

# LA TRADUCCIÓN E INTERPRETACIÓN DE NOMBRES DE ORGANIZACIONES EN EL EUROLECTO INGLÉS Y ESPAÑOL: UN ESTUDIO BASADO EN CORPUS

Fernando Sánchez Rodas

Universidad de Málaga, Instituto Universitario de Investigación de Tecnologías Lingüísticas Multilingües, Edificio de Institutos Universitarios, C/ Severo Ochoa, 4, Parque Tecnológico de Andalucía, 29071, Málaga, España  
fersanchez@uma.es

## Translation and interpreting of organisation names in the English and Spanish Eurolect: a corpus-based study

**Abstract:** Multi-word named entities (MWNEs) are crucial bearers of information and often provide answers to major questions concerning the understanding of texts; however, work focusing exclusively on MWNEs is of very recent origin and comes mostly from a single field of study, Natural Language Processing (NLP). This paper aims to perform a contrastive analysis of a set of English and Spanish multi-word organisation names (MWORGs) extracted from two bilingual comparable corpora (written and oral) created from texts and speeches of the European Parliament. The methodology will follow a data-driven Constructionist approach. First, Named Entity Recognition (NER) and manual recall are employed to retrieve the MWORGs, and their performance is evaluated through metrics. Then, four items of metadata are annotated for each entity. The analysis unveils that, at this onomastic level, the non-translated English Eurolect in fact contains a considerable number of translations (17.3%), while non-interpreted English has an even higher percentage of interpretations (24.3%). In the translated and interpreted Spanish Eurolect, mediating schemes (translations, interpretations, borrowings, and blending) are predictably dominant (80.5% and 98.1% respectively), although they also coexist with native onomastic constructions. The results of the study boost the commonly disregarded potential of proper names as a key tool for understanding translation and interpreting as relative categories in which each (non-)translated and (non-)interpreted textual genre is ruled by its own set of constructional rules, establishing different quantities of mediating and non-mediating constructions depending on factors such as the language, the norms, and the events which condition textual production.

**Keywords:** named entity; European Parliament; translation; interpreting; mediating schemes

**Resumen:** Las entidades nombradas multipalabra (ENMP) constituyen importantes nodos de información lingüística que, en numerosas ocasiones, ofrecen claves fundamentales para la comprensión de los textos en los que aparecen. Sin embargo, los estudios dedicados en exclusiva a este tipo de unidades son de muy reciente cuño y provienen en su gran mayoría de un único campo de estudio, el Procesamiento del Lenguaje Natural (PLN). El siguiente trabajo tiene como objetivo realizar un análisis contrastivo de una serie de nombres de organizaciones multipalabra (ORGMP) extraídos de dos corpus comparables inglés-español (uno escrito y otro oral) de material del Parlamento Europeo. Se sigue para ello una metodología de corte construccionista. Primero, empleamos técnicas de reconocimiento automático de entidades nombradas y recuperación manual para extraer los nombres propios, y evaluamos el rendimiento de ambas tareas mediante métricas. Después, se anotan cuatro parámetros para cada entrada. El análisis desvela que, a este nivel onomástico, el eurolecto inglés no traducido contiene en realidad un número significativo de traducciones (17,3 %) y el inglés no interpretado cuenta con un porcentaje aún mayor de interpretaciones (24,3 %). En el eurolecto español traducido (80,5 %) e interpretado (98,1 %) dominan las traducciones, interpretaciones, préstamos y fusiones de ORGMP, aunque estos esquemas mediadores también conviven con construcciones del onomástico nativo. Los resultados de este estudio manifiestan el potencial desapercibido de los nombres propios como puerta de acceso al entendimiento de que la traducción e interpretación no son categorías absolutas, sino que cada género textual (no) traducido o (no) interpretado se rige por una gramática construccionista distinta que reparte cantidades dispares de construcciones mediadoras y no mediadoras en función de parámetros como el idioma, las normas y eventos que rodean la producción textual.

**Palabras clave:** entidad nombrada; Parlamento Europeo; traducción; interpretación; esquemas mediadores

## 1. Introducción<sup>1</sup>

Las expresiones multipalabra (EMP<sup>2</sup>) pueden ser definidas de manera amplia como «interpretaciones lingüísticas idiosincráticas que van más allá de los límites marcados por una palabra o espacio» (Jackendoff 1997: 156).<sup>3</sup> Algunas características compartidas por la mayoría de EMP son la variación interlingüística, la paráfrasis monoléxica, la proverbialidad y la prosodia (Baldwin y Nam Kim 2010). Las tareas más importantes con EMP que aspira a ejecutar el Procesamiento del Lenguaje Natural (PLN) son dos: la identificación y extracción de EMP a partir de corpus mediante la desambiguación de las reglas sintácticas internas y la consecuente interpretación de dichas expresiones. La práctica y evaluación de las tareas mencionadas han impulsado paulatinamente la celebración de congresos (Corpas Pastor *et al.* 2016) y la publicación de volúmenes especializados (Mitkov *et al.* 2018), además de un refinamiento

<sup>1</sup> Este artículo es una versión ampliada de una conferencia publicada en las actas de EUROPHRAS (Sánchez Rodas 2022a). La presente ampliación añade un componente oral al corpus bilingüe y expande los contenidos teóricos y metodológicos, a la vez que presenta un análisis inédito del reciclaje de nombres traducidos en los textos interpretados.

<sup>2</sup> Se emplean los siguientes acrónimos especializados, por orden alfabético: EMP (Expresión MultiPalabra), ENMP (Entidad Nombrada MultiPalabra), ORGML (ORGanización MonoLéxica), ORGMP (ORGanización MultiPalabra) y PLN (Procesamiento del Lenguaje Natural).

<sup>3</sup> «Idiosyncratic interpretations that cross word boundaries (or spaces)» (traducción propia).

en la ejecución de las tareas mediante la aplicación de técnicas de *parsing* y alineación de palabras, entre otras (Hatami, Mitkov y Corpas Pastor 2021).

Las EMP sustantivas son uno de los subtipos más generalizados por su frecuencia de *tokens*, *types* y aparición en prácticamente todas las lenguas del mundo (Lieber y Štekauer 2009). Como su nombre indica, una EMP sustantiva está formada por un núcleo sintagmático sustantivo o sustantivado que incluye uno o más modificadores («vehículo de tracción a las cuatro ruedas»). Las formas y funciones concretas de las EMP varían dependiendo de la lengua o familia de lenguas. Por ejemplo, en inglés son frecuentes las EMP completamente sustantivas (*golf club*, *computer science department*), pero también encontramos EMP cuyo núcleo es un verbo sustantivado (*investor hesitation*, *stress avoidance*) o un sustantivo acompañado de modificadores adjetivales o verbales (*open secret*, *connecting flight*). En lenguas romances como el español, por el contrario, existe un subtipo de EMP sustantivas formadas por una preposición o contracción situada entre el núcleo y un sustantivo modificador («zumo de naranja», «vistas al mar»).

