003_proc/altss/courses/molla/C96-1079.pdf)
- Grzybek, P. (2007). History and methodology of word length studies. En P. Grzybek (Ed.), *Contributions to the Science of Text and Language: Word Length Studies and Related Issues* (pp. 15-90). Springer.
- Guellil, I., Saâdane, H., Azouaou, F., Gueni, B., & Nouvel, D. (s. f.). Arabic natural language processing: An overview. *Journal of King Saud University - Computer and Information Sciences*. <https://doi.org/10.1016/j.jksuci.2019.02.006>
- Guillou, L. (2013). Analysing Lexical Consistency in Translation. En B. Webber, A. Popescu-Belis, K. Markert, & J. Tiedemann (Eds.), *Proceedings of the Workshop on Discourse in Machine Translation* (pp. 10-18). Association for Computational Linguistics. <http://www.aclweb.org/anthology/W13-3302>
- Guzmán, D., & Prieto Ramos, F. (2021). Assessing legal terminological variation in institutional translation: The case of national court names in the human rights monitoring procedures of the United Nations. *Translation and Translanguaging in Multilingual Contexts*, 7(2), 224-247. <https://doi.org/10.1075/ttmc.00067.guz>

- Habash, N. (2010). Introduction to Arabic Natural Language Processing. *Synthesis Lectures on Human Language Technologies*, 3(1), 1-187. <https://doi.org/10.2200/S00277ED1V01Y201008HLT010>
- Habash, N., Rambow, O., & Roth, R. (2009, enero). MADA+TOKAN: A toolkit for Arabic tokenization, diacritization, morphological disambiguation, POS tagging, stemming and lemmatization. *Proceedings of the Second International Conference on Arabic Language Resources and Tools (MEDAR)*. [https://www.researchgate.net/publication/228529836\\_MADATOKAN\\_A\\_toolkit\\_for\\_Arabic\\_tokenization\\_diacritization\\_morphological\\_disambiguation\\_POS\\_tagging\\_stemming\\_and\\_lemmatization](https://www.researchgate.net/publication/228529836_MADATOKAN_A_toolkit_for_Arabic_tokenization_diacritization_morphological_disambiguation_POS_tagging_stemming_and_lemmatization)
- Hanna, S., El-Farahaty, H., & Abdel-Wahab, K. (Eds.). (2020). *Routledge Handbook of Arabic Translation* (1.<sup>a</sup> ed.). Routledge Taylor & Francis Group.
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2006). Metodología de la investigación. En *Metodología de la investigación* (4.<sup>a</sup> ed.). McGraw-Hill. <https://doi.org/-> ISBN 978-92-75-32913-9
- International Federation of Library Associations and Institutions. (2016). *Declaración de principios internacionales de catalogación (PIC)* (p. 22). [https://doi.org/10.1300/J104v38n01\\_03](https://doi.org/10.1300/J104v38n01_03)
- Itagaki, M., Aikawa, T., & He, X. (2013). *Validation of the Consistency of Automatic Terminology Translation* (Patent N.º US Patent 8,548,791). Google Patents.
- Itagaki, M., Aikawa, T., & He, X. (2007). Automatic validation of terminology translation consistency with statistical method. En B. Maegaard (Ed.), *Proceedings of MT Summit XI* (pp. 269-274). European Association for Machine Translation. <http://www.eamt.org>
- Jacquemin, C., & Bourigault, D. (2003). Term extraction and automatic indexing. *The Oxford Handbook of Computational Linguistics, February 2017*, 599-615. <https://doi.org/10.1093/oxfordhb/9780199276349.013.0033>
- Kang, J.-H. (2009). Institutional translation. En M. Baker & G. Saldanha (Eds.), *Routledge Encyclopedia of Translation Studies* (2.<sup>a</sup> ed., pp. 141-145). Routledge

Taylor & Francis Group.

- Khalifa, S., Zalmout, N., & Habash, N. (2016). YAMAMA: Yet another multi-dialect Arabic morphological analyzer. *Proceedings of COLING 2016 - 26th International Conference on Computational Linguistics: System Demonstrations*, 223-227. <https://www.aclweb.org/anthology/C16-2047.pdf>
- Koskinen, K. (2000). Institutional Illusions: Translating in the EU Commission. *The Translator*, 6(1), 49-65.
- Koskinen, K. (2008). *Translating Institutions: An Ethnographic Study of EU Translation*. St. Jerome Publishing.
- Koskinen, K. (2011). Institutional translation. En Y. Gambier & L. van Doorslaer (Eds.), *Handbook of Translation Studies: Volume 2* (Vol. 2, pp. 54-60). John Benjamins Publishing Company. <https://doi.org/10.1080/0907676X.2012.698072>
- Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Dartmouth Publishing Group.
- Lafeber, A. (2012). *Translation at inter-governmental organizations the set of skills and knowledge required and the implications for recruitment testing* [Universitat Rovira I Virgili]. <http://www.tesisenred.net/handle/10803/83500?show=full>
- Leaman, R., & Lu, Z. (2016). TaggerOne: Joint named entity recognition and normalization with semi-Markov Models. *Bioinformatics*, 32(18), 2839-2846. <https://doi.org/10.1093/bioinformatics/btw343>
- Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, 10(8), 707-710. <https://doi.org/citeulike-article-id:311174>
- Li, M., Munkhdalai, T., Yu, X., & Ryu, K. H. (2015). A Novel Approach for Protein-Named Entity Recognition and Protein-Protein Interaction Extraction. *Mathematical Problems in Engineering*, 2015. <https://doi.org/10.1155/2015/942435>
- Llopis, R. F., Muñoz, A., Suárez, A., & Montoyo. (1998). *EXIT: Propuesta de un sistema de extracción de información de textos notariales* (p. 10).

