

UNIVERSIDAD DE LA REPUBLICA

PROYECTO DE GRADO DE INGENIERÍA EN COMPUTACIÓN

Traducción automática de WordNet al español

Autores:

Javier Gonzalez

Matias Herrera

Supervisores:

Dina Wonsever

Luis Chiruzzo

1 de julio de 2015

RESUMEN

La principal motivación del siguiente proyecto fue diseñar e implementar distintas estrategias para lograr transformar el contenido de la base de datos léxica WordNet del inglés al español. Se realizó un estudio detallado del estado del arte en la materia y se presentan aquellas heurísticas que fueron consideradas más interesantes. Además se elige una estrategia y se la desarrolla, la misma consiste en la utilización de una serie de selectores como paso inicial para luego aplicar una fase de expansión. Dicha fase busca agregar variantes que los selectores iniciales no habían sido capaces de agregar explotando las relaciones entre synsets. Por último se presentan los resultados obtenidos con la estrategia seleccionada.

ÍNDICE GENERAL

1. INTRODUCCIÓN	7
1.1. Wordnet	7
1.2. Objetivos	7
1.3. Resultados esperados	8
1.4. Organización del documento	8
2. ESTADO DEL ARTE	9
2.1. Wordnet	9
2.2. El Multilingual Central Repository	10
2.3. Carencias del MCR actual	11
2.3.1. Carencias encontradas	11
2.4. Construcción de WordNet por métodos automáticos	12
2.4.1. WordNet español	12
2.4.2. WordNet francés	14
3. EXPERIMENTOS PREVIOS	17
3.1. Experimento 1	17
3.1.1. Objetivo	17
3.1.2. Herramientas utilizadas	17
3.1.3. Procedimiento	18
3.1.4. Resultados	18
3.2. Experimento 2	21
3.2.1. Objetivo	21
3.2.2. Herramientas utilizadas	21
3.2.3. Procedimiento	21
3.2.4. Resultados	22
4. ESTRATEGIA DE TRADUCCIÓN	25
4.1. Proceso de traducción del WordNet	25
4.2. Métodos de traducción	25
4.3. Los selectores iniciales	26
4.4. Fase de expansión	28
4.5. Decisiones significativas	29
4.6. Plan preliminar de pruebas	30
5. EVALUACIÓN DE LOS RESULTADOS	31
5.1. Evaluación cuantitativa de los resultados de la fase uno	31
5.1.1. Procedimiento	31
5.1.2. Traducciones	31
5.1.3. Distribución de las traducciones obtenidas según diferentes selectores	32

ÍNDICE GENERAL

5.1.4.	Synsets para los cuales los selectores iniciales obtuvieron resultados	34	
5.1.5.	Comparación con WordNet actual	34	
5.2.	Evaluación de la precisión de los resultados de la fase uno	35	
5.2.1.	Procedimiento	35	
5.2.2.	Selector de monosemia	36	
5.2.3.	Selector de factorización de traducciones	36	
5.2.4.	Selector de adverbios derivados	37	
5.2.5.	Selector de traducción singular	37	
5.2.6.	Conclusión sobre la calidad de los selectores	38	
5.3.	Evaluación de los resultados de la fase dos	39	
5.3.1.	Lemas procesados	39	
5.3.2.	Lemas procesados en la fase 2 con relaciones y con traducciones para estas relaciones	40	
5.3.3.	Comparación con WordNet actual	41	
5.3.4.	Evaluación manual de los synsets desambiguados	41	
5.4.	Evaluación de los resultados en una tarea de semántica léxica	43	
6.	CONCLUSIONES Y TRABAJO FUTURO	45	
6.1.	Conclusiones	45	
6.2.	Trabajo futuro	46	
	Referencias	49	
A.	EVALUACIÓN MANUAL DE LOS RESULTADOS	53	
A.1.	Evaluación Manual de los Resultados de la Fase 1	53	
A.1.1.	Selector de traducción singular	53	
A.1.2.	Selector de factorización	58	
A.1.3.	Selector de monoesmia	63	
A.1.4.	Selector de adverbios derivados	67	
A.2.	Evaluación Manual de los Resultados de la Fase 2	69	
B.	PSEUDOCÓDIGO	71	

INTRODUCCIÓN

1.1 WORDNET

WordNet (Fellbaum, 1998) es una base de datos léxica de propósito general que cuenta con más de cien mil conceptos en inglés e incluye relaciones entre los mismos. En WordNet las palabras son agrupadas en conjuntos de sinónimos llamados “synsets”, proporcionando definiciones cortas y generales, y se almacenan las relaciones semánticas entre los conjuntos de sinónimos. Su uso está actualmente muy difundido y se ha vuelto indispensable para cierto tipos de tareas requeridas para el análisis automático de textos. Es por esto que es necesario contar con un WordNet completo, o al menos lo más completo posible para aumentar la performance de las tareas relacionadas con el análisis léxico, como por ejemplo la desambiguación de palabras.

1.2 OBJETIVOS

Los objetivos principales de este proyecto fueron diseñar e implementar distintas estrategias para lograr convertir el contenido de la base de datos léxica WordNet del inglés al español. Para lograr esto se plantearon los subobjetivos enumerados a continuación:

1. Estudiar y analizar las distintas técnicas que se han utilizado para traducir los conceptos de WordNet al español y a otros idiomas.
2. Evaluar el nivel de cobertura del WordNet en español existente. En particular tomamos como referencia el Multilingual Central Repository.
3. Diseñar e implementar heurísticas para transformar WordNet del inglés al español.
4. Construir una versión de WordNet en español que incorpore los nuevos términos traducidos.
5. Utilizar la base de datos léxica construida y la versión anterior en una tarea de semántica léxica para lograr una evaluación intrínseca de los resultados.

1.3 RESULTADOS ESPERADOS

1.3 RESULTADOS ESPERADOS

Como resultado principal se esperaba obtener un WordNet en español más completo que el original desde el cual se partió. Era de interés evaluar la calidad de los términos agregados y no solamente la cantidad, ya que realizar una expansión agregando una gran cantidad de términos incorrectos no hubiera sido muy útil. Una vez finalizada la obtención de las métricas deseadas y luego de haber evaluado los diferentes selectores utilizados, fue de interés evaluar y comparar los rendimientos del WordNet español original y del expandido en una tarea de semántica léxica. Dicha tarea consistió en extraer todos los lemas del corpus Corin (Grassi, Malcuori, Couto, Prada, y Wonsever, 2001), aplicar Freeling (*FreeLing Home Page*, s.f.) para obtener sus categorías gramaticales y luego consultar tanto en el WordNet original como en el expandido por la existencia de los mismos. Era de esperar que el WordNet expandido comprendiera un porcentaje mayor de lemas que el original.

1.4 ORGANIZACIÓN DEL DOCUMENTO

Este documento se encuentra dividido en seis capítulos. En el primer capítulo se presenta el problema, plantea los objetivos y los resultado esperados. En el segundo capítulo se describen las investigaciones realizadas en la búsqueda de una solución. En el tercero se explican algunos experimentos previos. En el cuarto se desarrolla el método propuesto, alguno problemas encontrados y sus soluciones. En el quinto capítulo se detallan los procedimientos efectuados para la evaluación de la propuesta y finalmente en el sexto y último capítulo se presentan las conclusiones y el trabajo futuro.

ESTADO DEL ARTE

A continuación se expone un estudio del estado del arte enfocado en la construcción de WordNet automáticamente.

