

Citación para la versión publicada

Mikelenic, B. & Oliver, A. El complemento directo y de régimen en español y sus equivalencias en croata: un estudio basado en corpus. *Revista Española de Lingüística Aplicada (RESLA)*.

Versión del documento

Esta es una versión enviada del manuscrito.
La versión en el Repositorio O2 de la Universitat Oberta de Catalunya puede ser diferente de la versión final publicada.

Derecho de uso y reutilización

Esta versión del manuscrito se hace disponible con una licencia Creative Commons del tipo Atribución No Comercial No Derivadas (CC BY-NC-ND) <https://creativecommons.org/licenses/by-nc-nd/3.0/es/>, que permite descargarla y compartirla siempre que se cite su autoría, pero sin modificarla ni utilizarla con fines comerciales.

Consultas

Si cree que este documento infringe los derechos de autor, contacte con el equipo de investigación: repositori@uoc.edu



El complemento directo y de régimen en español y sus equivalencias en croata – un estudio basado en corpus

Bojana Mikelenić
Universidad de Zagreb
bmikelen@ffzg.hr

Antoni Oliver
Universitat Oberta de Catalunya
aoliverg@uoc.edu

Resumen

El presente estudio se basa en la teoría sobre la relación del complemento directo (CD) y de régimen (CR) en español –la posibilidad de su coocurrencia en el mismo predicado (Alarcos, 1966; Bosque, 1983; Rojo, 1983) y los predicados en los que estos dos complementos pueden alternar– para analizar las diferencias en la traducción de estos verbos y sus argumentos al croata. Mediante búsquedas en un corpus paralelo español-croata y la herramienta ReSiPC (*Regular Expression Search in Parallel Corpora*) desarrollados ambos para esta investigación, se muestran ejemplos de diferentes soluciones al traducir dichas estructuras –verbos diferentes, el mismo verbo con el régimen diferente, cambio del verbo dependiendo del significado del complemento–, al igual que ciertos paralelismos entre los esquemas de los dos idiomas. Con esta aproximación se pretende contribuir en el desarrollo de la metodología y herramientas que facilitarán otros trabajos de esta índole, subrayando también su valor descriptivo, dado el escaso número de trabajos que comparan estos dos idiomas.

Palabras clave: español, croata, traducción, complemento de régimen, complemento directo, corpus paralelo

1. Introducción

En este trabajo presentamos un estudio empírico basado en corpus sobre tres estructuras del predicado español: *verbo + complemento directo (CD)*, *verbo + complemento de régimen (CR)* y *verbo + CD + CR*. A las primeras dos estructuras las denominaremos *biactanciales* y elegiremos los verbos que se pueden utilizar con ambos complementos; la tercera estructura se denomina *triactancial* (José María García-Miguel, 1995). Nos centraremos en el estudio de las traducciones de estas estructuras españolas al croata mediante ejemplos de un corpus paralelo formado por cuatro novelas

en español y sus correspondientes traducciones al croata. Este corpus ha sido segmentado y alineado a nivel de oración. Asimismo, utilizaremos la herramienta ReSiPC (*Regular Expression Search in Parallel Corpora*), desarrollada específicamente para esta investigación. Para obtener las descripciones y los esquemas sintáctico-semánticos de los verbos españoles, hemos utilizado la base ADESSE (García-Miguel et al., 2005).

En cuanto a la organización del artículo, en el siguiente apartado presentaremos los fundamentos teóricos sobre los complementos verbales en ambos idiomas y posteriormente hablaremos de los principales conceptos relacionados con la lingüística de corpus, así como de las herramientas de explotación y consulta de corpus más utilizadas. Después describimos el proceso de compilación y tratamiento del corpus. Seguiremos con una descripción de la herramienta de consulta de corpus que hemos desarrollado específicamente para este proyecto y que se distribuye bajo una licencia libre. A continuación, figura la parte experimental donde ofreceremos una muestra de los resultados obtenidos para las estructuras de predicado mencionadas al principio. Por último, presentaremos las conclusiones y las líneas de trabajo futuras.

2. Los complementos verbales en español y croata

Desde que Emilio Alarcos (1966) propuso la existencia del complemento de régimen (CR), un complemento verbal independiente de los tres tradicionales (complemento directo: CD, indirecto: CI y circunstancial: CC), el tema de la relación y diferenciación de estos complementos ha sido estudiado por diversos lingüistas, algunos de los cuales mencionaremos a continuación. Tratando el asunto de la transitividad, Alarcos (1966, p. 17) concluye que “la transitividad es característica de la estructura del predicado: consiste en la presencia de un *implemento* o un *suplemento*.”¹ La aparente similitud de la relación con el verbo de cada uno de estos dos elementos era justamente

¹En dicho artículo Alarcos utiliza su propia terminología: *implemento* para el complemento directo, *complemento* para el indirecto, *aditamento* para el circunstancial y *suplemento* para el de régimen. En este trabajo utilizaremos la terminología tradicional para los tres complementos anteriormente conocidos y el nombre de *complemento de régimen* propuesto por María Luz Gutiérrez (1987) para el cuarto.

uno de los argumentos principales para afirmar la imposibilidad de su coocurrencia en el mismo predicado. A pesar de que esto fuera refutado por Ignacio Bosque (1983) y luego admitido por Alarcos (en Hortensia Martínez, 1986 y Alarcos, 1994), el vínculo (en primer lugar semántico) entre estos dos complementos sigue siendo un asunto que suscita interés. Esto se refleja también en la denominación de *objeto preposicional* para el CR y en la postulación de la *transitividad preposicional* (Rafael Cano, 1987; Gutiérrez, 1987). Gutiérrez (1987, p. 381) concluye que “la diferencia entre la transitividad directa y la preposicional reside en el plano formal, más que en el semántico.”

Por otro lado, García-Miguel (1995) tratando la transitividad dentro de la estructura sintagmática de la cláusula, utiliza el término *biactancial* para referirse a ambos esquemas, considerando, sin embargo, el esquema transitivo (con el CD) el no marcado. De esta manera el mismo autor acentúa la relación de la rección entre el verbo y su complemento, donde tanto el CD como el CR son elementos valenciales² o regidos, por lo que se refiere a “la identidad de valencia, no de estructura gramatical” en estas cláusulas biactanciales (García-Miguel, 1995, p. 22). La prueba de los diferentes valores gramaticales (y funcionales) que aporta cada complemento se puede encontrar en las cláusulas triactanciales, en las que los dos coocurren. García-Miguel (1995, p. 22) concluye que en las cláusulas de esta índole “resulta inequívoco que se trata de funciones sintácticas diferentes”, añadiendo que en estos casos de compatibilidad de dos complementos, sus funciones semánticas también son diferentes. En cuanto a la identidad de significado entre las dos estructuras, este autor asume una posición diferente de la de Cano, afirmando que la aportación del significado de la preposición indica una construcción diferente.³

²La gramática de valencias es una aproximación a la que apelan varios autores en el tratamiento del CR. García-Miguel menciona a Bosque (1983), Guillermo Rojo (1983/1990) y José Álvaro Porto Dapena (1987).

³Advertimos que el significado o la falta de significado de la preposición en estas construcciones es otro tema polémico, relacionado también con el marco teórico que uno adopta. Tanto García-Miguel (1995) como p. ej., José Antonio Candalija Reina (2008) consideran que es muy difícil comprobar la llamada *desemantización preposicional*, mientras que p. ej., Luis González García y Ana Veleiro (1991) o Ana Serradilla Castaño (1997-98) argumentan que la preposición en estas construcciones ha perdido su significado y solo sirve de enlace y marco del régimen verbal.