- [https://www.researchgate.net/publication/228881698\\_EXIT\\_Propuesta\\_de\\_un\\_sistema\\_de\\_extraccion\\_de\\_informacion\\_de\\_textos\\_notariales](https://www.researchgate.net/publication/228881698_EXIT_Propuesta_de_un_sistema_de_extraccion_de_informacion_de_textos_notariales)
- Manhire, V. (Ed.). (2018). *United Nations Handbook* (56.<sup>a</sup> ed.). Ministry of Foreign Affairs and Trade.
- Manning, C. D., & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press.
- Medea-García, L. (2017). *Gramaticalización y cambio lingüístico en árabe El caso de hattà en lengua clásica y dialectal* [Universitat Autònoma de Barcelona]. [https://ddd.uab.cat/pub/tesis/2017/hdl\\_10803\\_459246/lmg11de1.pdf](https://ddd.uab.cat/pub/tesis/2017/hdl_10803_459246/lmg11de1.pdf)
- Melamed, I. D. (1997). Measuring semantic entropy. *ACL-SIGLEX Workshop Tagging Text with Lexical Semantics: Why, What, and How*, 41–46. <http://acl.ldc.upenn.edu/W/W97/W97-0207.pdf>
- Merkel, M. (1996). Consistency and Variation in Technical Translations: a Study of Translators' Attitudes. *Unity in Diversity? Current Trends in Translation Studies*, 137-149. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.449.4159&rep=rep1&type=pdf>
- Mesmia, F. Ben, Friburger, N., Haddar, K., & Maurel, D. (2015). Transducers Cascades for an Automatic Recognition of Arabic Named Entities in Order to Establish Links to Free Resources. En A. Gelbukh & K. Shaalan (Eds.), *First International Conference on Arabic Computational Linguistics (ACLing'15)* (pp. 61–67). IEEE Computer Society. <https://doi.org/10.1109/ACLing.2015.16>
- Mesmia, F. Ben, Haddar, K., Friburger, N., & Maurel, D. (2018). CasANER: Arabic Named Entity Recognition Tool. En *Intelligent Natural Language Processing: Trends and Applications* (pp. 173-198). Springer, Cham.
- Mohammed, N. F., & Omar, N. (2012). Arabic named entity recognition using artificial neural network. *Journal of Computer Science*, 8(8), 1285-1293. <https://doi.org/10.3844/jcssp.2012.1285.1293>

- Mohit, B. (2014). Named Entity Recognition. En Zitouni (Ed.), *Natural Language Processing of Semitic Languages* (pp. 221–245). Springer. <http://link.springer.com/10.1007/978-3-642-45358-8>
- Monroe, W., Green, S., & Manning, C. D. (2014). Word segmentation of informal Arabic with domain adaptation. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014*, 2, 206-211. <https://doi.org/10.3115/v1/p14-2034>
- Moreno Fernández, F., & Otero Roth, J. (2016). *Atlas de la lengua española en el mundo* (3.<sup>a</sup> ed.). Ariel: Fundación Telefónica. [https://www.researchgate.net/profile/Francisco\\_Moreno-Fernandez/publication/312192247\\_Atlas\\_de\\_la\\_lengua\\_espanola\\_en\\_el\\_mundo\\_3\\_ed\\_ampliada\\_y\\_actualizada/links/58756c2b08ae8fce49282383/Atlas-de-la-lengua-espanola-en-el-mundo-3-ed-ampliada-y-actualizada.pdf](https://www.researchgate.net/profile/Francisco_Moreno-Fernandez/publication/312192247_Atlas_de_la_lengua_espanola_en_el_mundo_3_ed_ampliada_y_actualizada/links/58756c2b08ae8fce49282383/Atlas-de-la-lengua-espanola-en-el-mundo-3-ed-ampliada-y-actualizada.pdf)
- Naciones Unidas. (s. f.-a). *Dependencia Común de Inspección del Sistema de las Naciones Unidas*. Supervisión independiente del Sistema para la eficiencia, la efectividad y la coordinación. <https://www.unju.org/es/es>
- Naciones Unidas. (s. f.-b). *Estados no Miembros*. Recuperado 8 de febrero de 2019, de <http://www.un.org/es/sections/member-states/non-member-states/index.html>
- Naciones Unidas. (s. f.-c). *Multilingüismo*. Recuperado 23 de octubre de 2018, de <https://www.un.org/sg/es/multilingualism/index.shtml>
- Naciones Unidas. (s. f.-d). *UNTERM portal - Frequently asked questions about the UN terminology search portal*. Recuperado 13 de noviembre de 2018, de <https://untermportal.un.org/UNTERM/Home/HowTo>
- Naciones Unidas. (1945). *Carta De Las Naciones Unidas*.
- Naciones Unidas. (2009). *Compilación de directrices relativas a la forma y el contenido de los informes que deben presentar los Estados Partes en los tratados internacionales de derechos humanos. HRI/GEN/2/Rev.6*.
- Nadeau, D., & Sekine, S. (2007). A survey of named entity recognition and classification.