2.1 WORDNET

WordNet es una base de datos léxica que cuenta con más de cien mil conceptos en inglés e incluye relaciones entre los mismos. En este las palabras son agrupadas en conjuntos de sinónimos llamados “synsets”, proporcionando definiciones cortas y generales, y además se almacenan las relaciones semánticas entre los conjuntos de sinónimos. Es también un recurso semántico de amplia cobertura y dominio general que se ha vuelto una herramienta muy utilizada para cierto tipos de tareas requeridas para el análisis automático de textos. Es por esto que es necesario contar con una versión de WordNet lo más completa posible para aumentar la performance de las tareas relacionadas con el análisis léxico, como por ejemplo la desambiguación de palabras. WordNet distingue las palabras por sus categorías gramaticales o POS (Part Of Speech). Contiene sustantivos, verbos, adjetivos y adverbios en subconjuntos organizados independientemente. Cada uno de los synsets contiene un conjunto de palabras que son sinónimos, los distintos significados de una palabra están en diferentes synsets. WordNet en su versión 3.0, la última liberada y usada en este proyecto contiene 155287 palabras organizadas en 118812 synsets.

Un synset está formado por un conjunto de lemas, una definición o glosa y ejemplos de uso, como se puede apreciar en el ejemplo siguiente:

Synset:

Lemas - good, right, ripe

Glosa - “most suitable or right for a particular purpose”

Ejemplos de uso - “a good time to plant tomatoes”; “the right time to act”; “the time is ripe for great sociological changes”

2.2 EL MULTILINGUAL CENTRAL REPOSITORY

La mayor parte de los synsets están asociados con otros mediante varias relaciones semánticas, estas relaciones varían en base al POS del synset. Las más importantes son las siguientes:

1. Hiperónimo: Y es un hiperónimo de X si cada X es un (del tipo de) Y (canino es un hiperónimo de perro)
2. Hipónimo: Y es un hipónimo de X si cada Y es un (del tipo de) X (perro es un hipónimo de canino)
3. Término coordinante: Y es un término coordinante de X si X e Y comparten un hiperónimo (lobo es un término coordinante de perro y perro es un término coordinante de lobo)
4. Holónimo: Y es un holónimo de X si X es parte de Y (edificio es un holónimo de ventana)
5. Merónimo: Y es un merónimo de X si Y es una parte de X (ventana es un merónimo de edificio)
6. Tropónimo: el verbo Y es un tropónimo del verbo X si la actividad Y está realizando X de alguna manera (susurrar es un tropónimo de hablar)

2.2 EL MULTILINGUAL CENTRAL REPOSITORY

El Multilingual Central Repository (*MCR MediaWiki*, s.f.) (MCR) sigue el modelo propuesto por el proyecto EuroWordNet. EuroWordNet (Vossen, 1998) es una base de datos léxica multilingüe con wordnets para varios idiomas europeos, que están estructurados como el WordNet. El MCR se integra en cinco idiomas diferentes: inglés, español, catalán, euskera y gallego. El Inter-Lingual-Index (ILI) permite la conexión de las palabras en un idioma, con traducciones equivalentes en cualquiera de los otros idiomas, gracias a las asignaciones generadas automáticamente entre versiones de WordNet. Por ejemplo: el identificador ILI "ili-30-02084071-n" corresponde tanto al synset inglés "eng-30-02084071-n" con lemas "dog, domestic dog" como al synset español "spa-30-02084071-n" con lemas "can, perro". Al mismo tiempo se corresponde con el synset vasco "eus-30-02084071-n" cuyos lemas son "zakur, or, txakur", con el synset "cat-30-02084071-n" para el catalán de lemas "ca, canis familiaris" y también "glg-30-02084071-n" para el gallego con lemas "can, Canis familiaris". La versión actual ILI corresponde a WordNet 3.0, todos los identificadores provienen del synset original en inglés. El ejemplo anterior posee una traducción para cada uno de los idiomas, pero esto no es lo más común, el MCR está incompleto. Para determinar dichas carencias se investigó la herramienta y los resultados obtenidos se detallan en la siguiente sección.

2.3 CARENCIAS DEL MCR ACTUAL

Con el objetivo de buscar las carencias de la herramienta Multilingual Central Repository (MCR), en una primera aproximación, se decidió utilizarla y experimentar con ella. Se utilizó la versión 3.0, la última disponible hasta el momento. Para la realización de esta etapa se utilizó la interfaz web provista por el MCR (*Web MCR Interface MEANING*, s.f.). En todas las búsquedas realizadas se le indicaba al MCR que devolviera los resultados tanto para el inglés como para el español para poder compararlos.

2.3.1 *Carencias encontradas*

A continuación se muestran algunos ejemplos:

2.3.1.1 *Falta de palabras comunes*

Se encontró que faltaban varias palabras comunes, como por ejemplo el nombre “cargador” y el adverbio “no”.

2.3.1.2 *Synsets vacíos*

Observamos además synsets en español que eran accesibles a través de la interfaz web pero que estaban vacíos. Por ejemplo el synset “spa-30-00396699-r” no contenía ninguna variante pero su correspondiente en inglés “eng-30-00396699-r” sí. Esto indicaba que no existían en el MCR en español traducciones para los lemas “meagerly”, “sparingly”, “slenderly” y “meagrely”. Al buscar por ejemplo el adverbio “escasamente”, el correspondiente a una posible traducción de “sparingly”, no se encontraba.

2.3.1.3 *Categoría gramatical adverbios con muy pocas entradas*

Una vez estudiado el MCR, se concluyó que el cubrimiento de los adverbios era muy bajo. Ya mencionamos el caso del adverbio “no”, se encontró además que no existían por ejemplo los adverbios “recién” y “rápidamente” los cuales son de uso común en el español.

2.3.1.4 *Carencia de glosas en español*

Para muchas de las palabras buscadas no se encontró glosa en español. Como ejemplo, en el resultado para el nombre “cuchillo”, “spa-30-03623556-n ” y “spa-30-03624134-n”, observamos que para ninguna de las dos synsets existía una glosa en español.

2.4 CONSTRUCCIÓN DE WORDNET POR MÉTODOS AUTOMÁTICOS

2.3.1.5 *Falta de frases que muestre el uso de los lemas o synset*

Se notó para el idioma español una gran carencia de frases que muestren el uso de los lemas o synset. Dichas frases sí se encontraban disponibles para la versión en inglés. Un ejemplo de esta carencia se observó por ejemplo en el verbo “cocinar”, dicho verbo contenía la frase en inglés “Cook me dinner, please; can you make me an omelette?; fix breakfast for the guests, please”. Sin embargo para el español no existía dicha frase.

2.4 CONSTRUCCIÓN DE WORDNET POR MÉTODOS AUTOMÁTICOS

En esta sección se realiza un estudio sobre el estado actual de la traducción de WordNet a diferentes idiomas. Se exploraron diferentes enfoques que fueron tenidos en cuenta para la construcción del WordNet en español así como también una metodología seguida para la mejora del WordNet francés en la que basamos el trabajo realizado en este proyecto.

Según (Vossen, 1998) se pueden distinguir dos metodologías para la construcción de WordNets:

1. El modelo de fusión, en el cuál se construye una nueva ontología para el lenguaje destino, y las relaciones entre el WordNet de Princeton y el local que se está construyendo son generadas manualmente.
2. El modelo de expansión, en el cual las variantes en inglés asociadas con los synsets del WordNet de Princeton son traducidas utilizando diferentes estrategias.