Para Sag y otros (2002: 5-6), las entidades nombradas<sup>4</sup> multipalabra (ENMP) presentan frecuentemente rasgos propios de EMP sustantivas con un grado de fijación medio y una sintaxis idiosincrática. Los autores ilustran esta afirmación con ejemplos de nombres de clubes deportivos en inglés estadounidense. Estos nombres se suelen formar a partir de un núcleo constituido por un topónimo o nombre de organización (que a su vez puede estar formado por más de una palabra, como *San Francisco*) y un modificador que ubica al club en el deporte correspondiente (*San Francisco 49ers*, *Oakland Raiders*). Este tipo de construcciones onomásticas genera obstáculos importantes para el PLN, ya que presentan fenómenos idiosincráticos de elisión (*the [San Francisco] 49ers*) y selección y coordinación de determinantes (*an/the [[Oakland] Raiders] player*, *the Raiders and 49ers*), que suelen combinarse con lexicalizaciones alternativas (*the San Francisco 49ers*, *those San Francisco 49ers*, *San Francisco 49ers*, *the 49ers*, *those 49ers*, *49ers*), modificaciones adicionales (*the league-leading [San Francisco] 49ers*) y casos de sobreexplotación lingüística (*the Oakland 49ers*).

No es coincidencia que, de entre la gran variedad de clases onomásticas descritas en la bibliografía especializada (Zgusta 1998), Sag *et al.* (2002) elijan los nombres de organizaciones como categoría prototípica de las ENMP. Knappová (2017) indica que los nombres de organizaciones multipalabra (ORGMP) están formadas por un núcleo sustantivo de carácter apelativo o no onímico («Banco») y una serie de modificadores de diversa complejidad («Banco Central Europeo», «Banco Europeo de Inversiones», etc.). Los modificadores no solo sirven para distinguir la institución de otra similar, sino que aportan a las ORGMP una significación social y de prestigio y, por ende, un importante valor enciclopédico (Viimaranta y Mustajoki 2020).

Debido a la dependencia de las ORGMP de las condiciones jurídicas, económicas, comerciales y culturales que conforman una sociedad determinada, resulta lógico que los nombres de organizaciones sean una de las categorías onomásticas más recurrentes en el discurso de la Unión Europea o eurolecto (Goffin 1994), tal y como se

<sup>4</sup> Término alternativo empleado en el PLN para referirse a los nombres propios (véase Spossito *et al.* 2021).

ha comprobado de manera empírica en estudios anteriores (Steinberger *et al.* 2011; Corpas Pastor y Sánchez Rodas 2022; Corpas Pastor y Sánchez Rodas, en prensa). Como bien es sabido, el colectivo de los traductores e intérpretes es uno de los que más contribuyen a la construcción de la sociedad de sociedades que es la Unión Europea (Ortega Arjonilla *et al.* 2018), por lo que tiene que enfrentarse a diario a la mediación de estos nombres de organizaciones. Desde la perspectiva semiótica de Stecconi (2009: 262-263), la mediación, junto con la similitud y la diferencia, es una de las condiciones indispensables de la traducción; «es a todas luces imposible calificar de traducción [o interpretación] un texto en el que no se perciba que alguien habla en representación de otro, es decir, que no medie entre un contexto origen y otro meta».<sup>5</sup> Desde nuestro punto de vista, al menos en el discurso de la UE, la mediación es mucho más que un transvase interlingüístico a manos de un experto; constituye una función pragmática autónoma destinada a la resolución de conflictos sociales o nacionales mediante la lengua, función que traductores, intérpretes y otras figuras de la Unión ejercen consciente o inconscientemente en su trabajo diario, en el marco de unas normas y eventos de traducción e interpretación específicos.

El presente artículo pretende contribuir al estudio de las ENMP en general y las ORGMP en particular con un análisis del ejercicio de la mediación a través de los nombres de organizaciones en el eurolecto inglés y español. Partimos de la hipótesis de que la mediación, como acto pragmático de elevada ilocución y perlocución, se manifiesta en las construcciones lingüísticas a un nivel de abstracción relativamente elevado. Por esa razón, el primer objetivo es la extracción en masa de ORGMP de dos corpus bilingües comparables de textos del Parlamento Europeo, uno escrito y otro oral, mediante técnicas PLN avanzadas (reconocimiento de entidades nombradas). Una vez obtenidos estos datos, el segundo objetivo es la identificación de denominadores comunes y/o posibles diferencias en la mediación de los nombres extraídos. Para ello se ejecuta un análisis contrastivo semiconstruccional, en el sentido de que no disecciona con detalle los emparejamientos forma-función de las construcciones onomásticas, sino que emplea esquemas básicos para la caracterización de la mediación como un fenómeno poliédrico, que implica FUSIÓN, TRADUCCIÓN, INTERPRETACIÓN, PRÉSTAMO y combinaciones más complejas de estos cuatro, como TRADUCCIÓN + (PRÉSTAMO).

## 2. Metodología

Esta sección es la más extensa del artículo, ya que el método empleado para la extracción y el estudio bilingüe de nombres propios aprovecha al máximo las posibilidades de las nuevas tecnologías multilingües al respecto. En primer lugar, se describe el corpus elegido para la extracción de datos y las motivaciones de dicha selección (2.1.); posteriormente, se explica cómo se han reconocido las organizaciones multi-palabra de entre el conjunto de candidatos obtenido y se calculan métricas especializadas de precisión, recuperación y puntuación F-1 que evalúan el rendimiento de

<sup>5</sup> «[...] it is logically impossible to label as translation a text that is not perceived as speaking on behalf of another - i.e. that does not mediate between source and target environments» (traducción propia).

nuestro método (2.2.); para cerrar la sección, se detalla el proceso de anotación de metadatos de los nombres propios objeto de estudio, lo que sirve para tender puentes entre la metodología y el análisis (2.3.).