En croata tradicionalmente se distinguen los objetos (directo e indirecto) de los complementos adverbiales (~ circunstanciales). Eugenija Barić et al. (2005, p. 429) destacan que la diferencia entre estos dos grupos reside en que los complementos adverbiales pueden parafrasearse (usando preposiciones y casos diferentes) y los objetos no, es decir, su forma es fija. Josip Silić & Ivo Pranjković (2007, pp. 304-5), desarrollando más el tema, afirman que los objetos están más estrechamente conectados con el verbo, forman parte de la predicación y en general son obligatorios en la cláusula, a pesar de que existen excepciones. Por otro lado, los complementos adverbiales se introducen dependiendo del significado de la oración entera, designan circunstancias y por eso no participan en la predicación ya que son elementos generalmente facultativos. Los objetos se definen (Barić et al., 2005, p. 431) como palabras nominales en uno de los casos oblicuos (genitivo: *Gen.*, dativo: *Dat.*, acusativo: *Acc.*, locativo: *Loc.* o instrumental: *Inst.*), con o sin preposición. La forma del objeto depende de y está condicionada por la rección o régimen verbal. En Silić & Pranjković (2007, p. 264) se distingue entre *la rección fuerte* y *la rección débil*, donde la forma del elemento dependiente viene condicionada por las características léxico-gramaticales del verbo (u otro elemento principal) en el primer caso, o solamente del significado léxico del verbo en el segundo. Como se puede observar en las gramáticas del croata (Silić & Pranjković, 2007, pp. 300-1; Barić et al., 2005, p. 431; Radoslav Katičić, 2002, pp. 95, 109), el tipo de objeto está estrechamente conectado con su forma, es decir, su caso, de manera que el objeto directo adquiere la forma del acusativo sin preposición (u otros casos pero sustituible con acusativo) y el indirecto se presenta en el resto de los casos oblicuos, con o sin preposición. En las tres gramáticas mencionadas también figura la conexión entre la presencia del objeto directo y la transitividad. Puesto que el objeto directo es siempre un sustantivo, pronombre o sintagma nominal sin preposición, podemos concluir que en la tradición croata no se deja lugar para postular la transitividad preposicional.

Sin embargo, recientemente se ha empezado a analizar más profundamente la relación de régimen entre el verbo y la preposición (Daniela Katunar, 2015). Partiendo de la posición de las preposiciones “en la frontera entre el lexicón y la gramática, la sintaxis y la semántica”, Katunar (2015, p. 1) identifica las unidades que denomina *construcciones verbo-preposicionales*, en las que opera la rección fuerte, denominando

preposiciones regidas a las que aparecen en estas construcciones. La autora (Katunar, 2015, p. 6) arguye que la relación de régimen aquí es un reflejo de los procesos de idiomatización y gramaticalización, que son dos caras de la misma moneda y mediante los que se construye el significado de la preposición. Siguiendo el grado de influencia de los dos procesos sobre cada construcción verbo-preposicional (o presencia/ausencia de metaforización), Katunar (2015, pp. 147-150) propone la división entre las construcciones metafóricamente motivadas e idiomatizadas y las construcciones gramaticalizadas, en las que la preposición tiene un significado más esquemático (gramatical, funcional).⁴

Puesto que en este trabajo comparamos dos idiomas diferentes en cuanto a la riqueza de la flexión nominal, es importante notar la relación entre las preposiciones y los morfemas flexivos de caso. En su exposición sobre esta relación, Katunar (2015, p. 79) destaca que especialmente en los idiomas con tradición de interpretar las preposiciones como los *casos analíticos* (p. ej., inglés y francés), se subrayan las conexiones entre las funciones de los dos grupos de elementos. Abordando el tema del régimen en español, Eva Spitzová (1974, p. 51) recuerda que, debido a la falta de declinación de los sustantivos en español, las preposiciones asumieron muchas funciones que en los idiomas con declinación desempeñan los morfemas flexivos de caso. Daniela Katunar & Ida Raffaelli (2018, p. 56) destacan que en croata, al igual que en otras lenguas eslavas, las preposiciones rigen el caso del sustantivo que las acompaña y así cumplen con la función de demarcación de caso que poseen (en. *case-marking function*). Además de los sufijos flexivos, las autoras (Katunar & Raffaelli, 2018, p. 58) subrayan la importancia de los prefijos en croata, que se acercan funcionalmente a algunas preposiciones, principalmente, y como se puede adivinar, a las que se ven más gramaticalizadas. En la parte de análisis tendremos en cuenta las relaciones mencionadas y su papel en la traducción de diferentes estructuras argumentales del español al croata.

⁴Para la conexión con el tema del significado de la preposición en el CR en la lingüística española, v. nota a pie de página nº 3.

3. La lingüística de corpus

En este artículo se presentan los resultados de un estudio empírico basado en corpus y por este motivo en este apartado se introducen los aspectos más relevantes de la disciplina denominada lingüística de corpus. También comentamos las características principales de algunos de los programas informáticos para la explotación y búsqueda en corpus lingüísticos. Como veremos, ninguno de ellos se adapta plenamente a nuestras necesidades, por lo que hemos decidido desarrollar nuestra propia herramienta, que distribuimos bajo una licencia libre.

Un corpus es una colección de elementos lingüísticos seleccionados y ordenados de acuerdo con criterios lingüísticos explícitos con la finalidad de ser usado como muestra de la lengua (John Sinclair, 1996, p. 4). Dado que en la mayoría de los casos los corpus se analizan mediante herramientas informáticas, podemos definir un corpus automatizado (o informatizado) como aquel que se ha codificado de manera estándar y homogénea para distintas tareas de recuperación de información. El objetivo de la lingüística de corpus es obtener datos objetivos sobre las estructuras y las categorías (sintácticas, léxicas, morfológicas, etc.) de una lengua, si se trata de un corpus monolingüe; o de forma comparada entre lenguas, si se trata de un corpus bilingüe o multilingüe. Podemos distinguir también dos tipos de corpus bilingües (o multilingües): los corpus paralelos y los corpus comparables (Tony McEnery et al., 2006, p. 47). Un corpus paralelo es un corpus formado por textos en una lengua y sus correspondientes traducciones a una o más lenguas. Estos corpus a menudo se presentan alineados oración con oración; es decir, cada oración de los textos originales está relacionada con la correspondiente oración de los textos traducidos. No siempre esta relación entre oraciones es una a una y a menudo una oración original está relacionada con más de una oración traducida, o al revés, diversas oraciones originales están relacionadas con una oración traducida. Hemos hecho referencia a oraciones, aunque para ser más exactos sería mejor hablar de segmentos originales y traducidos, ya que no siempre hay una relación directa entre segmento y oración gramatical. Un corpus comparable, en cambio, es un corpus que contiene textos en una lengua y textos similares en otra u otras lenguas. Esta similitud puede venir dada por diversos criterios, como por ejemplo el

ámbito de especialidad de los textos. Además de los tipos de corpus presentados, cabe destacar también la diferenciación entre corpus sin anotar –donde los textos se presentan tal y como son sin ninguna información añadida– de los corpus anotados –donde los textos aparecen acompañados de algún tipo de información adicional–. En nuestro trabajo utilizamos los corpus anotados morfosintácticamente, donde cada palabra viene acompañada con información sobre su lema y su categoría gramatical. Las anotaciones pueden ser de índoles muy diversas: anotaciones semánticas sobre el sentido de las palabras, sobre opiniones y sentimientos, sobre la existencia de palabras o expresiones utilizadas metafóricamente, sobre marcadores del discurso, etc.