Como fue planteado anteriormente, uno de los objetivos del trabajo era construir una versión de WordNet en español con mayor cubrimiento que la existente, es decir, se quería mantener la estructura original del WordNet de Princeton. Por esta razón el enfoque estuvo dedicado a trabajos que siguen el modelo de expansión.

2.4.1 *WordNet español*

En (Oliver y Climent, 2011) se exploró la construcción de WordNet basándose en la traducción automática de un corpus en inglés desambiguado por sentidos. Para el modelo de expansión en particular, la idea era traducir las variantes asociadas a los synsets del WordNet de Princeton utilizando diccionarios bilingües u otras estrategias. Posteriormente se revisó si las relaciones entre synsets impuestas por la estructura del WordNet de Princeton eran válidas para la lengua de llegada. Se utilizaron dos algoritmos distintos. El primero tomaba todos los synsets presentes en el corpus, ordenados por frecuencia de aparición de forma descendente. Para cada synset se obtenía el conjunto de todas las oraciones en las que aparecía en la lengua deseada. El lema más frecuente de este conjunto que tuviera la misma categoría gramatical que el synset que

se estaba tratando era elegido como resultado, es decir, como variante apropiada para el synset. El segundo algoritmo era idéntico al primero pero se tomaba únicamente la propuesta más frecuente, siempre y cuando su frecuencia fuera, como mínimo, el doble que la del siguiente candidato. Se probaron estos algoritmos para el español y también para el inglés (para el cual no se debió traducir el texto), lo que permitió utilizarlo como medida de comparación. Se observó que los resultados obtenidos para el idioma inglés fueron notablemente mejores que para el español, lo que dejó en evidencia que el método dependía fuertemente del traductor utilizado. Además se concluyó que los resultados obtenidos son similares a los obtenidos mediante el uso de diccionarios bilingües.

Un enfoque distinto fue utilizado en (Oliver y Climent, 2012). En este trabajo la construcción de WordNet se realizó basándose en la explotación de corpus paralelos con anotaciones semánticas del texto origen en inglés. Debido a la dificultad encontrada para obtener grandes corpus con anotaciones semánticas, los autores propusieron dos estrategias para lograr su objetivo

1. Utilizar corpus etiquetados con sentidos monolingües y traducción automática.
2. Utilizar corpus paralelos y etiquetar automáticamente el texto en inglés con sus sentidos.

Una vez obtenidos los corpus paralelos con la parte en inglés etiquetada, el problema de construir WordNet se reducía a un problema de alineamiento de palabras.

Para este trabajo en particular se utilizaron los corpus Semcor y el corpus de glosas del WordNet de Princeton que contenía las glosas de WordNet 3.0 anotadas semánticamente. Para la alineación de palabras se utilizó el Berkeley Aligner. En cuanto a los corpus paralelos se utilizaron varios subsets del European Parliament Proceedings Parallel Corpus.

La evaluación fue realizada automáticamente utilizando la versión preliminar del WordNet 3.0 en español. Esto tenía como gran desventaja que al estar incompleto el WordNet español, muchas propuestas correctas podían ser evaluadas como incorrectas. Para el método de corpus etiquetados con sentidos monolingües y traducción automática se obtuvieron variantes para 3880 synsets con una precisión de al menos 80 % y 8866 con una precisión de 75 %. Considerando probabilidades de alineamiento de al menos 0.9, se obtuvieron variantes para 7996 synsets con 80 % de precisión y 10306 con una precisión del 75 %. Esto representaba un número muy bajo considerando que el WordNet inglés tiene 117659 synsets. Para el método de corpus paralelos y etiquetado automático del texto en inglés con sus sentidos el resultado obtenido fue un poco inferior, se pudieron obtener variantes para 3390 synsets con una precisión de al menos 80 % utilizando un subset de oraciones de un millón.

Concluyeron que el método basado en traducción automática de corpus desambiguados obtenía resultados similares a los que proveen los basados en diccionarios bilingües. Se observó además que el método basado en etiquetado automático tenía peor rendimiento que el de traducción automática.

2.4.2 *WordNet francés*

En (Pradet, de Chalendar, y Desormeaux, 2014) se propuso mejorar el WordNet francés utilizando como base al existente JAWS (Mouton y de Chalendar, 2010), el cual es una traducción de los sustantivos en francés construido utilizando diccionarios bilingües y un modelo de lenguaje sintáctico. Con WoNeF se logró un WordNet con mayor precisión y cubrimiento además de tener traducciones para todas las categorías gramaticales.

El procedimiento de construcción fue el siguiente:

1. Se creó un WordNet vacío, con la estructura WordNet, pero sin literales asociado a los synsets.
2. Se seleccionaron las traducciones más fáciles entre diccionarios candidatos para comenzar a llenar espacios.
3. Se amplió WoNeF progresivamente usando el modelo de lenguaje, las relaciones entre synsets y el JAWS existente.

Una vez construida la estructura del WordNet se inició la etapa 2 con los siguientes cuatro selectores iniciales:

1. El selector monosemia que seleccionaba las palabras que aparecían en un solo synset ya que no eran ambiguas: todas sus traducciones eran agregadas al WordNet Francés.
2. El selector de singularidad que identificaba palabras con una sola traducción y seleccionaba esta traducción en todos los synsets donde aparecen las palabras. Por ejemplo, como la única traducción de "pill" era la palabra "pilule", entonces se completaron los cinco synsets en los cuales aparecía la palabra "pill" con la palabra "pilule".
3. El selector de múltiples fuentes que consideraba traducciones candidatas procedentes de varias palabras diferentes de inglés en un synset dado.
4. El selector de Levenshtein que utilizaba dicha distancia de edición. Si la distancia entre una palabra en Inglés y su traducción era pequeña, se podía considerar que tenían el mismo sentido. Para mejorar este selector se utilizó el hecho de que el 28% del vocabulario inglés es de origen francés. Esto implicaba que era posible aplicar los cambios en el literal francés candidato antes de calcular la distancia de Levenshtein para acercar las palabras. Por ejemplo, se podía invertir la "r" y la "e" para acercar las palabras "tiger" y "tigre".

Para la expansión propuesta en el tercer paso se aprovecharon las relaciones entre synsets de WordNet para proponer nuevas traducciones. Se utilizó el modelo de

lenguaje para elegir entre las traducciones posibles. Por ejemplo, si un synset S_1 era merónimo de un synset S_2 en WordNet y existía un contexto en el que un literal seleccionado en S_1 era merónimo de un literal C candidato en S_2 , entonces el literal C era considerado correcto. La tarea de traducción se reducía de esta manera a comparar las relaciones léxicas entre synsets de WordNet y las relaciones léxicas entre lexemas franceses. El modelo de lenguaje no permitía elegir entre dos palabras muy diferentes, pero sí era capaz de elegir la traducción correcta de una palabra polisémica.

Para seleccionar los candidatos a traducción, definieron un umbral de aprendizaje para cada categoría gramatical y selector. Una vez definidos los umbrales, se seleccionaron todos los candidatos por encima de los mismos. De esta manera se conservaron la mayor cantidad de candidatos válidos.

Para la evaluación generaron un gold standard con la anotación manual de dos lingüistas, para los sustantivos, verbos y adjetivos, se anotaron 300 synsets. Para realizar este trabajo, los lingüistas tuvieron que decidir para cada candidato propuesto por los diccionarios bilingües, si pertenecía o no al synset.