## 2.1. Selección de corpus

Como se ha explicado en la sección anterior, el conjunto de datos del que parte este trabajo fue extraído de un corpus comparable inglés-español de textos procedentes de la Comisión de Peticiones del Parlamento Europeo. Estos textos son a su vez parte de PETIMOD, un corpus intermodal más grande con subcorpus de todos los tipos de mediación lingüística practicados en la Comisión PETI: traducciones, no-traducciones, interpretaciones y no-interpretaciones. Hasta donde alcanza nuestro conocimiento, PETIMOD es el único corpus compuesto en exclusiva por material escrito y oral de una comisión del Parlamento Europeo. Aunque las dos primeras versiones del corpus eran paralelas (Corpas Pastor y Sánchez Rodas 2022; Copras Pastor y Sánchez Rodas, en prensa), en la fase actual (PETIMOD 3.0) el corpus es comparable, ya que la adición indiscriminada de muestras ha revelado que la distribución de varios géneros textuales es asimétrica (no todas las muestras están traducidas o interpretadas del inglés al español, y viceversa). La Figura 1 muestra de manera jerárquica la arquitectura de PETIMOD y los subcorpus empleados para el presente estudio (marcados por un círculo). A continuación, las Tablas 1 y 2 diseccionan las estadísticas básicas de estos cuatro subcorpus que, una vez enfrentados, forman dos corpus bilingües comparables: PETIMOD 3.0 N-T\_EN/T\_ES (No Traducido\_inglés/Traducido\_ESpañol) y PETIMOD 3.0 N-I\_EN/I\_EN (No Interpretado\_inglés/Interpretado\_ESpañol).

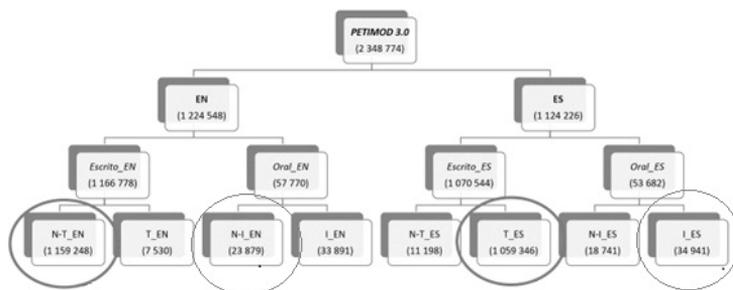


Figura 1. Arquitectura y tamaño de PETIMOD (en tokens). Los subcorpus marcados con círculos forman los dos corpus bilingües comparables empleados en este estudio

PETIMOD 3.0 N-T_EN/T_ES			
Subcorpus	N.º documentos	Types	Tokens
N-T_EN	880	29 273	1 159 248
T_ES	716	31 528	1 059 346
<b>TOTAL</b>	<b>1596</b>	<b>60 801</b>	<b>2 218 594</b>

Tabla 1. Estadísticas básicas del corpus comparable PETIMOD 3.0 N-T\_EN/T\_ES

PETIMOD 3.0 N-I_EN/I_ES			
Subcorpus	N.º documentos	Types	Tokens
N-I_EN	155	2931	23 879
I_ES	206	4415	34 941
<b>TOTAL</b>	<b>361</b>	<b>7346</b>	<b>58 820</b>

Tabla 2. Estadísticas básicas del corpus comparable PETIMOD 3.0 N-I\_EN/I\_ES

## 2.2. Reconocimiento de las ORGMP

El reconocimiento de ORGMP es un proceso complejo que requiere trabajar en dos fases: primero se deben dividir los resultados brutos (candidatos) ofrecidos por el sistema según sean nombres organizaciones monoléxicas (ORGML) o multipalabra, y después medir la precisión y recuperación propiamente dichas. Esto es así porque la mayoría de los sistemas de reconocimiento actuales no permiten filtrar los resultados por número de palabras, tan solo por el etiquetado semántico (PER, ORG, etc.).

### 2.2.1. Extracción de candidatos

De manera resumida, la primera fase (extracción de candidatos ORGMP) se llevó a cabo en los siguientes pasos:

1. subida de los dos corpus comparables a la plataforma de gestión de corpus VIP (Corpas Pastor 2021);
2. reconocimiento de entidades nombradas masivo con la librería DeepPavlov, integrada en VIP<sup>6</sup>;
3. exportación en archivo Excel de los resultados de la categoría ORG;
4. filtrado de los candidatos por número de palabras mediante Kutools<sup>7</sup>.

Este método, aun simple, demostró ser efectivo para la identificación de expresiones complejas, ya que el 75,15 % del total de candidatos ORG extraídos de todos los corpus (968 de 1288) resultó ser multipalabra. La Tabla 3 detalla los resultados parciales y totales de la extracción de candidatos ORG en todos los subcorpus.

	N-T_EN	T_ES	N-I_EN	I_ES	TOTAL
Candidatos ORGML	147	133	22	18	320
Candidatos ORGMP	475	420	37	36	968
<b>Candidatos ORG</b>	<b>622</b>	<b>553</b>	<b>59</b>	<b>54</b>	<b>1288</b>

Tabla 3. Candidatos ORGML y ORGMP extraídos de los dos corpus comparables

<sup>6</sup> Aunque VIP también permite el reconocimiento de entidades nombradas basado en la librería SpaCy, para este estudio decidimos optar directamente por DeepPavlov, debido a que publicaciones anteriores similares habían demostrado un rendimiento ligeramente superior de esta librería (Sánchez Rodas 2022b; Corpas Pastor y Sánchez Rodas, en prensa).

<sup>7</sup> Kutools es un conjunto de herramientas adicionales para Excel con un periodo de prueba gratuito de 30 días (<<https://www.extendoffice.com/download/kutools-for-excel.html>>). El conjunto incluye un asistente de funciones que permite introducir de manera automática una fórmula compleja para el cómputo de palabras por celda. La fórmula deduce el número total de palabras mediante el conteo de espacios, un método que en algunos casos requiere de revisiones manuales para evitar posibles fallos.

### 2.2.2. Métricas de precisión, recuperación y F-1

Tras extraer una lista considerable de candidatos, los múltiples condicionantes del objeto de estudio elegido (entidades nombradas multipalabra que además debían pertenecer a una categoría semántica concreta) hacían necesario seguir depurando los resultados. La medición de la precisión debía ser muy exacta, ya que debía excluir dos tipos de errores del reconocimiento automático: EMP que no fueran nombres propios (*Emergency Assistance*) y EMP que sí fueran nombres propios, pero no de organizaciones (*Habitats Directive*). Con objeto de enriquecer nuestro conjunto de datos y de respetar la idiosincrasia sintáctica de este tipo de unidades, se corrigieron y consideraron relevantes las ORGMP con fallos de conversión menores y reconocidas junto con determinantes (*a Court of Justice*), así como los préstamos, fundamentales para el estudio de los contactos interlingüísticos a nivel onomástico (*Universität Frankfurt*, «la Agenzia regionale per la protezione ambientale»).<sup>8</sup> Por el contrario, se descartaron los genitivos discontinuos en inglés (*European Parliament's*), las construcciones preposicionales y plurales inacabadas (*National Contact Points for [Transport]*, «las Carreras Judicial [y Fiscal]») y una serie de nombres propios de interpretación ambigua que, tras las respectivas consultas manuales en el corpus, resultaron ser reconocimientos parciales de entidades más complejas («la Confederación Comarcal de Organizaciones» como parte de «la Confederación Comarcal de Organizaciones Empresariales de Cartagena»). Los reconocimientos parciales, que fuera de contexto pueden ser interpretados como una organización distinta, ilustran a la perfección la naturaleza constructiva e imbricada de las ORGMP.