Los corpus lingüísticos bilingües y multilingües se han utilizado intensivamente en diversas técnicas de Procesamiento de Lenguaje Natural. Entre estas técnicas destacan la traducción automática, en la que se utilizan corpus paralelos para el entrenamiento de sistemas de traducción automática estadísticos (Philipp Koehn et al., 2007) y neuronales (Dzmitry Bahdanau et al., 2015). Recientemente, además, se están utilizando también corpus comparables para el entrenamiento de sistemas de traducción automática (Mikel Artetxe et al., 2018). Los corpus lingüísticos también se utilizan en las técnicas de extracción automática de terminología (María Teresa Pazienza et al., 2005). Estas técnicas pretenden descubrir los términos más relevantes presentes en una colección de textos (o corpus). En la extracción terminológica monolingüe, donde se pretenden descubrir términos en una lengua, se utilizan corpus monolingües. En cambio, en la extracción terminológica bilingüe o multilingüe, donde se pretenden descubrir términos en una lengua junto con sus equivalentes de traducción en una o más lenguas, se utilizan habitualmente corpus paralelos alineados (Els Lefever et al., 2009), aunque también se usan con cierto éxito corpus comparables (Pascale Fung, 1998), especialmente en los últimos años cuando se han empezado a utilizar *word embeddings* bilingües para esta tarea (Amir Hazem & Emmanuel Morin, 2017).

Pero la utilidad de la lingüística de corpus va más allá de los avances técnicos, y tal y como comenta Mona Baker (1993), esta disciplina está teniendo un gran impacto en los estudios de traducción, tanto desde el punto de vista teórico como de la enseñanza de la traducción. Además, los corpus paralelos en forma de memoria de traducción son

recursos muy valiosos para los traductores profesionales que ayudan a mejorar tanto la calidad del resultado como la eficacia de todo el proceso.

Existen diversas herramientas para realizar consultas a corpus lingüísticos. Un ejemplo de herramienta gratuita muy utilizada es AntConc,⁵ que permite realizar diversos tipos de búsquedas en corpus monolingües sin anotar. Entre estas búsquedas posibles destacan las concordancias, es decir, la búsqueda de ciertas palabras o expresiones junto a su contexto de aparición; cálculo de n-gramas, que son combinaciones de n palabras; o las colocaciones o conjuntos de dos o más palabras que acostumbran a aparecer juntas. Otra herramienta destacable es la IMS Open Corpus Workbench (CWB),⁶ que es una herramienta de código abierto que está diseñada específicamente para poder realizar consultas a corpus de gran tamaño que contengan anotaciones lingüísticas. Esta herramienta puede trabajar con corpus paralelos considerando los corpus en cada lengua como corpus diferentes y codificando la información sobre la alineación de segmentos como un atributo más. Otras herramientas, como TigerSearch⁷ o TGrep⁸ permiten realizar búsquedas en corpus analizados sintácticamente. Y por último, como herramienta muy completa para analizar y consultar corpus de un tamaño muy grande podemos citar SketchEngine,⁹ que además proporciona cerca de 500 corpus en más de 90 lenguas ya cargados en el sistema para su uso inmediato.

Aunque existen diversas herramientas de consulta de corpus, ninguna se adaptaba plenamente a nuestras necesidades: la posibilidad de realizar búsquedas complejas en corpus paralelos donde al menos una de las lenguas disponga de anotación morfosintáctica y que no precise de transformaciones complejas de los formatos de los corpus. Por este motivo, como explicaremos en el apartado 5, hemos diseñado e implementado nuestra propia herramienta.

5<http://www.laurenceanthony.net/software/antconc/>

6<http://cwb.sourceforge.net/>

7<http://www.ims.uni-stuttgart.de/forschung/ressourcen/werkzeuge/TIGERSearch/manual.html>

8<http://tedlab.mit.edu/~dr/Tgrep2/tgrep2.pdf>

9<https://www.sketchengine.eu/>

4. Constitución y procesamiento del corpus

Se ha constituido un corpus con los textos de cuatro novelas españolas y sus correspondientes traducciones al croata.¹⁰ Los corpus se han segmentado en oraciones y se han alineado a nivel de oración utilizando la herramienta LF-Aligner.¹¹ Este programa es una interfaz gráfica que facilita el uso de la herramienta de alineación Hunalign¹² (Dániel Varga et al., 2005). El resultado del proceso son los libros alineados a nivel de oración, es decir, archivos donde cada oración del texto español está relacionada con la oración correspondiente de la versión croata. En la Tabla 1 podemos ver la medida en segmentos y palabras de cada una de las obras, así como el total.

Tabla 1 Número de segmentos y palabras del corpus

Obra	Segmentos	Palabras español	Palabras croata
Marina	5.656	63.168	56.571
Ejército iluminado	2.702	60.361	52.658
El mapa del tiempo	7.785	182.268	167.679
El asombroso viaje de Pomponio Flato	2.617	45.515	40.349
TOTAL	18.760	351.312	317.257

El corpus constituido de esta manera será procesado y explotado con una herramienta propia. Este corpus es de un tamaño suficientemente grande para los objetivos de esta investigación. No obstante, el corpus puede ser ampliado con nuevas obras literarias u otros tipos de textos en un futuro.

¹⁰Los textos españoles: Ruiz Zafón, C. (1999). *Marina*. Barcelona: Edebé; Toscana, D. (2006). *Ejército iluminado*. México, D.F.: Tusquets Editores México; Palma, F.J. (2008). *El mapa del tiempo*. Sevilla: Algaída; Mendoza, E. (2008). *El asombroso viaje de Pomponio Flato*. Barcelona: Seix Barral. Los textos croatas: Ruiz Zafón, C. (2011). *Marina*. Zagreb: Fraktura; Toscana, D. (2015). *Prosvijetljena vojska*. Zagreb: Edicije Božičević; Palma, F.J. (2012). *Vremenska karta*. Zagreb: Fraktura; Mendoza, E. (2011). *Čudesno putovanje Pomponija Flata*. Zagreb: Fraktura.

¹¹<https://sourceforge.net/projects/aligner/>

¹²<https://github.com/danielvarga/hunalign>

5. La herramienta de búsqueda en el corpus

Para llevar a cabo esta investigación hemos desarrollado una herramienta propia llamada ReSiPC (*Regular Expression Search in Parallel Corpora*). El objetivo principal de esta herramienta es realizar búsquedas complejas en corpus paralelos donde al menos una de las versiones lingüísticas disponga de análisis morfosintáctico. Esta herramienta se distribuye bajo una licencia libre (GNU GPL) y se puede descargar de su página web de Sourceforge.¹³ Asimismo, está desarrollada en Python 3 y para poder utilizarla es necesario tener el intérprete instalado en el sistema.¹⁴ La herramienta permite realizar las siguientes acciones: cargar los corpus paralelos, etiquetado morfosintáctico, búsqueda en la versión etiquetada y guardar los resultados en ficheros. En los siguientes subapartados vamos a explicar estas acciones con más detalle.

5.1 Carga de los corpus paralelos

El programa admite diversos formatos de entrada:

- a. Formato Moses: son archivos de texto alineados y separados para cada lengua. La correspondencia de segmentos entre las versiones de cada lengua es línea a línea. Del mismo modo, las versiones etiquetadas morfosintácticamente tendrán la misma correspondencia línea a línea.
- b. Formato TSV: formato de texto separado por tabuladores. En este caso el primer campo tiene que ser el segmento original, el segundo el segmento traducido, el tercero el segmento original etiquetado y el cuarto, que es opcional, el segmento traducido etiquetado.

El programa puede abrir uno o más archivos, que serán tratados simultáneamente.