*Linguisticae Investigationes*, 30(1), 3–26. <https://doi.org/10.1075/li.30.1.03nad>

Neifar, W., Hamon, T., Zweigenbaum, P., Khemakhem, M. E., & Hadrich Belguith, L. (2018). Adaptation of a Term Extractor to Arabic Specialised Texts: First Experiments and Limits. En Alexander Gelbukh (Ed.), *Computational Linguistics and Intelligent Text Processing. 17th International Conference, CICLing 2016* (pp. 242-253). Springer. <https://link.springer.com/content/pdf/10.1007%2F978-3-319-75477-2.pdf>

Nóbrega, M. (2008). La traducción en las Naciones Unidas: la Torre de Babel a orillas del East River. *Traducción: contacto y contagio, Actas del III Congreso Internacional «El español, lengua de traducción»*, 135–142. [http://cvc.cervantes.es/lengua/esletra/pdf/03/013\\_nobrega.pdf](http://cvc.cervantes.es/lengua/esletra/pdf/03/013_nobrega.pdf)

North Atlantic Treaty Organization. (2010). *Final Communiqué of the first Session of the North Atlantic Council - (Terms of Reference and Organisation)*. [https://www.nato.int/cps/en/natolive/official\\_texts\\_17117.htm](https://www.nato.int/cps/en/natolive/official_texts_17117.htm)

Oficina de las Naciones Unidas en Ginebra. (s. f.-a). *About UNOG - Annual Report*. Recuperado 9 de noviembre de 2018, de [https://www.unog.ch/C125781000589799/\(httpPages\)/96B4529FBC62AF1E80256EF3005DC06B?OpenDocument](https://www.unog.ch/C125781000589799/(httpPages)/96B4529FBC62AF1E80256EF3005DC06B?OpenDocument)

Oficina de las Naciones Unidas en Ginebra. (s. f.-b). *What is the difference between the official language and the working language?* Recuperado 23 de octubre de 2018, de [https://www.unog.ch/80256EE60057CB67/\(httpPages\)/0CB2852A9D3C6635C1257EF80074780A?OpenDocument](https://www.unog.ch/80256EE60057CB67/(httpPages)/0CB2852A9D3C6635C1257EF80074780A?OpenDocument)

Oficina de las Naciones Unidas en Ginebra. (2012). *Annual Report 2011 - Meeting Current Challenges Together*. [http://www.unog.ch/80256EDD006B8954/\(httpAssets\)/22E1AAC4CBDC37E7C12579EC00323EBE/\\$file/Whole+report\\_Eng.pdf](http://www.unog.ch/80256EDD006B8954/(httpAssets)/22E1AAC4CBDC37E7C12579EC00323EBE/$file/Whole+report_Eng.pdf)

Oficina de las Naciones Unidas en Ginebra. (2015). *Annual Report 2014 - Working for Peace Rights Well-being*. [http://www.unog.ch/80256EDD006B8954/\(httpAssets\)/7106F4F9939E641FC1257E0A00338037/\\$file/15\\_02039E\\_PDF\\_2.pdf](http://www.unog.ch/80256EDD006B8954/(httpAssets)/7106F4F9939E641FC1257E0A00338037/$file/15_02039E_PDF_2.pdf)

- Oficina de las Naciones Unidas en Ginebra. (2016). *UNOG Annual Report 2015*.  
[https://www.unog.ch/80256EDD006B8954/\(httpAssets\)/41E55FE6507FE262C1257F7A00417CB8/\\$file/1603591\\_INTERACTIF\\_18\\_05\\_PDF.pdf](https://www.unog.ch/80256EDD006B8954/(httpAssets)/41E55FE6507FE262C1257F7A00417CB8/$file/1603591_INTERACTIF_18_05_PDF.pdf)
- Oficina de las Naciones Unidas en Ginebra. (2017). *United SDGs - Annual Report 2016*.  
[https://www.unog.ch/80256EDD006B8954/\(httpAssets\)/CC631258F86F2728C12580E90031DA91/\\$file/AR2016\\_interactivePDF.pdf](https://www.unog.ch/80256EDD006B8954/(httpAssets)/CC631258F86F2728C12580E90031DA91/$file/AR2016_interactivePDF.pdf)
- Oficina de las Naciones Unidas en Ginebra. (2018). *Change - Annual Report 2017*.  
[https://www.unog.ch/80256EDD006B8954/\(httpAssets\)/18BFA24FE0439B50C125826D00457747/\\$file/ENGLISH+RA17\\_ODG\\_signets.pdf](https://www.unog.ch/80256EDD006B8954/(httpAssets)/18BFA24FE0439B50C125826D00457747/$file/ENGLISH+RA17_ODG_signets.pdf)
- Oficina de las Naciones Unidas en Ginebra. (2020). *Unity & Resilience: 75 years of global cooperation - Annual Report 2020*.  
<https://sites.ungeneva.org/annualreport/2020/>
- Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos. (s. f.). *Guía práctica para la sociedad civil. Examen Periódico Universal*.  
[https://www.ohchr.org/Documents/AboutUs/CivilSociety/Universal\\_Periodic\\_Review\\_SPA.pdf](https://www.ohchr.org/Documents/AboutUs/CivilSociety/Universal_Periodic_Review_SPA.pdf)
- Oyarzún Hernández, N., & Vergara Salas, P. (2014). *Sistemas de Extracción de Información. Construcción automática de diccionarios de patrones para sistemas EI* (pp. 1–10).
- Pasha, A., Al-Badrashiny, M., Diab, M., El Kholy, A., Eskander, R., Habash, N., Pooleery, M., Rambow, O., & Roth, R. M. (2014). MADAMIRA: A fast, comprehensive tool for morphological analysis and disambiguation of Arabic. En N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the 9th International Conference on Language Resources and Evaluation, LREC 2014* (Vol. 14, pp. 1094-1101). [http://www.lrec-conf.org/proceedings/lrec2014/pdf/593\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2014/pdf/593_Paper.pdf)
- Patman, F., & Thompson, P. (2003). Names: A new frontier in text mining. *International Conference on Intelligence and Security Informatics*, 27-38.  
[https://link.springer.com/chapter/10.1007/3-540-44853-5\\_3](https://link.springer.com/chapter/10.1007/3-540-44853-5_3)