En cuanto a la recuperación de las posibles ORGMP no detectadas de manera automática, se hizo con la ayuda de listas de palabras frecuentes. Más concretamente, generamos en VIP cuatro listas de los cien sustantivos más frecuentes en cada uno de los corpus. La recuperación de las entidades completas se llevó a cabo mediante la búsqueda individual de las concordancias de cada sustantivo frecuente, en las que se aplicaron filtros básicos como la distinción de mayúsculas, la búsqueda lematizada o cambios de orden para una visualización más precisa de los resultados. Si bien es cierto que este método es ágil y los resultados de recuperación obtenidos son aceptables (véase Tabla 4 debajo), no está exento de riesgos. En las listas de frecuencia generadas con VIP encontramos sustantivos de distinta calidad para nuestro fin; presuponemos que algunos son núcleos sintagmáticos más productivos en la construcción de nombres institucionales, ya sea en su forma flexiva o lematizada (*authorities*, «Comisión»), mientras que otros son términos de construcción monoléxica no onímica (*petitioner*, «miembros») o sustantivos de elevada banalización (*time*, «lugar»). Además, el empleo de una lista de palabras

<sup>8</sup> Las erratas aceptadas incluyen fallos ortográficos y ortotipográficos que no dificultan la comprensión y que fueron generados de forma accidental durante la conversión a texto plano de los archivos del corpus. La mayoría son fallos relacionados con acentuación («las Policías del Estado»), paréntesis (*the WHO()*), conversión de caracteres individuales («Comisión Europea», «Cuerpos de Seguridad del Eeiado»), notas al pie (*the Solidarity Transport Hub2*) y espaciado («la Comisión dePeticiones»).

extraídas del mismo corpus sometido con anterioridad a reconocimiento de entidades nombradas puede generar solapamientos entre el método automático y el manual. La Figura 2 ilustra un ejemplo de recuperación de ORGMP mediante VIP en el corpus de inglés no traducido. A continuación, la Tabla 4 recoge las métricas de precisión recuperación, y F-1 (puntuación combinatoria) de todos los corpus, junto con el número final de ORGMP identificadas una vez completado todo el proceso.

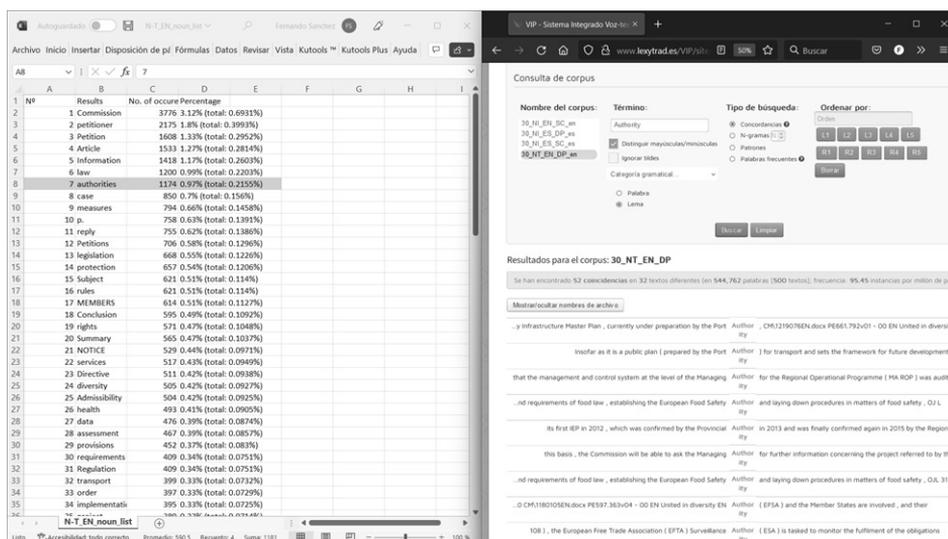


Figura 2. Ejemplo de recuperación de ORGMP en VIP mediante lista de sustantivos y búsqueda de concordancias

Subcorpus	Precisión	Recuperación	Puntuación F-1	ORGMP relevantes
N-T_EN	0,40	0,79	0,53	243
T_ES	0,43	0,63	0,51	287
N-I_EN	0,65	0,65	0,65	37
I_ES	0,69	0,49	0,57	51

Tabla 4. Métricas de recuperación, precisión y F-1 para los cuatro subcorpus de PETIMOD

Los resultados de la Tabla 4 muestran que las puntuaciones F-1 fueron por lo general aceptables para los cuatro corpus, sin rebasar en ninguno de los casos el umbral F-1 0,70. El F-1 resultó ligeramente más elevado para los corpus orales (+0,12 en inglés y +0,06 en español); sin embargo, las métricas de precisión se situaron por debajo de la media en el caso de los dos corpus más grandes (0,40 para N-T\_EN y 0,43 para T\_ES). Son números significativamente bajos si tenemos en cuenta dos factores: 1) entre las opciones técnicas disponibles se escogió expresamente la librería DeepPavlov por haber demostrado anteriormente un mejor rendimiento con las entidades ORG; 2) se establecieron una serie de criterios *ad hoc* para considerar relevantes el mayor número posible de candidatos ORGMP reconocidos de manera

automática. Las métricas de precisión ponen de manifiesto que, al menos en el sistema VIP/DeepPavlov, todavía es posible mejorar el reconocimiento de entidades nombradas con condicionantes (ENMP con etiquetas semánticas específicas), ya que solo en uno de los corpus (I\_ES) el valor de precisión supera al de recuperación (+0,20). Las métricas de recuperación, si bien más satisfactorias, también son mejorables, ya que su rendimiento es dispar (p. ej. N-T\_EN obtiene +0,30 en relación con I\_ES).

### 2.3. Anotación de las ORGMP

Con el propósito de aportar agilidad y solidez al análisis, la última fase metodológica consistió en una anotación semiautomática en Excel de algunos metadatos de las ORGMP. Recontamos cuántas palabras había en cada entidad mediante Kutools (véase subsección 2.2.1.) y añadimos tres metadatos más, que se detallan a continuación.