5.2 Etiquetado morfosintáctico

¹³<https://sourceforge.net/projects/resipc>

¹⁴<https://www.python.org/>

Nuestro programa puede conectarse con Freeling¹⁵ (Lluís Padró, 2011) para llevar a cabo el análisis morfosintáctico. Este analizador está disponible para las siguientes lenguas: alemán, asturiano, catalán, esloveno, español, francés, galés, gallego, inglés, italiano, noruego, portugués y ruso. Es posible trabajar con otras lenguas utilizando otros analizadores, siempre que se transforme el formato del texto analizado al formato requerido por el programa.

La conexión con Freeling se puede realizar de dos maneras:

a. Mediante la API de Freeling, lo que implica tener instalado Freeling y su API en la máquina donde se va a utilizar el programa. Para realizar el etiquetado utilizando la API se tiene que escribir:¹⁶

```
(1) python3 ReSiPC.py --st original-spa.txt --tag_freeling_api /path/to/freeling/api --output etiquetado-spa.txt --freeling_api_lang="es"
```

b. Mediante la conexión a un servidor de Freeling, que puede estar funcionando tanto en la propia máquina como en un servidor remoto:¹⁷

```
(2) python3 ReSiPC.py --st original-spa.txt --tag_freeling_server 6543 --output etiquetado-spa.txt
```

Ahora ya disponemos del texto en español, el texto español etiquetado morfosintácticamente y el texto en croata. Estos archivos están alineados a nivel de segmento. ReSiPC puede funcionar con los tres archivos por separado o bien con archivos de texto separado por tabuladores que contengan los tres campos: segmento en español, segmento analizado morfosintácticamente en español y segmento en croata. Podemos observar un ejemplo en (3).

¹⁵<http://nlp.lsi.upc.edu/freeling>

¹⁶Es necesario substituir *original-spa.txt* por el nombre real del archivo a etiquetar, */path/to/freeling/api* por la ruta real a la API de Freeling, *etiquetado-spa.txt* por el nombre del archivo de salida deseado y "es" por el código de lengua a etiquetar.

¹⁷En este caso se considera que el servidor está en la máquina local e iniciado en el puerto 6543. Si está en la máquina con IP 213.74.32.71 y en el puerto 8798 se puede indicar poniendo: `--tag_freeling_server 213.73.32.71:8798`

(3) Al atardecer del quinto día de viaje avistamos un campamento romano. A|a|
SP e|e|DA0MS0 atardecer|atardecer|NCMS000 de|de|SP e|e|DA0MS0 quinto|5|
AO0MS00 día|día|NCMS000 de|de|SP viaje|viaje|NCMS000 avistamos|avistar|
VMIP1P0 un|uno|DI0MS0 campamento|campamento|NCMS000 romano|romano|
AQ0MS00 .|.|Fp U predvečerje petoga dana razaznali smo rimski logor.

Obsérvese que las palabras analizadas morfosintácticamente se componen de tres campos separados por "|": forma, lema y etiqueta morfosintáctica. Las etiquetas son las utilizadas por Freeling, que están basadas en EAGLES¹⁸ y se pueden encontrar en la documentación de Freeling.¹⁹ Es posible utilizar cualquier otro analizador morfosintáctico siempre que transformemos la salida al formato de texto analizado que requiere el programa ReSiPC.

5.3 Búsqueda en los corpus etiquetados

El programa permite realizar búsquedas en el corpus de segmentos que contengan un determinado patrón morfosintáctico. Este patrón, en nuestro trabajo, se buscará en el segmento en español analizado morfosintácticamente. El formalismo permite especificar secuencias de formas, lemas o etiquetas morfosintácticas. Por ejemplo, el patrón (4) seleccionará todos los segmentos que en español contengan verbos seguidos de preposición y seguidos de sustantivo.

(4) ||V.* ||SP.* ||N.*

Como se puede observar en el ejemplo, en los patrones se pueden utilizar expresiones regulares estándar, lo que confiere una gran potencia al formalismo. También se pueden expresar las formas o los lemas concretos que se deseen buscar. Por ejemplo, el patrón en (5) seleccionará todos los segmentos que en español contengan el

¹⁸<http://www.ilc.cnr.it/EAGLES96/annotate/node9.html>

¹⁹<https://talp-upc.gitbooks.io/freeling-4-1-user-manual/content/tagsets/tagset-es.html>

verbo *tender* (en cualquier forma, ya que hemos especificado el lema), seguido de *a* y seguido de verbo en infinitivo.

(5) |tender| |a| ||VMN.*

El formalismo, además, permite indicar una serie de elementos cualquiera entre dos elementos definidos. Por ejemplo, la expresión en (6) indicaría que entre cualquier forma del verbo *dar* y *a* pueden ver desde 0 (ninguno) a 5 tokens cualesquiera. Esta expresión detectaría todos los siguientes casos: *...dió a...*, *...dió algo a...*, *...dió un regalo a...*, *...dió un gran regalo a...*, *...dió un regalo muy bonito a...*, *dió un regalo realmente muy bonito a...*

(6) |dar| *{0,5} |a|

Para realizar la búsqueda con ReSiPC, necesitaremos tener la expresión o expresiones regulares que queramos buscar en un archivo de texto (*expresiones.txt* en el ejemplo). Si trabajamos con el corpus en tres archivos diferentes, escribiremos como en (7), donde *traduccion-hrv.txt* es el fichero que contiene los segmentos de la traducción croata. En el archivo *resultado.txt* (o cualquier otro nombre que indiquemos), tendremos los resultados de la búsqueda.

(7) python3 ReSiPC.py --st original-spa.txt --tt traduccion-hrv.txt --tst etiquetado.txt --patterns expresiones.txt --output resultado.txt

Si trabajamos con un corpus en formato tabulado (*corpus-tabulado.txt*), escribiremos como en (8), donde *corpus-tabulado.txt* es el corpus en formato de texto tabulado.

(8) python ReSiPC.py --tsv corpus-tabulado.txt --patterns expresiones.txt --output file.txt

Es posible indicar más de un archivo separando estos por “:”, lo que se puede observar en (9).

(9) --tsv corpus1.txt:corpus2.txt:corpus3.txt

5.4 Formato de los resultados

El programa ofrece los resultados en un fichero de texto donde se muestra la oración original y traducida que satisfaga alguno de los patrones indicados por el usuario. Esta información se complementa con el patrón encontrado y el nombre del archivo donde se encuentra la oración, como se puede ver en (10).

(10) |pensar| *{0,5} |en|SP.* corpus1-spa.txt —La verdad es que yo no
pienso mucho en eso —dije. “Ja zapravo ne razmišljam mnogo o tome”, rekao
sam.

Los resultados se pueden ofrecer agrupados y ordenados por diversos criterios. También se puede indicar un número máximo de ejemplos para cada patrón en cada archivo.

6. Parte experimental

El objetivo de este estudio es analizar de manera empírica tres estructuras del predicado español y sus traducciones al croata. A dos de las estructuras analizadas, *verbo + complemento directo* (CD) y *verbo + complemento de régimen* (CR), las denominaremos biactanciales y nos centraremos en los verbos que se pueden utilizar con ambos complementos. La tercera estructura, donde los dos complementos coocurren (*verbo + CD + CR*), se denomina triactancial.

Para realizar este estudio, hemos utilizado el corpus y las herramientas de búsqueda descritos en los apartados anteriores, así como la base ADESSE: *Base de datos de*

*Verbos (Alternancias de Diátesis y Esquemas Sintáctico-Semánticos del Español).*²⁰ No se ha realizado por el momento un análisis de las posibles variaciones dialectales.