- Piskorski, J., & Yangarber, R. (2013). Information Extraction: Past, Present and Future. En T. Poibeau, H. Saggion, J. Piskorski, & R. Yangarber (Eds.), *Multi-source, multilingual information extraction and summarization* (pp. 23-49). Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-28569-1>
- Pouliquen, B., Elizalde, C., Junczys-Dowmunt, M., Mazenc, C., & García-Verdugo, J. (2013). Large-scale multiple language translation accelerator at the United Nations. *Proceedings of the XIV Machine Translation Summit*, 345-352. <http://www.mtsummit2013.info/files/proceedings/main/mt-summit-2013-pouliquen-et-al.pdf%5Cnpapers3://publication/uuid/D2124836-2693-47A6-BF86-8D214702D946>
- Prieto Ramos, F. (2018). Surveying Institutional Translation. En F. Prieto Ramos (Ed.), *Institutional Translation for International Governance: Enhancing Quality in Multilingual Legal Communication*. Bloomsbury.
- Prieto Ramos, F. (2020). Ensuring Consistency and Accuracy of Legal Terms in Institutional Translation: The Role of Terminological Resources in International Organizations. En F. Prieto Ramos (Ed.), *Institutional Translation and Interpreting: Assessing Practices and Managing for Quality* (pp. 128-149). Routledge. <https://doi.org/10.4324/9780429264894-10>
- Prieto Ramos, F., & Guzmán, D. (2018). Legal Terminology Consistency and Adequacy as Quality Indicators. En F. Prieto Ramos (Ed.), *Institutional Translation for International Governance: Enhancing Quality in Multilingual Legal Communication* (pp. 81-100). Bloomsbury.
- Prieto Ramos, F., & Guzmán, D. (2021). Examining institutional translation through a legal lens: A comparative analysis of multilingual text production at international organizations. *Target. International Journal of Translation Studies*, 33(2), 254-281. <https://doi.org/10.1075/target.21003.pri>
- Rafalovitch, A., & Dale, R. (2009). United Nations General Assembly Resolutions: A Six-Language Parallel Corpus. *Proceedings of the MT Summit XII*, 12, 292-299. [https://www.researchgate.net/profile/Robert\\_Dale4/publication/228579662\\_United\\_nations\\_general\\_assembly\\_resolutions\\_A\\_six-](https://www.researchgate.net/profile/Robert_Dale4/publication/228579662_United_nations_general_assembly_resolutions_A_six-)



language\_parallel\_corpus/links/5463068a0cf2837efdb00b3a/United-nations-general-assembly-resolutions-A-six-language-parallel-corpus.pdf

Rogers, M. (2008). Terminological Equivalence: Probability and Consistency. *LSP Translation Scenarios*, 2, 101-108.  
[http://www.translationconcepts.org/pdf/MuTra\\_Journal2\\_2008.pdf#page=101](http://www.translationconcepts.org/pdf/MuTra_Journal2_2008.pdf#page=101)

Roldán, J., & Fera, M. (s. f.). Sentence splitting in Arabic to Spanish translation. *Revista Española de Lingüística Aplicada*.

Roldán Verdejo, J. J. (s. f.). *La adaptación sintáctica en la traducción del árabe al español. Compilación y análisis de un corpus paralelo de textos de Naciones Unidas*. Universidad de Granada.

Sáez Guerrero, M. (2009). *Diseño de un sistema de extracción de información de artículos de Wikipedia* [Proyecto Fin de Carrera Ingeniería de Telecomunicación, Universidad Carlos III de Madrid]. [http://e-archivo.uc3m.es/bitstream/handle/10016/5874/PFC\\_Miguel\\_Saez\\_Guerrero.pdf?sequence=1](http://e-archivo.uc3m.es/bitstream/handle/10016/5874/PFC_Miguel_Saez_Guerrero.pdf?sequence=1)

Sainz-Quinn, C., & Fera, M. (2020). Translating Arabic named entities into English and Spanish: translation consistency at the United Nations. En S. Hanna, H. El-Farahaty, & A.-W. Khalifa (Eds.), *Routledge Handbook of Arabic Translation* (1.<sup>a</sup> ed., pp. 381-396). Routledge Taylor & Francis Group.

Salhi, H. (2010). Small Parallel Corpora in an English-Arabic Translation Classroom: No Need to Reinvent the Wheel in the Era of Globalization. En S. M. Shiyab, M. G. Rose, J. House, & J. Duval (Eds.), *Globalisation and Aspects of Translation* (pp. 53-67). Cambridge Scholars Publishing.