- *Subcategoría organizacional*: de acuerdo con la descripción que ofrece VIP de la etiqueta ORG, los nombres de organizaciones pueden referirse a «corporaciones» o empresas, (ORG<sub>CO</sub>), «agencias» o entidades públicas/sin ánimo de lucro de rango menor (ORG<sub>AGCY</sub>) e «instituciones» o entidades públicas/sin ánimo de lucro de rango mayor (ORG<sub>INST</sub>).<sup>9</sup>
- *Lenguas*: si la entidad es multilingüe, se anotan las lenguas por estricto orden sintáctico. Por ejemplo, en la entrada «Greenalia Biomass Power Curtis-Teixeiro S.LU» se anota INGLÉS + GALLEGO + ESPAÑOL.
- *Esquemas de mediación*: en numerosos nombres de organizaciones, se observan y anotan patrones pragmáticos de mediación. Para definirlos, recurrimos a terminología clásica de los estudios de traducción («traducción», «interpretación», «préstamo»), pero también de la lingüística cognitiva («fusión»<sup>10</sup>). Mientras que los tres primeros se explican por sí mismos, las fusiones aúnan palabras nativas y extranjeras con funciones distintas dentro de una misma construcción («Bundestag alemán»). Estas cuatro categorías básicas se combinan entre ellas para formar esquemas más complejos, normalmente señalizados por paréntesis o comas que forman parte de la propia construcción. También existen construcciones no mediadas, como instituciones y compañías nacionales de larga tradición (*National Health Service*, «Comisión Nacional del Mercado de Valores de España»), o en el caso de los textos no traducidos ni interpretados, entidades supranacionales de relevancia fosilizadas por motivos de estandarización, traducción acuñada, etc. (*the International Civil Aviation Organisation, European Economic and Social Committee*).

La Figura 3 muestra parte de la anotación de las entidades del corpus de español no traducido, mientras que la Tabla 5 lista todos los esquemas de mediación y no mediación observados junto con ejemplos adicionales de cada corpus.

<sup>9</sup> <[http://www.lexytrad.es/VIP/site/paq\\_complementario/entities\\_text.php](http://www.lexytrad.es/VIP/site/paq_complementario/entities_text.php)>.

<sup>10</sup> Traducción libre de *blending* (Fauconnier y Turner 2002).

	N-T_EN	T_ES	N-I_EN	I_ES
<b>ESQUEMAS MEDIADORES</b>				
FUSIÓN	the 'Petón do Lobo' Environmental Association	AG de Frankfurt	LIBE Committee	Environmental Glasnost
INTERPRETACIÓN	-	-	Catalan Supreme Court	Banco de Inversiones Europeo
PRÉSTAMO	Sociedad de Agricultores de la Vega	Bulgarian Helsinki Committee	Sauvons nos Enfants Japon	-
TRADUCCIÓN	Basque Department of Health	Parlamento de Baja Sajonia	-	-
FUSIÓN + (PRÉSTAMO)	-	Conselleria de Agricultura, Desarrollo Rural, Emergencia Climática y Transición Ecológica (Generalitat Valenciana)	-	-
PRÉSTAMO + PRÉSTAMO	Air France-KLM	Pannon GSM	-	-
PRÉSTAMO + (TRADUCCIÓN)	Associazione Autisti Soccorritori Italiani (Association of Italian Road Rescue)	Comitato volontario dei cittadini contro la discarica nell'ex cava Viti (Comité volontario de ciudadanos contra el vertedero en la antigua cantera Viti)	-	-
TRADUCCIÓN + (PRÉSTAMO)	Netherlands Institute for Human Rights (College voor de Rechten van de Mens)	banco del Estado de Luxemburgo (BCEE)	-	-
<b>ESQUEMAS NO MEDIADORES</b>				
NO TRADUCIDOS	British Airlines	Tribunal Superior de Justicia de Andalucía	-	-
NO INTERPRETADOS	-	-	Disability Rights International	Cruz Roja

Tabla 5. Tipología de esquemas mediadores y no mediadores en las ORGMP<sup>11</sup>

<sup>11</sup> Las denominaciones cortas de las comisiones del Parlamento Europeo se consideran **fusiones** porque el acrónimo proviene del nombre oficial en francés. Por ejemplo, PECH es la *Commission de la*

	MWORG	WORD COUNT	SUBCATEGORY	LANGUAGE(S)	MEDIATING SCHEME
1					
2	Wogtlaendlicher Bauernmarkt e. V.»	6	ORG-AGCY	German	Borrowing
3	Academia Europea de las Ciencias	5	ORG-INST	Spanish	Translation
4	Administración Pública española	3	ORG-INST	Spanish	Non-translated
5	Afinsa Bienes Tangibles S.A	4	ORG-CO	Spanish	Non-translated
6	AG de Frankfurt	3	ORG-CO	German + Spanish	Blending
7	Agencia Búlgara de Seguridad Alimentaria	3	ORG-AGCY	Spanish	Translation
8	Agencia Comunicación y Género S.L.	5	ORG-CO	Spanish	Non-translated
9	Agencia de los Derechos Fundamentales de la Unión Europea	9	ORG-AGCY	Spanish	Translation
10	Agencia Estatal de Resolución d'Entitats Bancàries	6	ORG-AGCY	Catalan	Borrowing
11	Agencia Europea de Seguridad Aérea	5	ORG-AGCY	Spanish	Translation
12	Agencia Europea de Seguridad Aérea	5	ORG-AGCY	Spanish	Translation
13	Agencia Europea de Seguridad Marítima	5	ORG-AGCY	Spanish	Translation
14	Agencia irlandesa para la Protección del Medio Ambiente	8	ORG-AGCY	Spanish	Translation
15	Agrupación Nacional de Víctimas de la Intervención de Fórum y Afinsa	11	ORG-AGCY	Spanish	Non-translated
16	AIOP – Associação de Industriais do Concelho de Pombal	9	ORG-AGCY	Portuguese	Borrowing
17	al Comité de las Regiones	5	ORG-INST	Spanish	Translation
18	al Parlamento Europeo	3	ORG-INST	Spanish	Translation
19	Animal Welfare Foundation	3	ORG-AGCY	English	Borrowing
20	Armadora Pareros S. L.	6	ORG-CO	Spanish	Non-translated
21	ARPA Friuli Venezia Giulia	4	ORG-AGCY	Italian	Borrowing
22	Asociación Española de Protección Civil para el Clima Espacial	9	ORG-AGCY	Spanish	Non-translated
23	Asociación Europea de Libre Comercio	5	ORG-AGCY	Spanish	Translation
24	Associação Sindical dos Juízes Portugueses	5	ORG-AGCY	Portuguese	Borrowing
25	Astarcom Telecomunicaciones	2	ORG-CO	Spanish	Non-translated
26	Audienca Provincial	2	ORG-INST	Spanish	Non-translated
27	Audienca Provincial (Landgericht)	3	ORG-INST	Spanish + German	Translation + (Borrowing)
28	Audienca Territorial (Oberlandesgericht)	3	ORG-INST	Spanish + German	Translation + (Borrowing)
29	Autoridad de Protección de Datos Personales Italiana	7	ORG-INST	Spanish	Translation
30	Autoridad de Seguridad de la Aviación Civil de Australia	9	ORG-INST	Spanish	Translation
31	Autoridad Europea de Seguridad Alimentaria	5	ORG-INST	Spanish	Translation
32	Autoridad Europea de Valores y Mercados	6	ORG-INST	Spanish	Translation
33	Autoridad Portuguesa	2	ORG-INST	Spanish	Non-translated
34	Ayuntamiento de Real	3	ORG-INST	Spanish	Non-translated
35	Banco Central Europeo	3	ORG-INST	Spanish	Non-translated
36	Banco Central portugués	3	ORG-INST	Spanish	Translation
37	banco del Estado de Luxemburgo (BCEE)	6	ORG-INST	Spanish + French	Translation + (Borrowing)