Para los resultados, calcularon la precisión y *recall* (exhaustividad) basándose en la intersección de los synsets presentes en WoNeF y el gold standard. La precisión era la intersección de los dos conjuntos de pares sobre el conjunto de pares obtenidos por el WoNeF. El recall por su parte era la intersección de los dos conjuntos sobre el conjunto del gold standard. La precisión obtenida fue del 93 % mientras que la exhaustividad fue aproximadamente del 50 %.

Debido a los buenos resultados obtenidos en (Pradet y cols., 2014) y considerando que las técnicas allí mencionadas eran novedosas y no habían sido aplicadas para la construcción de WordNet en español, decidimos tomar ese trabajo como base para nuestro propio trabajo.

EXPERIMENTOS PREVIOS

Antes de realizar una propuesta para el desarrollo del proyecto se realizaron algunos experimentos.

3.1 EXPERIMENTO 1

3.1.1 *Objetivo*

El primer experimento realizado tuvo como principal objetivo explorar las herramientas disponibles para realizar el trabajo y su cubrimiento para el idioma español. Para llevar a cabo esta tarea se tomó como base el corpus Corin para investigar lo siguiente:

1. El porcentaje de lemas disponibles en la versión en español de WordNet.
2. El porcentaje de synsets que tienen glosa en la versión en español de WordNet.

3.1.2 *Herramientas utilizadas*

Las herramientas utilizadas en esta etapa fueron:

1. NLTK v2.0 (*NLTK 3.0 documentation*, s.f.), una plataforma para la creación de programas en Python para tratar con datos del lenguaje humano.
2. WordNet construido a partir del MCR 3.0, el MCR integra consistentemente el framework EuroWordNet y diferentes versiones de WordNet para cinco idiomas diferentes: Inglés, Español, Catalán, Vasco y Gallego.
3. Freeling, una herramienta de análisis del lenguaje de código abierto, liberado bajo la GNU (General Public License) de la Free Software Foundation.
4. Corpus Corin, un corpus de corte sincrónico que comprende el período 1996-2000. Contiene textos de tipo literario de autores uruguayos (ensayo y ficción) y textos de la producción periodística publicada en Montevideo (artículos y reportajes)

3.1 EXPERIMENTO 1

3.1.3 Procedimiento

Se construyó un diccionario con los lemas de cada palabra del corpus Corin. Para realizar esto procesamos cada texto del corpus y para cada lema del mismo, se obtuvo su POS, sus synsets y sus glosas de WordNet en español. Este procedimiento permitió obtener datos de los porcentajes de cubrimiento que tiene el WordNet en español para el corpus estudiado, tanto en synsets como en glosas, y discriminado por POS.

3.1.4 Resultados

A continuación detallamos los resultados obtenidos.

3.1.4.1 Porcentaje de lemas disponibles de Corin en la versión en español de WordNet

POS	Lemas no encontrados		Lemas encontrados		Lemas procesados
N	69,29 %	2780	30,71 %	1232	4012
A	51,00 %	840	49,00 %	807	1647
V	75,35 %	1235	24,65 %	404	1639
R	32,79 %	121	67,21 %	248	369
Total	48,70 %	3734	51,30 %	3933	7667

En la tabla anterior se observa del total de lemas procesados, qué cantidad se encontraron en WordNet y además se puede ver la información detallada según la POS de los lemas. A partir de los datos anteriores se construyó la siguiente gráfica donde se pueden observar más claramente los resultados.

3.1 EXPERIMENTO 1

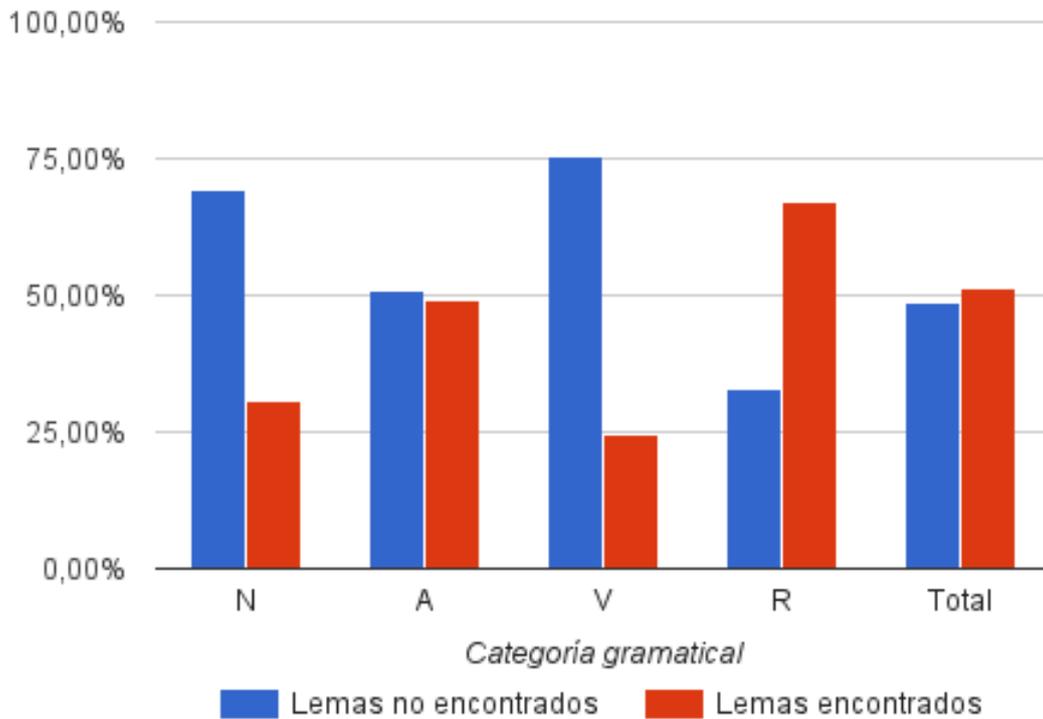


Gráfico 1: Lemas procesados

Se concluye entonces que los verbos son la categoría gramatical que presenta el menor cubrimiento con un 25%. Las restantes POS presentan un cubrimiento mayor, siendo los adverbios los mejores.

3.1.4.2 Porcentaje de synsets de Corin que tienen glosa en la versión en español de WordNet.

POS	Synset con glosa	Synset sin glosa	Synset pocesados
N	31,64 % 1719	68,36 % 3714	5433
A	14,22 % 326	85,78 % 1967	2293
V	40,04 % 1604	59,96 % 2402	4006
R	100,00 % 115	0,00 % 0	115
Total	31,77 % 3764	68,23 % 8083	11847

De manera análoga al punto anterior, se construyó la siguiente gráfica con los datos obtenidos