6.1 Metodología

Tomando una aproximación ascendente, es decir, guiada por los datos, hemos utilizado ADESSE para buscar los verbos que admiten la estructura con el CR. En esta base los verbos están divididos en clases semánticas y para cada clase es posible obtener los esquemas sintácticos (construcciones) más frecuentes, junto con sus ejemplos en el corpus. Esto nos ha permitido buscar el esquema “S R (*prep*)” y combinarlo con los esquemas “S D”, para los verbos bivalentes y con el esquema “S D R (*prep*)” para los trivalentes, donde el orden de los complementos no tiene importancia.²¹ Para limitar nuestro análisis, hemos tenido en cuenta los verbos que tenían más de veinte ocurrencias en el corpus con cada esquema.

Puesto que queríamos observar las traducciones de estas estructuras al croata, hemos buscado ejemplos de cada uno de los verbos en el corpus paralelo unidireccional español-croata, utilizando la herramienta de búsqueda descrita en el apartado 5.

En este paso se llevaron a cabo tres búsquedas diferentes: 1) para la estructura transitiva (con el CD) se consideraron ejemplos sin una preposición inmediatamente detrás del verbo (11); 2) para la estructura con el CR se buscaron las construcciones de *verbo + preposición*, donde algunos verbos admiten varias preposiciones (12); 3) para la estructura triactancial (con ambos complementos) se hizo la búsqueda de verbos seguidos de preposición la cual puede encontrarse hasta cinco posiciones detrás del verbo (13).²² Los ejemplos de las expresiones regulares que se utilizaron para cada grupo de verbos aparecen a continuación:

²⁰Sobre el proyecto y la base ver: García-Miguel et al., 2005, y otras publicaciones disponibles en: <http://adesse.uvigo.es/index.php/ADESSE/>.

²¹En estos esquemas *S* es sujeto, *R* complemento de régimen con su preposición específica entre paréntesis y *D* complemento directo.

²²Se limitó el número de posiciones en las que puede aparecer la preposición para simplificar el análisis manual. Hay que advertir que de esta manera no se recogen todos los ejemplos en el corpus, ni es posible llevar a cabo un análisis estadístico.

- (11) |separar| |[^[S].*
- (12) |separar| |S.*
- (13) |separar| *{0,5} |de|S.*

Cada grupo de ejemplos se analizó manualmente y se omitieron los verbos que en el corpus paralelo aparecieron en menos de cinco ejemplos en cualquiera de las tres estructuras, así como los verbos muy ambiguos, por la dificultad de discernir entre sus acepciones. Se seleccionaron tanto los verbos en construcción activa, como los verbos pronominales con *se*.²³ Finalmente, se marcaron las traducciones de todas las estructuras al croata.

6.2 Resultados

Las categorizaciones teóricas no se corresponden necesariamente con la manera más simple y lógica de procesar la lengua automáticamente. Creemos que partiendo de los datos y usando expresiones regulares deliberadamente muy genéricas para su extracción, tenemos la posibilidad de ir refinando las herramientas y el análisis en la dirección ascendente, conservando siempre una relación bidireccional con la teoría lingüística.

Concretamente, se ha podido observar en el apartado anterior que la metodología usada para la extracción de datos tenía varios pasos independientes que aún quedan por automatizar. Por esta razón ofreceremos a continuación una muestra del análisis con algunos ejemplos seleccionados, haciendo hincapié en las líneas de investigación que se abren.

Según se explica en Francisco Albertuz (2007, p. 2016), la clasificación de verbos en ADESSE empieza con el etiquetado de papeles semánticos de los constituyentes sintácticos identificados en la BDS (*Base de datos sintácticos del español actual*),²⁴ lo

²³En ADESSE: *voz activa y voz media*. Para las dificultades del análisis de *se* y su relación con la transitividad, v. Cano, 1987, pp. 256-301.

²⁴En la construcción de ADESSE se utilizó el análisis sintáctico hecho anteriormente en la BDS, así que la BDS sirvió como una base para ADESSE. Ver más en Albertuz (2007) y en las publicaciones

que “conduce inmediatamente a tareas de clasificación verbal y de separación en acepciones”. Podemos concluir que los papeles semánticos tendrán el rol central en la clasificación de los verbos, porque dentro de una misma clase verbal los verbos compartirán los papeles semánticos, que, por su lado, se asociarán con el significado básico de la clase entera. Albertuz (2007, p. 2015) aclara que esto se debe a que en el proceso de etiquetado semántico del corpus en ADESSE se siguió “un nivel intermedio de generalidad.” De esta manera se aumenta el número de roles en comparación con los generales de p. ej., Agente o Paciente, pero se evita la especificación completa de los distintos papeles para cada verbo, facilitando la comparación entre ellos.

En este trabajo hemos seleccionado ejemplos de construcciones biactanciales y triactanciales, es decir, trabajamos con dos grupos de verbos: 1) los verbos bivalentes que se pueden construir tanto con el CR como con el CD, lo que hemos denominado la *alternancia* entre dos complementos (p. ej., *olvidar* + CD y *olvidarse* + CR(*de*); *pensar* + CD y *pensar* + CR(*en*))²⁵ y 2) los verbos trivalentes que en su estructura argumental aceptan ambos complementos a la vez (p. ej., *obligar* + CD + CR(*a*); *convertir* + CD + CR(*en*)).

6.2.1 Grupo 1: estructuras biactanciales: alternancia entre el CD y el CR

Como ya se ha subrayado, las diferentes acepciones de un verbo en ADESSE no son las mismas que las acepciones de este verbo p. ej., en el *DRAE*, porque los criterios, la perspectiva y el punto de partida son diferentes. En ADESSE las diferentes acepciones de un verbo pertenecen a clases semánticas diferentes, sus argumentos tienen papeles semánticos diferentes (comunes para los verbos de esta clase) y por extensión pueden aparecer en estructuras sintácticas diferentes. Esto significa que la alternancia entre el CD y el CR puede observarse, por un lado, con la misma acepción, donde tendremos el mismo rol semántico para los dos complementos o roles diferentes para cada uno y, por el otro, con acepciones diferentes. Destacaremos aquí ciertas peculiaridades que, sin embargo, deben investigarse más detalladamente con un número más elevado de verbos,

disponibles en: <http://adesse.uvigo.es/index.php/ADESSE/Textos>.

²⁵Para más información sobre este tema v. Bojana Mikelenić (2018).

para poder obtener conclusiones sobre las tendencias en la traducción y su relevancia estadística.

En el caso de la misma acepción y el mismo rol semántico, en la traducción de los verbos de este grupo se han encontrado las mismas soluciones para ambas estructuras. Por ejemplo, *olvidar* + CD y *olvidarse* + CR(*de*), rol semántico: Contenido, traducción: *zaboraviti* + Acc. o *zaboraviti na* + Acc; *gozar* + CD y *gozar* + CR(*de*), rol semántico: Estímulo, traducción: *uživati u* + Loc. o *uživati* + Acc. Sin embargo, se han encontrado diferencias en significado, en primer lugar de carácter metafórico, que se reflejan también en la traducción, p. ej., *acabar* + CR(*con*) con el significado de ‘matar a alguien’ (cro. *oduzeti život: quitar la vida, dokrajčiti nekoga/zgotoviti s nekim*; literalmente *acabar con alguien* o *terminar a alguien*). El papel de los prefijos en croata se puede observar en el ejemplo de *apresurar* + CD y *apresurarse* + CR(*a*), donde ambas estructuras se han traducido frecuentemente con un verbo prefijado que unía el significado del verbo y su complemento (p. ej., *apresurar el paso: pohitati*).