Figura 3. Ejemplo de anotación de ORGMP (corpus PETIMOD 3.0 T\_ES)

### 3. Análisis y resultados

Este apartado de análisis cualitativo-cuantitativo está dividido en dos subsecciones, una por cada uno de los corpus comparables inglés-español analizados (escrito y oral).

#### 3.1. Esquemas mediadores y no mediadores en N-T\_EN/T\_ES

	N-I_EN	I_ES
<b>ESQUEMAS MEDIADORES</b>	<b>83</b>	<b>231</b>
FUSIÓN	7	18
INTERPRETACIÓN	0	0
PRÉSTAMO	28	36
TRADUCCIÓN	42	163
FUSIÓN + (PRÉSTAMO)	0	1
PRÉSTAMO + PRÉSTAMO	1	4
PRÉSTAMO + (TRADUCCIÓN)	3	1
TRADUCCIÓN + (PRÉSTAMO)	2	8
<b>ESQUEMAS NO MEDIADORES</b>	<b>160</b>	<b>56</b>
NO TRADUCIDOS	160	56
NO INTERPRETADOS	0	0
<b>TOTAL</b>	<b>243</b>	<b>287</b>

Tabla 6. Esquemas mediadores y no mediadores en las ORGMP de N-T\_EN/T\_ES

Como cabría esperar, los resultados del primer corpus comparable demuestran que los esquemas onomásticos mediadores, que incluyen –pero no se limitan a– traducciones,

pêche y AFCO es la *Commission des affaires constitutionnelles*. Puede consultarse la lista completa en <[https://fr.wikipedia.org/wiki/Commissions\\_du\\_Parlement\\_europ%C3%A9en](https://fr.wikipedia.org/wiki/Commissions_du_Parlement_europ%C3%A9en)>.

son más numerosos en el español traducido (80,5 %) que en el inglés no traducido (34,2 %). Sin embargo, no deja de resultar sorprendente que el corpus no traducido incluya, de hecho, bastantes TRADUCCIONES (17,3 %). La mayoría son versiones inglesas de organizaciones nacionales extranjeras, de las que en numerosas ocasiones se ha podido registrar que emplean exactamente la misma traducción en su página web o documentación oficial (*French Agency for Food, Environmental and Occupational Health & Safety*)<sup>12</sup>. El segundo esquema mediador más presente en ambos corpus son los PRÉ - TAMOS (11,5 % en N-T\_EN y 12,5 % en T\_ES), procedentes de lenguas como el portugués (*Tribunal Central Administrativo Sul*) o el alemán (*Radlberger Getränkegesellschaft*). El esquema PRÉSTAMO + PRÉSTAMO, por el contrario, es bastante infrecuente; la mayoría de los ejemplos los conforman nombres de empresas en al menos una lengua minoritaria, como neerlandés (*Air France-KLM*) o húngaro (*Pannon GSM*).

Si comparamos de manera más detenida algunos de los esquemas mediadores restantes de este corpus, podemos extraer conclusiones interesantes sobre las distintas maneras en las que las dos lenguas y culturas ejercen la función pragmática de la mediación en el contexto de la UE. La categoría PRÉSTAMO + (TRADUCCIÓN), por ejemplo, funciona ligeramente mejor en N-T\_EN (3) que en T\_ES (1). Este resultado podría significar que, al menos en el discurso escrito de la Unión, el inglés prefiere comenzar la construcción onomástica con el elemento extranjero y aportarle una relevancia temática superior, como en *Audiencia Provincial (Court of Appeal)*. El español traducido prefiere el esquema inverso TRADUCCIÓN + (PRÉSTAMO), incluso cuando la lengua extranjera también es romance, como en «Ministerio de Medio Ambiente, Protección del Territorio y Medio Marino (Ministero dell' Ambiente e della Tutela del Territorio e del Mare)». Otra diferencia reseñable es la ausencia del esquema FUSIÓN + (PRÉSTAMO) en inglés, que en español traducido sí cuenta con un ejemplo: «la Conselleria de Agricultura, Desarrollo Rural, Emergencia Climática y Transición Ecológica (Generalitat Valenciana)». La primera entidad es una fusión de español y valenciano (nótese el encaje del sustantivo *Conselleria*), mientras que la segunda entidad entre paréntesis está en completo valenciano. Aunque en este esquema concreto los números son demasiado bajos para extraer conclusiones sólidas, se podría deducir que el eurolecto español busca compensar una cierta falta de diplomacia, evidenciada en la preferencia del esquema TRADUCCIÓN + (PRÉSTAMO), mediante la diversificación o creación de un esquema onomástico de mediación que, curiosamente, está orientado a la resolución de un conflicto nacional (la cooficialidad lingüística del Estado español) y no supranacional ni internacional.

### 3.2. Esquemas mediadores y no mediadores en N-I\_EN/I\_ES

El primer dato que llama poderosamente la atención en contraste con los corpus escritos es la completa ausencia de mediaciones complejas como FUSIÓN + (PRÉSTAMO) o PRÉSTAMO + PRÉSTAMO. En el caso de los esquemas con paréntesis, resulta lógica la

<sup>12</sup> Traducción oficial del organismo francés ANSES (*Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail*) que podemos encontrar en la versión inglesa de su propia página web (<<https://www.anses.fr/en/content/our-identity>>).

exclusión, ya que, como mencionamos más arriba, consideramos que estos símbolos gráficos forman parte íntegra de las construcciones onomásticas escritas en las que aparecen; en la oralidad no es posible su representación, quedando sustituidos por otros fenómenos como pausas del habla, elipsis, finales de frase, etc. Sin embargo, en el caso de PRÉSTAMO + PRÉSTAMO, se podría argumentar que la falta de complejidad es una estrategia económica que busca ahorrar esfuerzos cognitivos ante la instantaneidad de la producción oral (corpus no interpretado) o la actividad combinada de producción oral más transvase a otra lengua (corpus interpretado). En segundo lugar, el porcentaje de mediaciones parcialmente insertas en el discurso no mediado se mantiene inusitadamente alto, hasta un 40,5 % en N-I\_EN (+ 6,3 % en relación con N-T\_EN). En el corpus interpretado, los esquemas no mediados son directamente mínimos (1,9 %). La INTERPRETACIÓN es el esquema de mediación favorito en ambos corpus (24,3 % en N-I\_EN y 88,2 % en I\_ES), seguido de la FUSIÓN y el PRÉSTAMO.