3.1 EXPERIMENTO 1

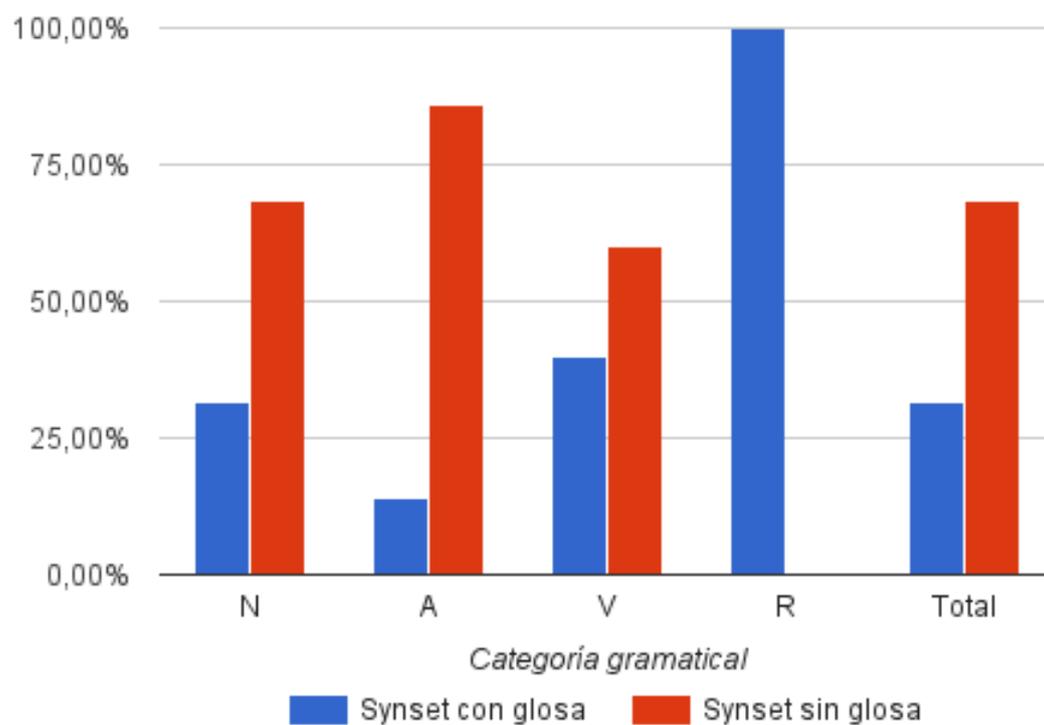


Gráfico 2: Synsets procesados

Se observa entonces que todos los adverbios presentan glosa. Además se ve que las demás categorías gramaticales presentan un cubrimiento muy bajo siendo los adjetivos los peores.

3.2 EXPERIMENTO 2

3.2.1 *Objetivo*

El objetivo de esta etapa fue investigar qué tanto podía ser mejorado el MCR español traduciendo lemas del MCR inglés. En particular interesaba saber qué porcentaje de palabras faltaban en el español pero su traducción sí se encontraba en el MCR inglés. Además en este experimento se comenzaron a utilizar herramientas de traducción. Consideramos relevante obtener las siguientes medidas:

- El porcentaje de lemas del corpus de los cuales se pudo obtener una traducción.
- El porcentaje de estos lemas que no estaba en el MCR español pero su traducción sí se encontraba en el MCR inglés.

3.2.2 *Herramientas utilizadas*

Nuevamente se utilizaron las herramientas del experimento anterior como:

- NLTK v2.0
- MCR 3.0
- Freeling
- Corpus Corin

Pero además se utilizó:

- Apertium (*Apertium*, s.f.), una plataforma de traducción automática de fuente abierta/libre.
- Wiktionary (*Wiktionary*, s.f.), un diccionario multilingüe de contenido abierto, escrito colaborativamente por gente de todo el mundo.

3.2.3 *Procedimiento*

En primer lugar se obtuvieron todos los lemas de cada palabra del corpus Corin, su correspondiente POS, sus synsets y sus glosas de MCR en español. Luego de esto cada lema fue traducido a través de los diccionario Apertium y Wiktionary, y su traducción fue consultada en el MCR Inglés. De esta forma se obtuvieron estadísticas, no sólo sobre la cantidad de lemas ingleses y sus traducciones que aparecen en el MCR Inglés y Español respectivamente, sino también del cubrimiento de los diccionario disponibles hasta el momento.

3.2 EXPERIMENTO 2

3.2.4 Resultados

Los resultados obtenidos se muestran a continuación

3.2.4.1 Lemas del corpus de los cuales se pudo obtener una traducción

POS	Que no se logró traducir		Que se logró traducir		Lemas
N	12,04 %	483	87,96 %	3529	4012
A	21,07 %	347	78,93 %	1300	1647
V	18,00 %	295	82,00 %	1344	1639
R	16,26 %	60	83,74 %	309	369
Total	15,46 %	1185	84,54 %	6482	7667

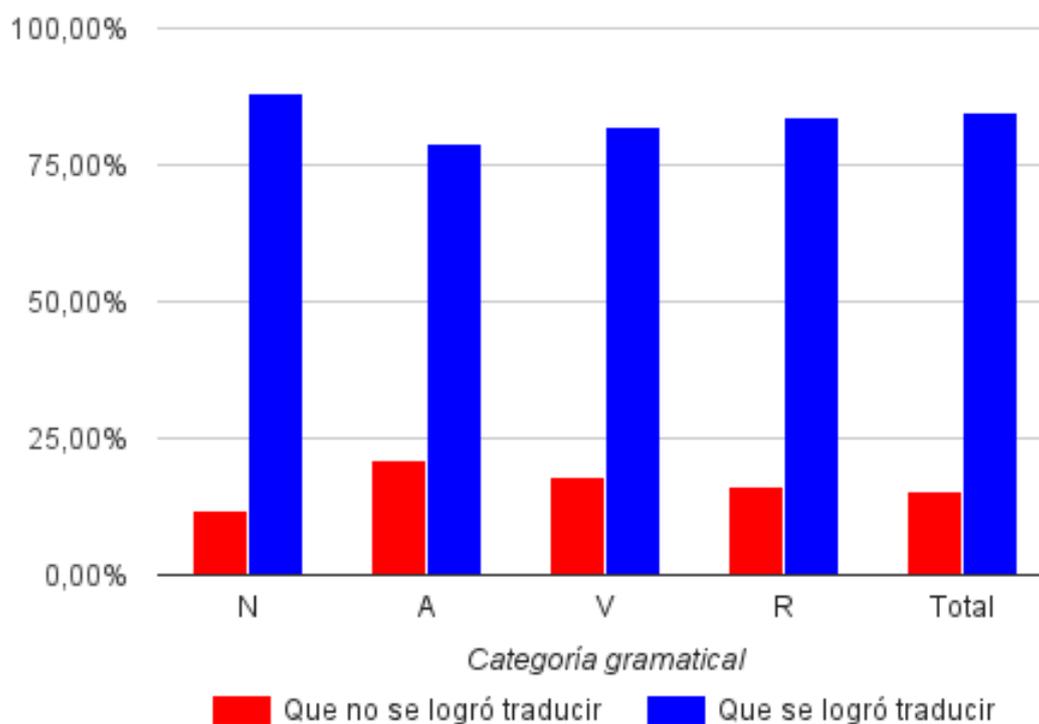


Gráfico 3: Traducciones obtenidas por POS

Observando el gráfico se puede concluir que al unir los dos diccionarios utilizados se logró cubrir un gran porcentaje de los lemas presentes en el corpus. Aun así estos resultados no aseguran la calidad de las traducciones, por esta razón será necesario mejorar los recursos utilizados para realizar las mismas.

3.2 EXPERIMENTO 2

3.2.4.2 *Lemas que no estaba en el MCR español pero su traducción se encontraba en el MCR inglés*

Del los 6477 lemas traducidos al inglés se hizo foco en aquellos que se encontraban en el MCR inglés, de esta forma fue posible comparar los lemas que no se encontraban en el MCR español pero su traducción se encontraba en el MCR inglés.