Salhi, H. (2013). Investigating the Complementary Polysemy and the Arabic Translations of the Noun Destruction in EAPCOUNT. *Meta: Journal des traducteurs*, 58(1), 227-246. <https://doi.org/10.7202/1023818ar>

Samy, D. (2005). Named Entities : Structure and Translation. A study based on a Parallel Corpus (Arabic-English-Spanish). *Proceedings from the Corpus Linguistics Conference Series*. <http://www.lllf.uam.es/~doaa/Publications/NamedEntities->

CorpusLinguistics05.pdf

Samy, D., & González-Ledesma, A. (2008). Pragmatic Annotation of Discourse Markers in a Multilingual Parallel Corpus (Arabic- Spanish-English). *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*. <http://www.lrec-conf.org/proceedings/lrec2008/>

Samy, D., Moreno Sandoval, A., & Guirao, J. M. (2005). A Proposal for an Arabic Named Entity Tagger Leveraging a Parallel Corpus. *Proceedings of the International Conference Recent Advances in Natural Language Processing (RANLP'05)*, 459-465. [http://elvira.llf.uam.es/ESP/Publicaciones/ArabicNE\\_RANLP.pdf](http://elvira.llf.uam.es/ESP/Publicaciones/ArabicNE_RANLP.pdf)

Samy, D., Moreno Sandoval, A., & Guirao, J. M. (2004). An Alignment Experiment of a Spanish-Arabic Parallel Corpus. *Proceedings of the International Conference on Arabic Resources and Tools (NEMLAR'04)*, 85-89. <http://elvira.llf.uam.es/ESP/Publicaciones/AlignmentPaper04.pdf>

Samy, D., Moreno Sandoval, A., Guirao, J. M., & Alfonseca, E. (2006). Building a Parallel Multilingual Corpus (Arabic-Spanish-English). *Proceedings of the 5th International Conference on Language Resources and Evaluations (LREC'06)*. <http://www.llf.uam.es/ING/Publicaciones/MultiCorpusPaperMod.pdf>

Sarairoh, M. A. T. (2001). Inconsistency in technical terminology: A problem for standardization in Arabic. *Babel*, 47(1), 10-21. <https://doi.org/10.1075/babel.47.1.03sar>

Sección de Traducción al Español - Paginilla. (s. f.-a). *El árabe en la traducción al español de las Naciones Unidas*. Recuperado 16 de noviembre de 2018, de <https://ls-sts.unog.ch/basic-page/el-arabe-en-la-traducion-al-espanol-de-las-naciones-unidas>

Sección de Traducción al Español - Paginilla. (s. f.-b). *Generalidades del EPU*. Recuperado 23 de octubre de 2018, de <https://ls-sts.unog.ch/committee/examen-periodico-universal>

Sección de Traducción al Español - Paginilla. (s. f.-c). *Guía de la Sección*. Recuperado 15 de noviembre de 2018, de <https://ls-sts.unog.ch/basic-page/guia-de-la-seccion>

- Sección de Traducción al Español de las Naciones Unidas - Paginilla. (2015). *Orientaciones para la traducción de textos jurídicos*. [https://ls-sts.unog.ch/sites/default/files/2020-03/Traducción de textos jurídicos - Orientaciones 2015.pdf](https://ls-sts.unog.ch/sites/default/files/2020-03/Traducción%20de%20textos%20jur%C3%ADdicos%20-%20Orientaciones%202015.pdf)
- Sekine, S. (2004). Named Entity: History and Future. *Project notes, New York University*, 4. [https://ccc.inaoep.mx/~villasen/index\\_archivos/cursosTATII/EntidadesNombradas/Sekine- NEsHistory04.pdf](https://ccc.inaoep.mx/~villasen/index_archivos/cursosTATII/EntidadesNombradas/Sekine-NEsHistory04.pdf)
- Sekine, S., & Nobata, C. (2004). Definition, dictionaries and tagger for Extended Named Entity Hierarchy. *LREC*, 1977-1980. [https://pdfs.semanticscholar.org/b243/4644b7178a01f97235a75bddd87b614313af.pdf?\\_ga=1.48684064.702852063.1484739534](https://pdfs.semanticscholar.org/b243/4644b7178a01f97235a75bddd87b614313af.pdf?_ga=1.48684064.702852063.1484739534)
- Servicio de Traducción al Español de las Naciones Unidas en Nueva York. (s. f.). *Manual del traductor del Servicio de Traducción al Español de las Naciones Unidas*. Recuperado 15 de noviembre de 2018, de <https://onutraduccion.wordpress.com/manual/>
- Servicio de Traducción al Español de las Naciones Unidas en Nueva York. (2015). *Nuevo portal UNTERM*. <https://onutraduccion.wordpress.com/2015/06/09/nuevo-portal-unterm/>
- Shaalán, K. (2014). A Survey of Arabic Named Entity Recognition and Classification. *Computational Linguistics*, 40(2), 469-510. [https://doi.org/10.1162/COLI\\_a\\_00178](https://doi.org/10.1162/COLI_a_00178)
- Shaalán, K., & Raza, H. (2009). NERA: Named Entity Recognition for Arabic. *Journal of the American Society for Information Science and Technology*, 60(8), 1652-1663. <https://doi.org/10.1002/asi.21090>
- Shahrour, A., Khalifa, S., Taji, D., & Habash, N. (2016). CamelParser: A system for Arabic syntactic analysis and morphological disambiguation. *Proceedings of COLING 2016 - 26th International Conference on Computational Linguistics: System Demonstrations*, 228-232. <https://www.aclweb.org/anthology/C16-2048.pdf>
- Tabory, M. (1980). *Multilingualism in International Law and Institutions*. Sijthoff &

Noordhoff.