En el grupo de verbos donde el CD y el CR con la misma acepción no desempeñan el mismo rol semántico, se han detectado diversas dificultades. En primer lugar, la relación entre los roles semánticos en ADESSE, donde se han encontrado traducciones equivalentes en los casos donde los roles son semánticamente más cercanos, p. ej., Contenido para el CD y Asunto para el CR en *pensar* + CD y *pensar* + CR(*en*), traducción con ambos complementos: *(po)misliti* + *(na)* + Acc., *razmišljati o* + Loc. En cuanto a los verbos cuyo complemento está encabezado por la preposición *a*, surge el problema de la diferenciación entre los argumentos, puesto que en español esta preposición puede introducir tanto el CD como el CI y el CR. Asimismo, se han notado en este grupo varios verbos que admiten la estructura triactancial, en la que el CD (o el CI) y el CR coocurren. Estos dos casos, junto con más ejemplos de este grupo, los abordaremos en el subapartado siguiente.

Con acepciones diferentes, los dos complementos siempre desempeñarán roles semánticos diferentes, según se ha descrito arriba. Asimismo, las diferentes acepciones de los verbos en este grupo presentan una diferencia de significado menor o mayor. En la Tabla 2 mostramos tres maneras de expresar esa diferencia de significado en la traducción al croata: traducción con verbos diferentes (*disponer*), traducción con verbos

diferentes según el contenido del complemento (*negar*), y traducción con el mismo verbo que entra en diferentes estructuras para reflejar las diferentes acepciones españolas (*servir*).

Tabla 2 Traducción de varias acepciones de *disponer*, *servir* y *negar* en estructuras diferentes

disponer	ESP	<i>disponerse</i> + CR(a) 'prepararse'	<i>disponer</i> + CD 'preparar'	<i>disponer(se)</i> + CR(<i>de</i>) 'tener'	<i>disponer</i> + CD 'mandar'
	CRO	<i>spremiti se, biti spreman</i>	<i>postaviti</i>	<i>raspolagati</i>	<i>narediti, odrediti</i>
servir	ESP	<i>servirse</i> + CR(<i>de</i>) 'utilizar, aprovechar'	<i>servir(se)</i> + CD 'actuar en servicio de otros'	<i>servir</i> + CR (<i>para</i>) <i>servir</i> + CR (<i>de</i>) 'ser útil, valer'	
	CRO	<i>poslužiti</i> + Dat.	<i>poslužiti</i> + Acc.	<i>(po)služiti</i> + Dat.; <i>za (para)</i> + Acc.; <i>kao (como)</i> + N	
negar	ESP	<i>negar</i> + CD 'decir [que no es verdad]'	<i>negarse</i> + CR(a) 'no querer realizar una acción'	<i>negar</i> + CD 'no conceder [algo] que se pretende o se pide'	
	CRO	<i>zanijekati/poreći</i> (algo verbal)	<i>odbiti</i> (una acción)	<i>uskratiti</i> (un objeto, posesión)	

En el caso de *disponer*, de sus cuatro acepciones, dos admiten la estructura con el CD y dos con el CR y sus significados varían de 'preparar' y 'prepararse' a 'tener' y 'mandar'. Por otro lado, las tres acepciones de *servir* tienen significados similares y se traducen al croata con el verbo *poslužiti* cuyo régimen cambia, lo que se puede ver en la diferencia de caso y/o preposición que sigue a cada verbo. Hemos encontrado ejemplos de *poslužiti* + Dat., *poslužiti* + Acc., *poslužiti* + preposición *za (para)* y *poslužiti* + conjunción *kao (como)*. El tercer tipo de traducción es el del verbo *negar*, donde el significado básico es común, pero las acepciones se distinguen entre sí según lo que se niega –algo verbal, una acción o un objeto (posesión)–. Los verbos en croata, aunque diferentes, también tienen sus significados relacionados de la misma manera, con respecto a lo negado –*zanijekati/poreći*, *odbiti*, *uskratiti*– respectivamente.

6.2.2 Grupo 2: estructuras triactanciales: coocurrencia del CD y el CR

Al analizar los verbos donde el CD y el CR alternan con la misma acepción pero desempeñan roles semánticos diferentes, hemos notado que en la mayoría de los casos estos verbos también admiten ambos complementos a la vez. Dado que en estos casos estamos frente a verbos trivalentes, queríamos profundizar el análisis de este grupo y comparar la traducción de las estructuras biactanciales y las triactanciales. En la Tabla 3 se pueden observar los verbos analizados, junto con las estructuras en las que aparecen también sus traducciones al croata. Hay que acentuar que, como trabajamos con un corpus relativamente pequeño y como la parte de la diferenciación de estructuras requirió un análisis manual, no todas las construcciones en las que aparecen estos verbos según ADESSE han sido confirmadas en nuestro corpus.

Tabla 3 Traducciones de verbos con los que el CD y el CR pueden coocurrir

Verbo	Estructura	Traducción
transformar	<i>transformar</i> + CD	<i>izmijeniti/preobraziti/narušiti</i> + Acc.
	<i>transformarse</i> + CR(en)	<i>pretvoriti/preobraziti se u</i> + Acc., <i>postati</i> + N., <i>prijeći u</i> + Acc.
	<i>transformar</i> + CD + CR(en)	<i>pretvoriti/preobličiti/pretočiti/preoblikovati</i> + Acc. + <i>u</i> Acc.
convertir	<i>convertir/convertido</i> + CR(en)	<i>pretvoriti/preobraziti u</i> + Acc., <i>preobražen/pretvoren u</i> + Acc., <i>nalik na</i> + Acc., <i>postati</i> + Inst./N.
	<i>convertirse</i> + CR(en)	<i>postati</i> + N./Inst., <i>pretvoriti se/prijeći</i> (+ N.) + <i>u</i> Acc., <i>postići</i> + Acc.
	<i>convertir</i> + CD/CI + CR(en)	<i>postati</i> + N., <i>pretvoriti</i> + Acc. + <i>u</i> Acc., <i>učiniti</i> + Acc. + Inst.
separar	<i>separar</i> + CD	<i>razdvajati/rastaviti/dijeliti</i> + Acc.
	<i>separarse/separado</i> + CR(de)	<i>odvojiti se/odvojen od</i> + Gen.
	<i>separar</i> + CD + CR(de)	<i>dijeliti/odvajati/odvojiti</i> + Acc. + <i>od</i> Gen.
unir	<i>unir</i> + CD	<i>ujediniti/(po)vezati/udružiti/objediniti</i> + Acc.

	<i>unir/unido</i> + CR(a)	<i>povezan, u spoju</i>
	<i>unirse</i> + CR(a)	<i>pridružiti/priključiti se</i> + Dat., <i>sjediniti se</i> + Inst.
	<i>unir</i> + CR(a) + CD	<i>pridružiti/pripojiti</i> + Acc. + Dat.
ayudar	<i>ayudar</i> + CD/CI	<i>pomoći/pomagati</i> + Dat.
	<i>ayudar</i> + CR(a)	<i>omogućiti</i> + Acc.
	<i>ayudar</i> + CD/CI + CR(a)	<i>pomoći</i> + Dat. + infinitivo/oración subordinada
obligar	<i>obligarse</i> + CR(a)	<i>prisiliti se na</i> + Acc., <i>truditi se</i> + infinitivo, <i>natjerati se</i> + oración subordinada
	<i>obligar</i> + CD/CI + CR(a)	<i>prisiliti/primorati/(na)tjerati</i> + Acc. + oración subordinada/na Acc.
	<i>verse/ser/estar obligado</i> + CR(a)	<i>biti/osjetiti se prisiljen/primoran/ponukanim</i> + infinitivo, <i>tjerati</i> + Acc. + oración subordinada

Los verbos *transformar* y *convertir* son sinónimos en muchos contextos, así que se suelen traducir al croata con los mismos verbos. Según la clasificación en ADESSE, ambos pertenecen a la clase CAMBIO DE ESTADO, en la que el CD tiene el rol semántico de Entidad y el CR de Atributo. La situación es similar con los verbos *separar* y *unir*, ambos de la clase UNIÓN, donde el CD también tiene el rol de Entidad, mientras que el rol del CR es de Entidad 2, semánticamente muy cercano al de Atributo de los dos primeros verbos. Los dos últimos verbos, *ayudar* y *obligar*, pertenecen a la clase INDUCCIÓN y OBLIGACIÓN respectivamente, donde el rol del CD es Causado para *ayudar* y Obligado para *obligar*, de nuevo con una similitud semántica evidente, y el del CR de Acción en ambos casos.