POS	Lemas no incluidos en MCR español		Lemas incluidos en MCR español		Total
N	43,40 %	1349	56,60 %	1759	3108
A	46,37 %	492	53,63 %	569	1061
V	15,02 %	176	84,98 %	996	1172
R	69,00 %	187	31,00 %	84	271
Total	39,27 %	2204	60,73 %	3408	5612

Con los datos de la tabla anterior se construyó el siguiente gráfico:

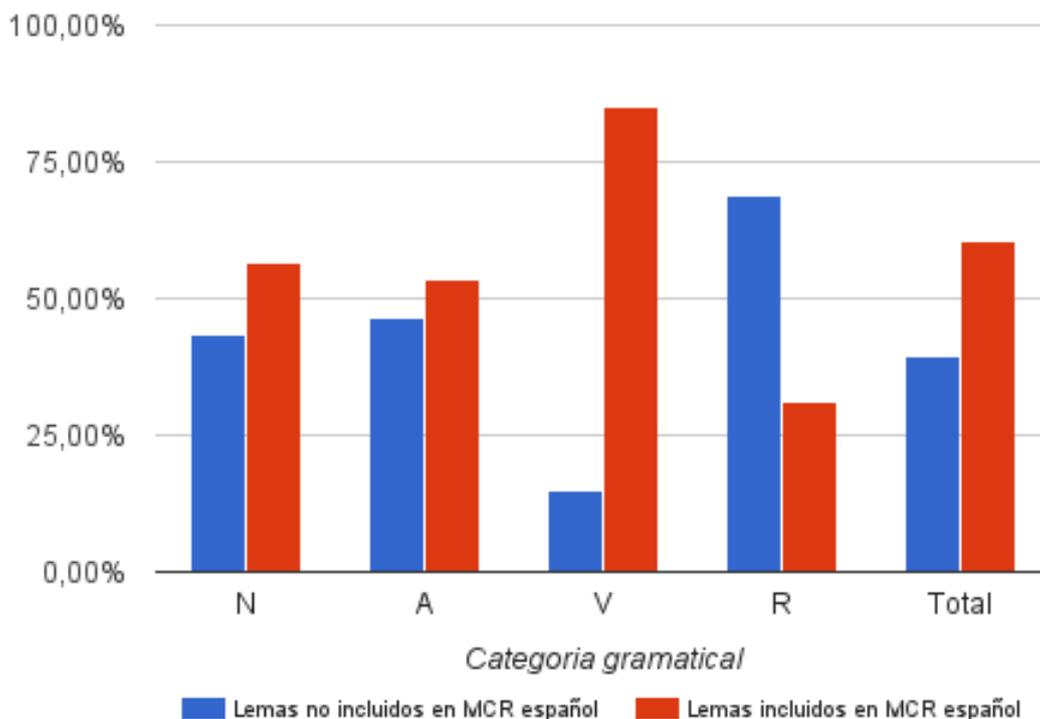


Gráfico 4: Lemas y sus traducciones

Se puede concluir que los verbos son la categoría gramatical con mayor cubrimiento, en cambio, los adverbios son los más incompletos. Además, los nombres y los adjetivos presentan aproximadamente un poco más de un 50% de cubrimiento. Esto dio la pauta de que se debía comenzar a trabajar con los adverbios.

ESTRATEGIA DE TRADUCCIÓN

4.1 PROCESO DE TRADUCCIÓN DEL WORDNET

Para mejorar el WordNet español existente se utilizaron “selectores”. Un selector es un mecanismo que al aplicarlo a un synset inglés se encargará de elegir la o las traducciones que tendrá el synset español generado a partir del original inglés. El proceso de mejora del WordNet español es el siguiente: se tomó como punto de partida el último WordNet Español que logramos conseguir, el correspondiente a la versión 3.0 del MCR. Se comenzó aplicando los selectores iniciales al WordNet inglés, de esta forma se generaron los términos faltantes en el WordNet español anteriormente mencionado y así se empezó a completarlo. Luego de terminado el proceso inicial se intentó ampliarlo explotando las relaciones entre synsets, es decir, se utilizó un selector por relaciones semánticas. Esta fase fue conocida como fase de expansión.

De los synsets en inglés, se decidió descartar los siguientes debido a que incluían mucho ruido en los procesos de traducción: por un lado los lemas que contenían os símbolos guión, guión bajo y apóstrofe, ya que con mucha frecuencia las traducciones eran erróneas. Luego de intentar varias aproximaciones debido al bajo porcentaje de lemas que contenían dichos caracteres y al esfuerzo que implicaría su traducción se decidió descartarlos. Por otro lado también se decidió descartar aquellos synset y lemas que comienzan en mayúsculas ya que la mayoría de los mismos corresponden a nombres propios; si bien algunos sí deberían ser traducidos, como por ejemplo los nombres de ciudades, estos no fueron priorizados para ser tratados en esta primera etapa.

4.2 MÉTODOS DE TRADUCCIÓN

Un paso clave para llevar a cabo esta técnica de creación del WordNet español a partir del WordNet inglés fue el proceso de traducción utilizado. En nuestro caso se utilizaron dos técnicas diferentes: traducción automática y diccionarios. En el caso de los diccionarios se utilizó el Wiktionary y un diccionario creado a partir de los archivos XML fuentes del diccionario Apertium. La traducción automática utilizada fue la provista por el traductor Bing (*Microsoft Translator*, s.f.). Se eligieron estas he-

4.3 LOS SELECTORES INICIALES

ramientas debido principalmente a su disponibilidad, ya que son gratis y/o libres. Wiktionary y Apertium fueron descargados desde sus respectivas páginas web, en cambio el traductor Bing fue consultado en línea a través de su API.

El traductor Bing de Microsoft no toma en cuenta la categoría gramatical de la palabra a traducir por lo tanto se encontraron casos en los que al traducir verbos, se obtenían nombres o en ocasiones el mismo verbo pero en alguna de sus formas conjugadas y no en infinitivo como se buscaba. Para solucionar este problema se decidió tomar las salidas del traductor y obtener su análisis morfológico aplicando Freeling. El procedimiento consiste en obtener todas las posibles categorías gramaticales de la palabra y su lema para luego seleccionar las palabras con la categoría gramatical que tenía la palabra en inglés que se tradujo originalmente.

Se decidió utilizar un diccionario creado a partir de los archivos XML fuentes del diccionario Apertium en lugar del diccionario Apertium ya procesado, debido a que, por alguna razón, al consultarlo solamente se obtenía una traducción posible, incluso cuando el archivo XML contenía más. Usando los archivos XML fuentes se extrajeron todas las traducciones posibles para cada palabra.

4.3 LOS SELECTORES INICIALES

A continuación se detallan los selectores iniciales:

MONOSEMIA Monosemia toma aquellas palabras que se encuentran en un único synset. Esta condición parece indicar que no es ambiguo y por lo tanto se agregan todas las traducciones obtenidas a los synsets correspondientes en el WordNet español.

Por ejemplo al aplicar este selector al synset “eng-30-00048268-r”, cuyo lema es “currently” las tres traducciones obtenidas por los traductores “hoy”, “ahora” y “actualmente” se seleccionan ya que “currently” solo se encuentra en un synset en el WordNet inglés.

TRADUCCIÓN SINGULAR Este selector toma todas las palabras que tengan traducción única al español y la coloca en todos los synsets de WordNet en español correspondientes. Por ejemplo al aplicar este selector al synset “eng-30-00061528-r” cuyo lema es “abruptly” y se obtiene como traducción “abruptamente” y esta se selecciona ya que es única.