	N-I_EN	I_ES
<b>ESQUEMAS MEDIADORES</b>	<b>15</b>	<b>50</b>
FUSIÓN	4	5
INTERPRETACIÓN	9	45
PRÉSTAMO	2	0
TRADUCCIÓN	0	0
FUSIÓN + (PRÉSTAMO)	0	0
PRÉSTAMO + PRÉSTAMO	0	0
PRÉSTAMO + (TRADUCCIÓN)	0	0
TRADUCCIÓN + (PRÉSTAMO)	0	0
<b>ESQUEMAS NO MEDIADORES</b>	<b>22</b>	<b>1</b>
NO TRADUCIDOS	0	0
NO INTERPRETADOS	22	1
<b>TOTAL</b>	<b>37</b>	<b>51</b>

Tabla 7. Esquemas mediadores y no mediadores en las ORGMP de N-I\_EN/I\_ES

Un aspecto muy relevante observado en numerosas INTERPRETACIONES tanto de N-I\_EN como I\_ES, y que merecería ser objeto de estudio aparte, es el descubrimiento de que existe un continuo de diferencia/similitud respecto a traducciones escritas de los mismos nombres de organizaciones. Por un lado, tenemos una serie de ORGMP que sí se pueden identificar más claramente como interpretaciones, ya que su construcción parece muy influenciada por la espontaneidad oral o la labor de interpretación. Dicha influencia se manifiesta en cambios de orden sintáctico («Banco de Inversiones Europeo» en lugar de «Banco Europeo de Inversiones»), acortamientos («Agencia de los Medicamentos» en lugar de «Agencia Europea de Medicamentos») o construcciones *ad hoc* que no se asemejan a ninguna traducción previa («el Congreso de la Dieta japonesa», «Oficina del Parlamento Europeo para los Niños»), entre otros fenómenos. Por otro lado, existe todo un abanico de construcciones onomásticas en las que la barrera entre interpretación y traducción es más difusa, y que presentan distintos

grados de acercamiento o fidelidad a traducciones escritas rastreables en fuentes de mayor o menor oficialidad. Por ejemplo, en N-I\_EN encontramos *The Family Court of Turkey*, una ORGMP de la que no se ha podido rastrear ninguna traducción oficial al inglés, aunque sí construcciones idénticas en prensa<sup>13</sup> y traducciones de páginas web de abogacía.<sup>14</sup> En el mismo corpus tenemos *The High Court of Tokyo*, una construcción muy parecida que sí cuenta con la misma forma estandarizada en la prestigiosa base de datos jurídica JustCite<sup>15</sup>. El nivel de similitud con las traducciones puede ser aún más elevado, ya que encontramos ejemplos que parecen directamente extraídos de las bases de datos terminológicas de la Unión Europea, como «coordinador del Parlamento Europeo para los derechos del niño»<sup>16</sup> en I\_ES, e incluso nombres reciclados de la misma petición escrita, como *Environmental Inspection Service*, entidad que ya aparecía en N-T\_EN. En definitiva, aunque sean necesarios más datos empíricos de los subcorpus restantes de PETIMOD e información adicional sobre el trabajo en la Comisión PETI, estos resultados parecen indicar que muchas de las ORGMP catalogadas con cautela en este estudio como interpretaciones podrían ser traducciones «recicladas» para construir el discurso oral. El nivel de contacto entre ambas modalidades de mediación podría ser mayor de lo esperado, quizás porque los ponentes de los discursos originales y/o los intérpretes de las comisiones combinan la preparación escrita previa con la argumentación oral espontánea.

#### 4. Conclusiones

Este estudio ha filtrado dos corpus comparables inglés-español de eurolecto escrito (2 218 594 *tokens*) y oral (58 820 *tokens*) en busca de nombres de organización multipalabra (ORGMP), mediante la funcionalidad de reconocimiento de entidades nombradas del sistema VIP. Por un lado, las puntuaciones F-1 de esta tarea están por encima de la media, pero en ningún caso superan el umbral del 0,70, siendo la más alta F-1 0,65 (corpus de inglés no interpretado). Dichas puntuaciones demuestran que el reconocimiento de entidades multipalabra de categorías específicas como ORG aún es una tarea compleja, tanto a nivel automático como manual, y que se debe avanzar en el entrenamiento de sistemas y el desarrollo de técnicas de recuperación más eficientes. Por otro lado, el análisis de las 618 ORGMP relevantes extraídas de este modo ha desvelado la existencia de numerosos patrones mediadores (fusión, interpretación, préstamo y traducción, más algunas combinaciones complejas de estos cuatro) y no mediadores (esquemas no traducidos y no interpretados).

El hallazgo más importante es que, a este nivel onomástico, el eurolecto inglés no traducido contiene en realidad un número significativo de traducciones (17,3 %). El eurolecto inglés no interpretado cuenta con un porcentaje aún mayor de interpretaciones (24,3 %), siendo incluso posible que varias de ellas sean realmente reciclajes de traduc-

<sup>13</sup> <<https://timesofmalta.com/articles/view/judges-order-child-removed-from-turkey-stays-in-malta.216092>>.

<sup>14</sup> <<http://en.baykurthukuk.com.tr/aile-hukuku/>>.

<sup>15</sup> <<http://www.justcite.com/kb/data-coverage/standardised-court-names/>>.

<sup>16</sup> <<https://iate.europa.eu/entry/result/2243717/es-en>>.

ciones previas, ya que el grado de similitud con las ORGMP escritas es en muchos casos idéntico. Como cabría esperar, en el eurolecto español traducido e interpretado dominan los esquemas mediadores (80,5 % y 98,1 % respectivamente), ya sea en forma de traducciones, interpretaciones, préstamos, fusiones u otros. Sin embargo, se puede apreciar que la mediación tampoco es hegemónica, ya que, por ejemplo, el español traducido cuenta con hasta un 19,5 % de unidades que provienen del onomástico nativo.