De seis verbos analizados, encontramos cuatro en la estructura transitiva, aunque en ADESSE todos los verbos admiten dicha estructura. En nuestro corpus no había ejemplos de los verbos *convertir* y *obligar* solo con el CD, pero ambos aparecieron en la estructura triactancial tanto con la forma acusativa del pronombre personal: *convertirlo*, *obligarlo*, como con la dativa: *convertirle*, *obligarle*. Puesto que la forma adecuada del pronombre personal nos revela si se trata del CD o del CI, en los ejemplos donde el complemento es un sustantivo o un sintagma nominal, la prueba de sustitución no funciona. Aunque aquí no abordaremos más este problema, estamos frente a una

dificultad de identificación y separación de los complementos verbales en español, relacionada también con la expansión del *leísmo*. Con el verbo *ayudar*, se encontraron igualmente ejemplos de la estructura con CD o CI, pero en el corpus se confirmó tanto el esquema biactancial como el triactancial. En todos estos casos tanto el CD como el CI se han traducido al croata con los mismos verbos que tenían el mismo régimen independientemente del tipo de complemento en español. Las estructuras transitivas de los verbos *transformar*, *separar* y *unir*, se han traducido al croata con los verbos también transitivos, es decir, con un complemento en acusativo sin preposición.

Todos los verbos menos *ayudar* han sido encontrados en la estructura pronominal con CR. Cuando comparamos estas traducciones entre sí, podemos observar que en todos los casos una de las opciones es la traducción con el verbo también pronominal en croata, lo que podría indicar ciertas similitudes estructurales entre las dos lenguas, pero tiene que investigarse con más verbos y más ejemplos. En cuanto al régimen de estos verbos croatas, es decir, el caso del complemento, tenemos varios y no apreciamos ningún paralelismo entre ellos.

Además de la estructura pronominal o reflexiva con *se*, encontramos cuatro verbos en la voz activa con el CR. En el caso de tres –*convertir*, *unir* y *obligar*– apareció también la forma del participio/adjetivo²⁶ que mantiene el mismo régimen y se traduce al croata con las formas del llamado *adjetivo verbal pasivo* (cro. *glagolski pridjev trpni/pasivni*) derivado del mismo verbo, p. ej., *convertir* – *convertido*: *pretvoriti* – *pretvoren*; *unir* – *unido*: *povezati* – *povezan*. Con el verbo *separar* seguido también del CR pero en la forma pronominal *separarse*, se ha notado la misma tendencia: *separarse* – *separado*: *odvojiti se* – *odvojen*.

Lo que era el criterio distintivo para este grupo de verbos es que todos puedan construirse con dos complementos a la vez –en esas estructuras el CD y el CR coocurren–. Con algunos (*convertir*, *obligar*, *ayudar*) tenemos la dificultad de diferenciar entre el CD y el CI, mencionada anteriormente, que no tiene ningún efecto para la traducción al croata. Podemos solamente notar que los verbos croatas en las traducciones de estas estructuras suelen ser transitivos, es decir, el CD/CI de estas

²⁶Nos referimos aquí a la similitud funcional entre el participio y el adjetivo y no entraremos en el tema de su diferenciación, ya que para este estudio nos importa su traducción al croata que es la misma en los dos casos.

estructuras suele traducirse con una forma en acusativo sin preposición, menos en el caso del verbo *ayudar* cuyo régimen es de dativo. En cuanto al CR, este se traduce con el mismo caso en la estructura biactancial y en la triactancial, lo que era de esperar puesto que los verbos croatas son los mismos y requieren los mismos casos de sus complementos.

Desde el punto de vista semántico es importante acentuar que todos los verbos analizados son trivalentes, es decir, requieren tres argumentos o actantes –el sujeto/agente que realiza la acción, el objeto o la entidad que cambia o resulta afectada por la acción y el segundo objeto o entidad final–. Por esto no sorprende que en todos los casos se pueda observar un paralelismo en la traducción de la estructura triactancial y las dos biactanciales. Lo ilustraremos con el ejemplo del verbo *transformar*. El esquema transitivo se traduce con varios verbos, pero todos también transitivos, es decir, con el régimen del CD en Acc. La estructura con el CR y el verbo pronominal (reflexivo) se traduce con los verbos pronominales o no pronominales seguidos por Acc. preposicional (la preposición *u* en ambos casos).²⁷ En esta estructura el *se* reflexivo en ambos idiomas muestra la similitud con el CD de la primera estructura, dado que en vez de transformar *algo* (el CD), uno se transforma a sí mismo (*se* reflexivo). La estructura donde los complementos coocurren se ve como una extensión o combinación de las primeras dos, así que no sorprende el paralelismo con los verbos trivalentes en croata que exigen un acusativo no preposicional (el CD) y otro preposicional (con la preposición *u*). El mismo modelo se puede observar con los verbos *convertir*, *separar* y *unir*. Por otro lado, el CR de los verbos *ayudar* y *obligar* en la estructura triactancial suele traducirse con una oración subordinada, así que en un futuro análisis se deberían tener en cuenta las oraciones subordinadas en función de diferentes complementos verbales, lo que también abre la polémica de la diferencia en clasificación de las oraciones en la tradición croata y en la española.²⁸

²⁷Tanto el verbo *transformar* como *convertir* pueden traducirse con el verbo *postati* (*llegar a ser*, *empezar a ser*) que requiere el uso del nominativo. En estos casos se detecta un matiz semántico diferente que tiene que tomarse en consideración al analizar más detalladamente las posibles traducciones.

²⁸Por ejemplo, las gramáticas croatas (Barić et al., 2005, p. 467; Katičić, 2002, p. 189; Silić & Pranjković (2007, pp. 329-330) diferencian entre las oraciones subordinadas de objeto (cro. *objektne rečenice*) y de complemento circunstancial (cro. *adverbne rečenice*), pero no mencionan los tipos de

7. Conclusiones

En el análisis de diferentes estructuras del predicado español, hemos dividido los verbos en dos grupos: los bivalentes, con los que el CD y el CR pueden alternar (con la misma acepción o con acepciones diferentes) y los trivalentes, con los que los dos complementos coocurren. Se han identificado diversas opciones de traducción de dichas estructuras al croata, donde hay que tener en cuenta el cambio de régimen de cada verbo (las preposiciones y los casos que estos requieren), el papel de los prefijos, la tendencia de traducir los verbos p. ej., transitivos o pronominales con los verbos del mismo tipo, igual que la relación entre la estructura y el significado. Asimismo, hemos visto que el rol semántico de los complementos españoles tiene un papel importante en su traducción al croata. Todas estas conclusiones, además del valor descriptivo que tienen, pueden ser útiles para el desarrollo de herramientas lingüísticas y de procesamiento del croata, como p. ej., el etiquetado semántico o identificación de los esquemas sintáctico-semánticos en este idioma. También sería oportuno etiquetar morfosintácticamente los segmentos en croata, con el objetivo de facilitar la recogida de datos para un análisis cuantitativo, disminuir el trabajo manual y automatizar la búsqueda de p. ej., posibles equivalencias traductoras de ciertos verbos o estructuras. Finalmente, hemos presentado la herramienta ReSiPC que permite realizar búsquedas complejas en corpus paralelos anotados morfosintácticamente mediante el uso de expresiones regulares. Esta herramienta se distribuye bajo una licencia libre y se puede aplicar al estudio de otros pares de lenguas. Creemos haber ofrecido una muestra de las posibles aplicaciones de este tipo de análisis y metodología en los trabajos contrastivos, que se puede extender a la traducción y el análisis traductológico, así como a la enseñanza de idiomas.