FACTORIZACIÓN El selector de factorización trabaja a nivel de synset. Este toma todos los synsets pertenecientes al WordNet inglés y para cada lema contenido en los anteriores se obtienen todas las traducciones posibles. Una vez formado el conjunto de traducciones de cada lema, el selector selecciona aquellas traducciones que se encuentran como traducción común de todos los lemas del synset, es decir con la intersección de los conjuntos de traducciones de cada lema. Por

4.3 LOS SELECTORES INICIALES

ejemplo el synset “eng-30-01309991-a” cuyos lemas son “artless” e “ingenuous”. Las traducciones de “artless” son: “inocente”, “ingenuo” y “cándido” y las de “ingenuous” son: “inocente” e “ingenuo”, en este caso al aplicar este selector se obtuvo “inocente” e “ingenuo”, como traducción común.

ADVERBIO DERIVADO Este selector obtiene adverbios del WordNet inglés y luego los adjetivos de los que derivan. Para obtener los adjetivos de los cuales derivan los adverbios se utilizó la propiedad “is_derived_from” que es provista por el MCR. Una vez obtenido los synsets adjetivos se obtiene todas las variantes y se traducen para posteriormente aplicarles reglas de derivación morfológica para construir adverbios en español. Aplicando este selector al synset “eng-30-00033562-r” cuyo lema es “mildly” y está relacionado al synset de POS adjetivo “eng-30-01508719-a” cuyo lema es a su vez “mild”, se obtiene “suavemente” y “levemente”. Estas últimas son generadas a partir de las dos traducciones de “mild”: “suave” y “leve” y la aplicación de las siguientes reglas de derivación morfológica.

Si el adjetivo terminaba con la letra “o” esta era remplazada por la secuencia “amente”, un ejemplo de esto es “lento” y el resultado es “lentamente”. Si la terminación del adjetivo era las letras “r” o “n” entonces se le adicionaba la secuencia “amente”, ejemplos son “encantador” y “fanfarrón” y sus respectivos resultados “encantadoramente” y “fanfarronamente”. El resto de los adjetivos que no caía en las categorías anteriores se les adicionaba la secuencia “mente”, por ejemplo “educada” y “educadamente”.

Dado que este selector construye palabras aplicando reglas de derivación morfológica, se observó que en ocasiones se generaban adverbios que no existían en el idioma español. Por esta razón se decidió validarlas contra un corpus de prensa proporcionado por el grupo de PLN. Para esto se extrajeron todos los adverbios del corpus de prensa mencionado, creando así una lista de adverbios contra la cual se validó la existencia de los adverbios construidos por el selector. Si bien esta validación tiene la desventaja de que puede descartar adverbios correctos por no encontrarse en el corpus de referencia, pareció mas valioso tener la certeza de agregar palabras correctas. Además siempre existe la posibilidad de utilizar una lista mayor de adverbios conocidos y para disminuir la cantidad de falsos negativos.

LEVENSHTEIN Este selector utiliza la distancia de edición Levenshtein, basándose en que si la distancia entre una palabra en Inglés y su traducción es pequeña, se puede considerar que tienen el mismo sentido. Para disminuir la distancia entre una palabra y su traducción se le realizan pequeñas transformaciones. Un ejemplo de estas transformaciones es la inversión de la “r” y la letra “e”, que puede ser aplicada en el caso de la palabra “tiger” y su traducción “tigre”. Aquí luego de realizar dicha transformación la distancia de Levenshtein pasa a ser

4.4 FASE DE EXPANSIÓN

cero. Al implementar los selectores iniciales se decidió no utilizar este ya que el mismo no arrojó buenos resultados en los experimentos iniciales. Se cree que una posible razón por la que pudo ocurrir esto, es que el español y el inglés no comparten tantos términos cognados como si lo hacen en el caso del inglés y el francés, como fue visto en el artículo WoNeF.

Los selectores de traducción singular, monosemia y Levenshtein fueron inspirados en (Pradet y cols., 2014), los selectores de factorización singular y adverbio derivado fueron una producción completamente propia.

4.4 FASE DE EXPANSIÓN

Para la etapa de expansión se propuso un selector que explotara las relaciones entre los synsets y la frecuencias de apariciones de ambas palabras dentro de un corpus, para determinar qué traducción es correcta en cada synset ambiguo. Cabe aclarar que el escenario de uso de este selector se da cuando se conoce los dos lemas relacionados en inglés, pero al momento de obtener las traducciones correspondientes al español, para uno de ellos se obtuvo una traducción y para el otro varias posibles traducciones.

A continuación se muestra en detalle cómo es aplicada esta fase:

Supongamos que se tienen un synset SA asociado con un synset SB en WordNet a través de la relación de hiperonimia. Además existen dos lemas LA y LB en inglés, de SA y SB respectivamente. Se tiene que las traducciones de LA son TA_1 y TA_2 , y las traducciones de LB son TB . Entonces para poder decidir qué traducción es correcta para este lema se busca en un corpus las ocurrencias de cada traducción, estas búsquedas serán tomadas como un función y denotadas con la letra Θ . A este proceso se le llamó desambiguación.

Por ejemplo en $\Theta(TA_1, TB)$ se buscó en el corpus las ocurrencias de las palabras TA_1 y la palabra TB , dentro de una misma oración.

$$O_1 = \frac{(\Theta(TA_1, TB))}{\Theta(TA_1) + \Theta(TB)}$$

$$O_2 = \frac{(\Theta(TA_2, TB))}{\Theta(TA_2) + \Theta(TB)}$$

En el caso que $O_1 \geq O_2 \implies$ elegimos como la traducción de LA a TA_1 .

En cambio si $O_1 < O_2 \implies$ elegimos como la traducción de LA a TA_2 .

A continuación mostramos un ejemplo de aplicación de esta fase:

Si se tiene que $SA = \text{"eng-30-09776346-n"}$ y $SB = \text{"eng-30-09816771-n"}$ están asociados por la relación de hiperonimia cuyos lemas son $LA = \text{"affiliate"}$ y $LB = \text{"associate"}$

4.5 DECISIONES SIGNIFICATIVAS

respectivamente. Además se sabe que $TB = \text{"asociado"}$ y los candidatos como traducción de "affiliate" son $TA_1 = \text{"filial"}$ y $TA_2 = \text{"afiliado"}$ dado que $O(\text{filial}, \text{asociado}) = 0,0$ y $O(\text{afiliado}, \text{asociado}) = 8,18129755379e^{-05}$ entonces como

$O(\text{afiliado}, \text{asociado}) \geq O(\text{filial}, \text{asociado}) \implies$ se elige como la traducción de LA a $TA_2 = \text{"afiliado"}$.

El resultado anterior es correcto dado que la glosa de $SA = \text{"eng-30-09776346-n"}$ es: "a subordinate or subsidiary associate; a person who is affiliated with another or with an organization".

Las relaciones sintácticas que se utilizaron fueron la meronimia, sinonimia, antagonismo, hiperonimia, hiponimia. WordNet dispone de más relaciones sintácticas pero se consideró que las más completas y frecuentes, y por lo tanto más importantes, son las anteriormente mencionadas.

La idea original era utilizar la web como corpus de desambiguación, por lo que se propuso realizar consultas a Google para obtener la cantidad de resultados retornados por la búsqueda. Lamentablemente cuando se comenzó a realizar las pruebas se encontró con que Google no permitió realizar tantas consultas de forma batch. Esto llevó a tener que buscar otra alternativa para realizar la fase de expansión, por lo que se decidió entonces utilizar el mismo corpus que se utilizó para la validación de los adverbios generados por el selector de adverbios derivados.