Tafalla Plana, M. (2010). El multilingüismo en la Organización de las Naciones Unidas. *Revista de Llengua i Dret*, 53, 137-162. <http://revistes.eapc.gencat.cat/index.php/rld/article/viewFile/2319/2810>

Tang, B., Feng, Y., Wang, X., Wu, Y., Zhang, Y., Jiang, M., Wang, J., & Xu, H. (2015). A comparison of conditional random fields and structured support vector machines for chemical entity recognition in biomedical literature. *Journal of Cheminformatics*, 7(Suppl 1), 4-9. <https://doi.org/10.1186/1758-2946-7-S1-S8>

Trabelsi, F. B. F., Zribi, C. B. O., & Kouki, W. (2015). Combined Classification for Extracting Named Entities from Arabic Texts. *Proceedings of the 2015 First International Conference on Arabic Computational Linguistics (ACLing'15)*, 55-60. <https://doi.org/10.1109/ACLing.2015.15>

Turmo, J., Ageno, A., & Català, N. (2006). Adaptive information extraction. *ACM Computing Surveys*, 38(2). <https://doi.org/10.1145/1132956/1132957>

United Nations Coordination Committee of Special Procedures. (2008). *Manual of Operations of the Special Procedures of the Human Rights Council*. [https://www.ohchr.org/Documents/HRBodies/SP/Manual\\_Operations2008.pdf](https://www.ohchr.org/Documents/HRBodies/SP/Manual_Operations2008.pdf)

Vasconcellos, M., Avey, B., Gdaniec, C., Gerber, L., León, M., & Mitamura, T. (2001). Terminology and Machine Translation. En S. Wright & G. Budin (Eds.), *Handbook for Terminology Management* (Vol. 2, pp. 697-723). John Benjamins Publishing Company.

Vathy-Fogarassy, A., Székér, S., Szolár, B., & Fogarassy, G. (2020). The Efficiency of Different Distance Metrics for Keyword-Based Search in Medical Documents: A Short Case Study. *Studies in health technology and informatics*, 271, 232-239. <https://doi.org/10.3233/SHTI200101>

Villán Durán, C. (2016). *Manual sobre el sistema universal de protección de los derechos humanos* (p. 183). [http://www.upeace.org/uploads/file/publication/MANUAL\\_DDHH\\_DIGITAL-2016.pdf](http://www.upeace.org/uploads/file/publication/MANUAL_DDHH_DIGITAL-2016.pdf)

- Villán Durán, C., & Faleh Pérez, C. (2016). *Manual de Derecho Internacional de los Derechos Humanos*. Ubijus.
- Wade, M. L. (2003). *Consistency of translation techniques in the tabernacle accounts of Exodus in the Old Greek*. Brill.
- Wagner, E., Bech, S., & M Martínez, J. (2002). *Translating for the European Union Institutions* (p. 162). Routledge Taylor & Francis Group.
- Wagner, E., Bech, S., & M Martínez, J. (2012). *Translating for the European Union Institutions* (2.<sup>a</sup> ed.). St. Jerome Publishing.
- Wang, Y., Qin, J., & Wang, W. (2017). Efficient Approximate Entity Matching Using Jaro-Winkler Distance. En A. Bouguettaya, Y. Gao, A. Klimenko, L. Chen, X. Zhang, F. Dzerzhinskiy, W. Jia, S. V Klimenko, & Q. Li (Eds.), *Web Information Systems Engineering -- WISE 2017* (pp. 231-239). Springer International Publishing. <https://jqin.gitee.io/files/wise2017-wang.pdf>
- Wroblewski, S. (2014). *eLUNa at a glance (ICTS/OUSG/DGACM)*. United Nations-DGACM. [https://cms.unov.org/meta/Media/Default/Files/eLuna Quick Reference Guide v1.0 11 Feb 2014\\_edited.pdf](https://cms.unov.org/meta/Media/Default/Files/eLuna Quick Reference Guide v1.0 11 Feb 2014_edited.pdf)
- Xiao, T., Zhu, J., Yao, S., & Zhang, H. (2011). Document-level Consistency Verification in Machine Translation. *Machine Translation Summit XIII, 13*, 131-138. <http://www.mt-archive.info/MTS-2011-Xiao.pdf>
- Xu, J., Fraser, A., & Weischedel, R. (2001). TREC 2001 Cross-lingual Retrieval at BBN. *Tenth Text REtrieval Conference (TREC'01)*, 68-77. <http://trec.nist.gov/pubs/trec10/papers/BBNTREC2001.pdf>
- Ziemski, M., Junczys-Dowmunt, M., & Pouliquen, B. (2016). The United Nations Parallel Corpus v1.0. *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC'16)*, 3530-3534.
- Zitouni, I., & Benajiba, Y. (2014). Aligned-parallel-corpora based semi-supervised learning for arabic mention detection. *IEEE Transactions on Audio, Speech and Language Processing*, 22(2), 314-324.