Referencias bibliográficas

Alarcos Llorach, E. (1966). Verbo transitivo, verbo intransitivo y estructura del predicado. *Archivum*, XVI, 5-17.

Alarcos Llorach, E. (1994). *Gramática de la lengua española*. Madrid: Espasa Calpe.

Albertuz Carneiro, F. (2007). Sintaxis, semántica y clases de verbos: Clasificación verbal en el proyecto ADESSE. En P. Cano López (Coord.), *Actas del VI Congreso de Lingüística General, Santiago de Compostela, 3-7 de mayo de 2004, Vol. 2, Tomo 2, Las lenguas y su estructura (Iib)* (pp. 2015-2030).

Artetxe, M., Labaka, G., Agirre, E. & Cho, K. (2018). Unsupervised neural machine translation. En *Proceedings of the Sixth International Conference on Learning Representations*. Vancouver, Canada.

Bahdanau, D., Cho, K. & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. En *3rd International Conference on Learning Representations, {ICLR} San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings* arXiv:1409.0473.

Baker, M. (1993). Corpus linguistics and translation studies: Implications and applications. En M. Baker, G. Francis & E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair* (pp. 233-250). John Benjamins Publishing Company.

Barić, E., Lončarić, M., Malić, D., Pavešić, S., Peti, M., Zečević, V. & Znika, M. (2005). *Hrvatska gramatika*. Zagreb: Školska knjiga.

Bosque, I. (1983). Dos notas sobre el concepto de ‘suplemento’ en la gramática funcional. *Dicenda*, 2, 147-156.

Candalija Reina, J. A. (2008). *Preposiciones vacías del español: complemento de régimen preposicional*. Universidad de Alicante: Biblioteca Virtual Miguel de Cervantes.

Cano Aguilar, R. (1987). *Estructuras sintácticas transitivas en el español actual*. Madrid: Gredos.

Fung, P. (1998). A statistical view on bilingual lexicon extraction: from parallel corpora to non-parallel corpora. En *Conference of the Association for Machine Translation in the Americas* (pp. 1-17). Berlin, Heidelberg: Springer.

García-Miguel, J. M. (1995). *Transitividad y complementación preposicional en español*. Santiago de Compostela: Universidade de Santiago de Compostela.

García-Miguel, J. M., Costas, L. & Martínez, S. (2005). Diátesis verbales y esquemas construccionales. Verbos, clases semánticas y esquemas sintáctico-semánticos en el proyecto ADESSE. En G. Wotjak & J. Cuartero Otal (Eds.), *Entre semántica léxica, teoría del léxico y sintaxis* (pp. 373-384). Frankfurt am Main: Peter Lang.

González, L. & Veleiro, A. (1991). Construcciones pronominales y suplemento (A propósito de las ‘construcciones reversibles’). *Verba*, 18, 411-429.

Gutiérrez, M. L. (1987). Sobre la transitividad preposicional en español. *Verba*, XIV, 367-381.

Hazem, A. & Morin, E. (2017). Bilingual Word Embeddings for Bilingual Terminology Extraction from Specialized Comparable Corpora. En *Proceedings of the Eighth International Joint Conference on Natural Language Processing, Volume 1: Long Papers* (pp. 685-693).

Katičić, R. (2002). *Sintaksa hrvatskoga književnog jezika*. Zagreb: Hrvatska akademija znanosti i umjetnosti: Nakladni zavod Globus.

Katunar, D. (2015). *Ustroj leksikona u konstrukcijskoj gramatici – primjer prijedloga u hrvatskom jeziku*. Doktorska disertacija. Zagreb: Filozofski fakultet.

Katunar, D. & Raffaelli, I. (2018). Contrastive study of governed prepositions in Croatian, English and French. En P. Biały & B. Cetnarowska (Eds.), *Studies in Contrastive Semantics, Pragmatics and Morphology* (pp. 55-70). Katowice: Wydawnictwo UŚ.

Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., et al. (2007). Moses: Open source toolkit for statistical machine translation. En *Proceedings of the 45th annual meeting of the ACL on interactive poster and demonstration sessions* (pp. 177-180). Association for Computational Linguistics.

Lefever, E., Macken, L. & Hoste, V. (2009). Language-independent bilingual terminology extraction from a multilingual parallel corpus. En *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 496-504). Association for Computational Linguistics.

Martínez García, H. (1986). *El suplemento en español*. Madrid: Gredos.

McEnery, T., Xiao, R. & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. Taylor & Francis.

Mikelenić, B. (2018). “La alternancia entre los complementos verbales (CD y CR) en español y la traducción al croata”. En M. Díaz, G. Vaamonde, A. Varela, M.C. Cabeza, J.M. García-Miguel & F. Ramallo (Eds.), *Actas do XIII Congreso Internacional de Lingüística Xeral* (pp. 621-628). Universidade de Vigo.

Padró, L. (2011). Analizadores Multilingües en FreeLing. *Linguamatica*, 3(2), 13-20.

Pazienza, M. T., Pennacchiotti, M. & Zanzotto, F. M. (2005). Terminology extraction: an analysis of linguistic and statistical approaches. En *Knowledge mining* (pp. 255-279). Berlin, Heidelberg: Springer.

Porto Dapena, J. A. (1987). Notas sobre el suplemento. *Thesaurus*, 42/1, 122-136.

Rojo, G. (1983/1990). Sobre los complementos adverbiales. En *Homenaje a Francisco Marsá* (pp. 153-171). Barcelona: Universidad.

Serradilla Castaño, A. (1997-98). El complemento de régimen preposicional. *Cauce*, 20-21, 1017-1051.

Silić, J. & Pranjković, I. (2007). *Gramatika hrvatskoga jezika: za gimnazije i visoka učilišta*. Zagreb: Školska knjiga.

Sinclair, J. (1996). *EAGLES Preliminary Recommendations on Corpus Typology*. s. l.: EAGLES Document EAG-TCWG-CTYP/P. Recuperado de: <http://www.ilc.cnr.it/EAGLES/corpusstyp/corpusstyp.html>

Spitzová, E. (1974). El régimen y los complementos de verbo. *Etudes Romanes de Brno*, VII, 45-57.

Varga, D., Halácsy, P., Kornai, A., Nagy, V., Nemeth, L. & Tron, V. (2005). Parallel corpora for medium density languages. En *Proceedings of the RANLP 2005* (pp. 590-596).

Summary

This article is based on the theory about the relationship between the direct object (spa. *complemento directo* – CD) and the prepositional object (spa. *complemento de régimen* – CR) in Spanish, mainly the possibility of their co-occurrence in the same predicate (Alarcos, 1966; Bosque, 1983; Rojo, 1983), as well as the predicates in which these two complements can alternate. We observe the differences in the

translation of these verbs and their arguments to Croatian. Using a Spanish-Croatian unidirectional parallel corpus and the ReSiPC tool (*Regular Expression Search in Parallel Corpora*), both developed for this research, we show examples of different solutions in translating these structures – different verbs, the same verb with different government, change of the verb depending on the meaning of the complement, as well as certain parallelisms between the structures of the two languages. This approach aims to contribute to the development of the methodology and tools which will facilitate other similar works, while also emphasizing its descriptive value, given the small number of works that compare these two languages.

Keywords: Spanish, Croatian, translation, prepositional object, direct object, parallel corpus

PREPRINT VERSION