Los números de este estudio ponen de relieve el potencial desapercibido de los nombres propios como puerta de acceso al estudio profundo de la mediación lingüística y, consecuentemente, al entendimiento de que la traducción y la interpretación no son etiquetas absolutas impuestas *ex machina*, sino de que cada género textual (no) traducido o (no) interpretado se rige por su propia gramática construccional, repartiendo un número determinado de construcciones mediadoras y no mediadoras en función de parámetros como el idioma, las normas y eventos que rodean a la producción textual, etc. Creemos que solo desde una visión parecida los investigadores serán capaces de seguir generando conjuntos de datos y análisis reutilizables para la mejora continua de aplicaciones PLN como los generadores de texto y resúmenes y, naturalmente, los sistemas de traducción e interpretación automáticos.

## Agradecimientos

El presente trabajo ha sido financiado por la beca FPU18/05803 del Ministerio de Universidades de España. También ha sido realizado (parcialmente) en el seno de varios proyectos de investigación relacionados con las tecnologías lingüísticas: VIP II (PID2020-112818GB-I00), TRIAJE (UMA18-FEDERJA-067) y las pruebas de concepto VIP PDC2021-121220-I00 y E3/04/21.

### Referencias bibliográficas

- BALDWIN, Timothy - NAM KIM, Su (2010) «Multiword Expressions», en INDURKHYA, N. - DAMERAU, F. J. (eds.), *Handbook of Natural Language Processing* (Second Edition), Boca Raton: CRC Press, 267-292.
- CORPAS PASTOR, Gloria (2021) «Technology Solutions for Interpreters: The VIP System», *Hermēneus. Revista de Traducción e Interpretación* 23, 91-123.
- CORPAS PASTOR, Gloria - SÁNCHEZ RODAS, Fernando (2022), «NLP-Enhanced Shift Analysis of Named Entities in an English<>Spanish Intermodal Corpus of European Petitions», en KAJZER-WIETRZNY, M. - BERNARDINI, S. - FERRARESI, A. - IVASKA, I. (eds.), *Empirical Investigations into the Forms of Mediated Discourse at the European Parliament*, Berlín: Language Science Press, 219-251.
- CORPAS PASTOR, Gloria - SÁNCHEZ RODAS, Fernando (en prensa), «EU Phraseological Verbal Patterns in the PETIMOD 2.0 Corpus: A NER-Enhanced Approach», en BIEL, L. - KOCKAERT, H. (eds.), *Handbook of Legal Terminology*, Ámsterdam: John Benjamins.
- CORPAS PASTOR, Gloria - MONTI, Johanna - SERETAN, Violeta - MITKOV, Ruslan (eds.) (2016), *Workshop Proceedings: Multi-Word Units in Machine Translation and Translation Technologies MUMTTT2015*, Ginebra: Tradulex.
- FAUCONNIER, Gilles, - TURNER, Mark (2002), *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*, Nueva York: Basic Books.

- GOFFIN, Roger (1994), «L' Eurolecte : Oui, Jargon Communautaire : Non», *Meta* 39 (4), 636-642.
- HATAMI, Ali – MITKOV, Ruslan – CORPAS PASTOR, Gloria (2021), «Cross-Lingual Named Entity Recognition via FastAlign: A Case Study», en *Proceedings of the Translation and Interpreting Technology Online Conference*, Shoumen: INCOMA Ltd, 85-92.
- JACKENDOFF, Ray (1997), *The Architecture of the Language Faculty*, Cambridge: MIT Press.
- KNAPPOVÁ, Miroslava (2017), «Chrématoymum», en *Nový Encyklopedický Slovník Čestiny Online* [disponible en <<https://www.czechency.org/slovník/CHRÉMATONYMUM>>, 06/11/2022].
- LIEBER, Rochelle – ŠTEKAUER, Pavol (eds.) (2009), *The Oxford Handbook of Compounding*, Oxford: Oxford University Press.
- MITKOV, Ruslan – MONTI, Johanna – CORPAS PASTOR, Gloria – SERETAN, Violeta (eds.) (2018), *Multiword Units in Machine Translation and Translation Technology*, Ámsterdam/Filadelfia: John Benjamins.
- ORTEGA ARJONILLA, Emilio – MARTÍNEZ LÓPEZ, Ana Belén – DELGADO PUGÉS, Iván (eds.) (2018), *La traducción e interpretación jurídicas en la Unión Europea: Retos para la Europa de los ciudadanos*, Granada: Comares.
- SAG, Ivan A. – BALDWIN, Timothy – BOND, Francis – COPESTAKE, Ann – FLICKINGER, Dan (2002), «Multiword Expressions: A Pain in the Neck for NLP», en *Proceedings of the Third International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2002)*, Ciudad de México, 1-15.
- SÁNCHEZ RODAS, Fernando (2022a), «A Corpus-Based Analysis of Mediation in EU Multi-Word Organization Names», en CORPAS PASTOR, G. – MITKOV, R. – KUNILOVSKAYA, M. – CARO QUINTANA, R. (eds.), *Proceedings of the International Conference EURO-PHRAS 2022*, Shoumen: INCOMA Ltd., 160-69.
- SÁNCHEZ RODAS, Fernando (2022b), «Introducing the PETIMOD Corpus: A Resource for the Analysis of Personification in Mediated and Non-Mediated Discourse», en ESTEVES-FERREIRA, J. – MITKOV, R. – RECORT RUIZ, M. – STEFANOV, O. (eds.), *Proceedings of 43rd Conference Translating and the Computer (TC43)*, Ginebra: Tradulex, 128-132.
- SPOSSITO, Osvaldo Mario – RYCKEBOER, Hugo Emilio – LEDESMA, Viviana – PROCOPIO, Gastón *et al.* (2021), «Propuesta Para La Construcción de Un Corpus Jurídico Utilizando Expresiones Regulares», en *XXVII Congreso Argentino de Ciencias de La Computación (CACIC) (Modalidad Virtual, 4 Al 8 de Octubre de 2021)*, 746-755.
- STECCONI, Ubaldo (2009), «Semiotics», en BAKER, M. – SALDANHA, G. (eds.), *Routledge Encyclopedia of Translation Studies*, Londres/Nueva York: Routledge, 260-263.
- STEINBERGER, Ralf – POULIQUEN, Bruno – KABADJOV, Mijail – BELYAEVA, Jenya – VAN DER GOOT, Erik (2011), «JRC-NAMES: A Freely Available, Highly Multilingual Named Entity Resource», en *International Conference Recent Advances in Natural Language Processing, RANLP*, 104-110.
- VIIMARANTA, Johanna – MUSTAJOKI, Arto (2020), «What Can Science, Religion, Politics, Culture and the Economy Do? A Corpus Study of Metonymical Conceptualization Combined with Personification», *Scando-Slavica* 66 (1), 71-85.
- ZGUSTA, Ladislav (1998), «The Terminology of Name Studies (In Margine of Adrian Room's Guide to the Language of Name Studies)», *Names: A Journal of Onomastics* 46 (3), 189-203. <https://doi.org/10.1179/nam.1998.46.3.189>.