<https://doi.org/10.1109/TASLP.2013.2287055>

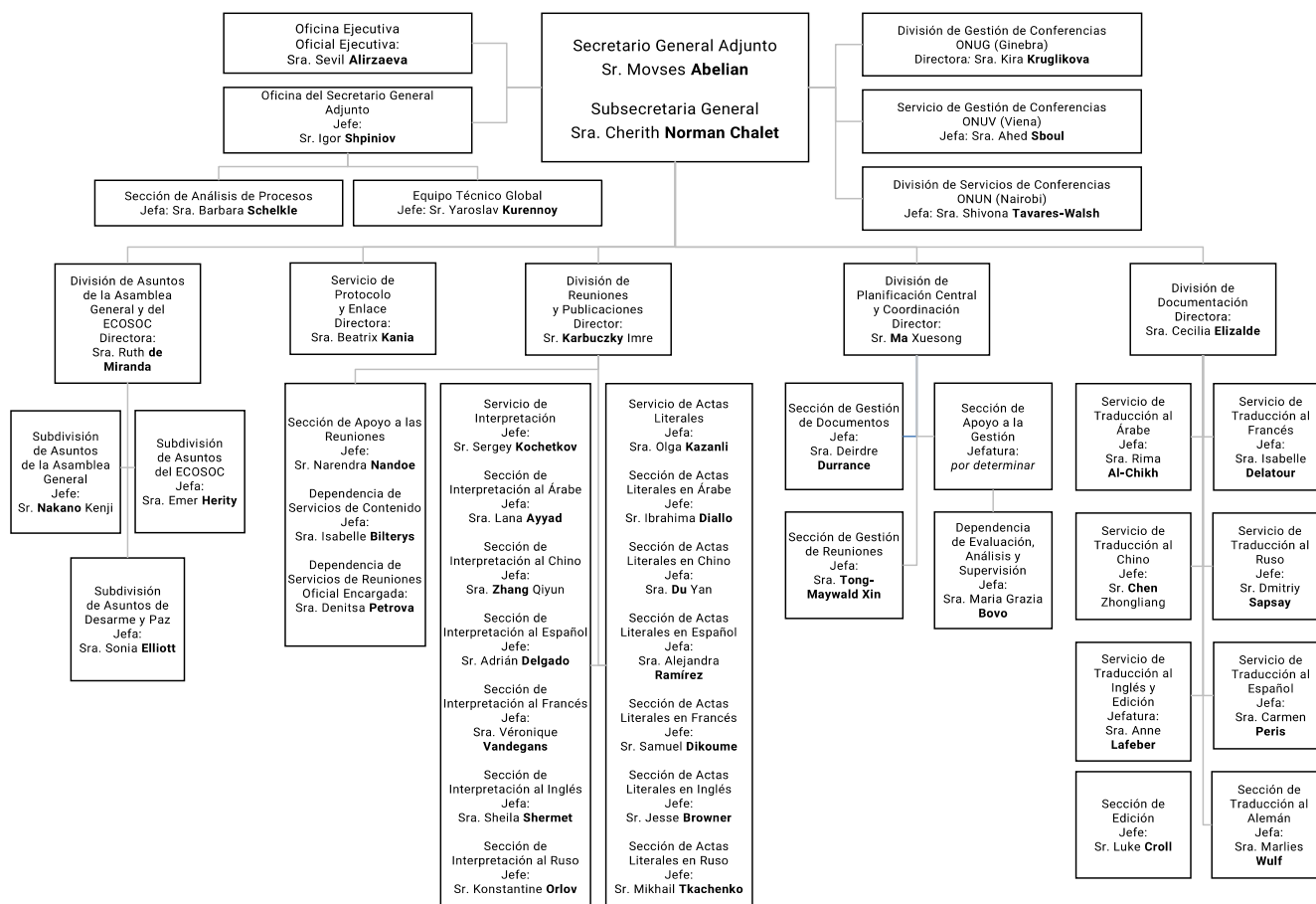
## **Anexos**







## 2. Anexo 2: Organigrama del Departamento de la Asamblea General y de Gestión de Conferencias



Actualizado el 1 de octubre de 2021/Laurence Hallé

Ilustración 5: Organigrama del Departamento de la Asamblea General y de Gestión de Conferencias

Fuente: <https://www.un.org/dgacm/sites/www.un.org.dgacm/files/chart-sp-dgacm.pdf> [Última consulta: 2 de octubre de 2021]

### 3. Anexo 3. Ejemplo de la portada de un informe periódico de Estado

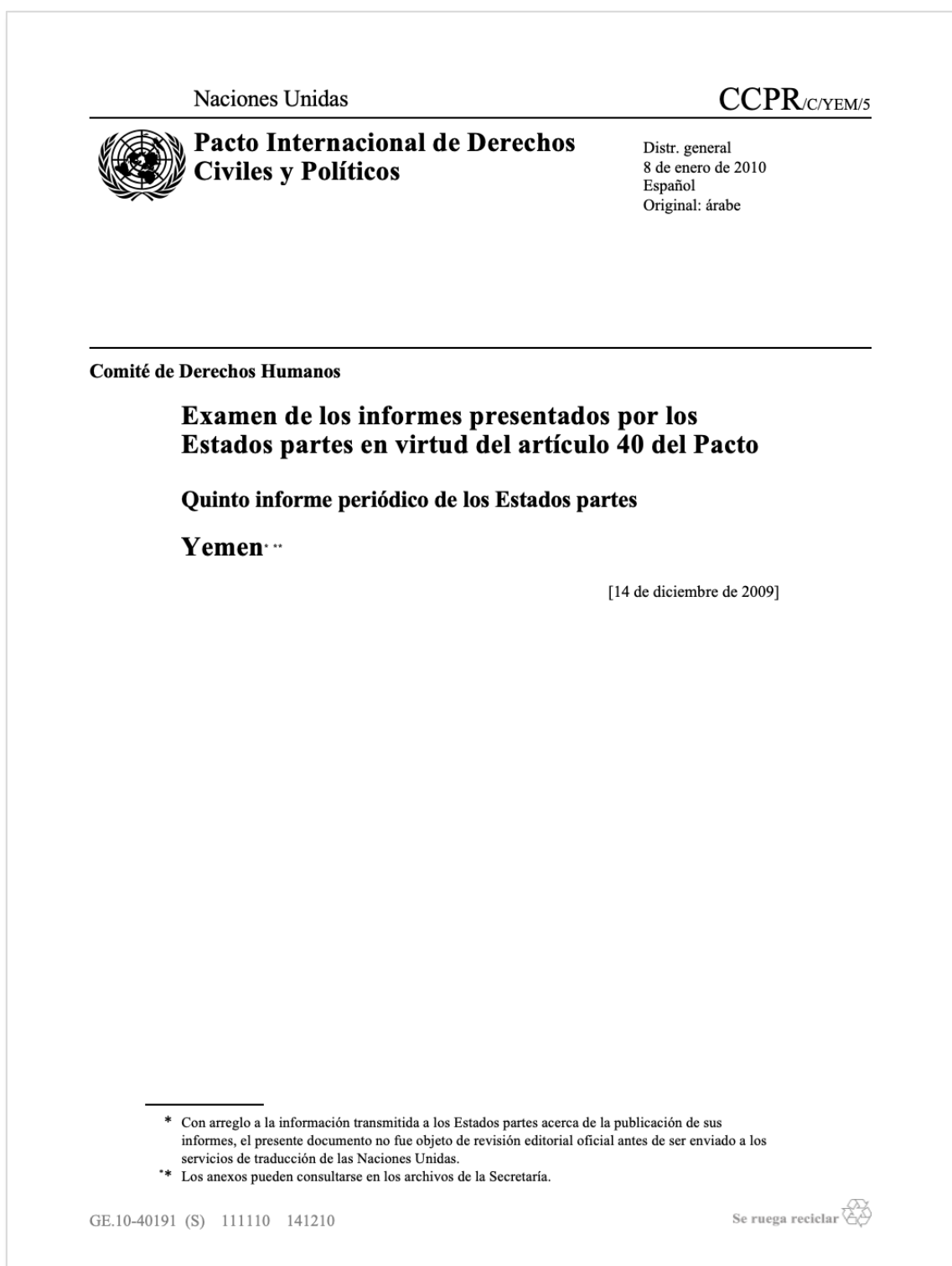


Ilustración 6: Ejemplo de la portada de un informe periódico del Estado (CCPR/C/YEM/5) en su versión en español


Fuente: <https://undocs.org/es/ccpr/c/yem/5> [Última consulta: 24 de abril de 2019]



United Nations

A/HRC/WG.6/32/YEM/1

---

 **General Assembly**

Distr.: General  
8 November 2018  
English  
Original: Arabic

---


**Human Rights Council**  
**Working Group on the Universal Periodic Review**  
**Thirty-second session**  
1 January–1 February 2019

**National report submitted in accordance with paragraph 15  
(a) of the annex to Human Rights Council resolution 5/1\***

**Yemen**

\* The present document has been reproduced as received. Its content does not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations.

GE.18-18894 (E) 271118 281118

  
\* 1 8 1 8 8 9 4 \*



Please recycle  

Ilustración 7: Ejemplo de la portada de un informe del EPU (A/HRC/WG.6/32/YEM/1 en su versión en inglés

Fuente: <https://documents-dds-ny.un.org/doc/UNDOC/GEN/G18/335/77/PDF/G1833577.pdf?OpenElement> [Última consulta: 09 de octubre de 2021]



#### 4. Anexo 4. Tratados internacionales y comités de expertos

Tratado internacional	Año de aprobación	Comité de expertos independiente
Pacto Internacional de Derechos Civiles y Políticos y sus protocolos facultativos	1966	Comité de Derechos Humanos (CCPR)
Pacto Internacional de Derechos Económicos, Sociales y Culturales	1966	Comité de Derechos Económicos, Sociales y Culturales (CESCR)
Convención Internacional sobre la Eliminación de Todas las Formas de Discriminación Racial	1965	Comité para la Eliminación de la Discriminación Racial (CERD)
Convención sobre la Eliminación de Todas las Formas de Discriminación contra la Mujer y su protocolo facultativo	1979 Protocolo facultativo: 1999	Comité para la Eliminación de la Discriminación contra la Mujer (CEDAW)
Convención contra la Tortura y Otros Tratos o Penas Crueles, Inhumanos o Degradantes	1984	Comité contra la Tortura (CAT)
Convención sobre los Derechos del Niño y sus protocolos facultativos	1989 Protocolo facultativo: 2000	Comité de los Derechos del Niño (CRC)
Convención Internacional sobre la Protección de los Derechos de Todos los Trabajadores Migratorios y de Sus Familiares	1990	Comité para la Protección de los Derechos de todos los Trabajadores Migratorios y Sus Familiares (CMW)
Convención Internacional sobre los Derechos de las Personas con Discapacidad	2006	Comité de los Derechos de las Personas con Discapacidad (CRPD)
Convención Internacional para la Protección de Todas las Personas contra las Desapariciones Forzadas	2006	Comité contra las Desapariciones Forzadas (CED)
Establecido de conformidad con las disposiciones del Protocolo Facultativo de la Convención contra la Tortura (OPCAT)	2002	Subcomité para la Prevención de la Tortura y Otros Tratos o Penas Crueles, Inhumanos o Degradantes (SPT)

Tabla 20: Relación completa de los tratados internacionales, y los respectivos años de aprobación y comités de expertos

Fuente: elaboración propia, basada en [www.ohchr.org/SP/HRBodies/Pages/HumanRightsBodies.aspx](http://www.ohchr.org/SP/HRBodies/Pages/HumanRightsBodies.aspx) [Última consulta: 24 de junio de 2018]