Por mas detalles sobre los algoritmos mencionados se puede observar el anexo B.

4.5 DECISIONES SIGNIFICATIVAS

A continuación se detallan algunas consideraciones que se tuvieron al momento de realizar la implementación:

1. El desarrollo fue realizado en el lenguaje de programación Python ya que es el lenguaje más reconocido en el área de procesamiento de lenguaje natural. El mismo cuenta con muchos módulos para el procesamiento de texto que simplifican enormemente la tarea. Se destaca además la rapidez con la que se pudo realizar prototipos para las diferentes pruebas realizadas.
2. Debido a que la estrategia fue implementada en fases y que los selectores iniciales iban a ser evaluados individualmente se decidió utilizar una base intermedia para guardar los resultados de dichos selectores. De esta forma se facilitaba las tareas mencionadas anteriormente y se permitía también intentar diferentes aproximaciones.
3. Debido a los extensos tiempos de ejecución solo se ejecutaron los algoritmos para un conjunto reducido de synsets del WordNet, mil synsets de cada POS. Ejecutar dicha cantidad tomó alrededor de 20 horas en computadoras con un procesador

4.6 PLAN PRELIMINAR DE PRUEBAS

Core i5 con 4Gb RAM, pero se consideró que esa cantidad era suficiente para realizar el estudio deseado.

4.6 PLAN PRELIMINAR DE PRUEBAS

El plan de pruebas pensado para el proyecto buscaba evaluar dos aspectos de los resultados. Por un lado la cobertura de las estrategias implementadas, es decir evaluar el porcentaje de synsets procesados para los cuales se ofrece una solución, y por otro lado la correctitud de los resultados para estos synset. De esta forma se utilizarían dos técnicas diferentes para dichos aspectos, la correctitud de los resultados se evaluaría manualmente sobre un conjunto reducido de synsets elegidos aleatoriamente. La cobertura de la estrategia en cambio será evaluada automáticamente midiendo el porcentaje de synsets traducidos sobre todo el resultado. Al mismo tiempo los dos aspectos se evaluarían en dos casos distintos, el primer caso sería luego de la fase inicial, o sea que solo consideraría los resultados de los selectores iniciales. El segundo caso consideraría los resultados luego de la segunda fase. Esto nos permitirá evaluar por separado las dos tipos de selectores. Cabe aclarar que debido a las limitaciones de tiempo, tanto para la ejecución de la estrategia propuesta como para la parte manual de la evaluación de los resultados, la estrategia sería aplicada a un conjunto de 1000 de los synsets del WordNet.

EVALUACIÓN DE LOS RESULTADOS

Este capítulo se encuentra dividido en tres secciones: la evaluación de la fase uno es decir la evaluación de los resultados de los selectores iniciales, la evaluación de los resultados de la fase dos y la evaluación del resultado total en una tarea de semántica léxica.

5.1 EVALUACIÓN CUANTITATIVA DE LOS RESULTADOS DE LA FASE UNO

5.1.1 *Procedimiento*

La evaluación consistió en elegir aleatoriamente mil synsets para cada POS (verbo, adverbio, nombre y adjetivo). Para cada uno de los lemas de todos los synsets sorteados se obtuvieron todas sus traducciones y se les aplicó cada uno de los cuatro selectores anteriormente presentados. Los resultados obtenidos fueron almacenados en una base de datos.

5.1.2 *Traducciones*

Resultados de los traductores al aplicarlos sobre los lemas pertenecientes a los cuatro mil synset aleatoriamente seleccionados

POS	Con traducción		Sin traducción	
R	82,80 %	1187	17,20 %	246
V	71,90 %	1226	28,10 %	478
A	59,50 %	969	40,50 %	659
N	71,20 %	1036	28,80 %	419
Todas	71,00 %	4418	29,00 %	1802

5.1 EVALUACIÓN CUANTITATIVA DE LOS RESULTADOS DE LA FASE UNO

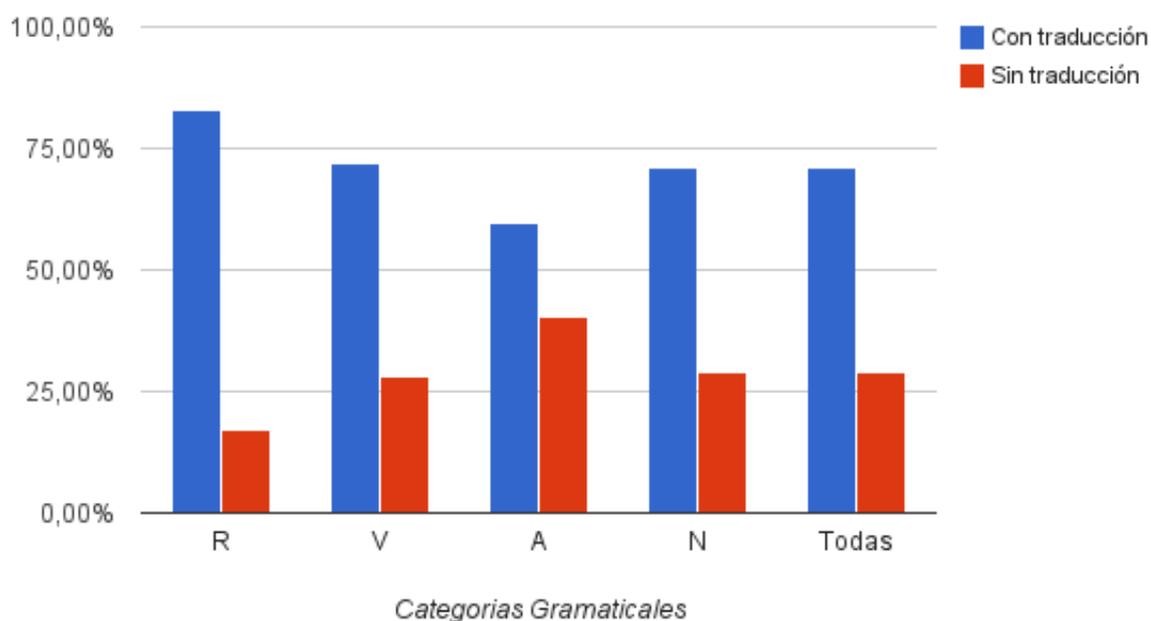


Gráfico 5: Lemas con traducción

Se puede observar en el Gráfico 5 que se logró obtener traducción para el 71 % de los lemas procesados. Si se desglosa el análisis a nivel de categoría gramatical se aprecia que la categoría que obtuvo mayor porcentaje de traducción es los adverbios con más del 80 %. Para las restantes categorías se observa un comportamiento similar entre ellas, siendo los adjetivos, con casi un 60 % de traducciones obtenidas, la categoría con menor cubrimiento.

5.1.3 Distribución de las traducciones obtenidas según diferentes selectores

La tabla a continuación contiene la distribución de las traducciones de los lemas de cada uno de los cuatro mil synsets seleccionados. Como se puede apreciar la idea fue obtener los resultados que obtuvieron cada uno de los selectores sobre el total de lemas traducidos pero sin que se solaparan sus resultados, para esto se les dio un orden de importancia. Dicho orden es el siguiente: selector singular, selector de monosemia, selector de factorización y resto. Aquí el resto contiene para las POS "V", "A", "N" las traducciones que no fueron seleccionadas por ningún selector, para el caso de la POS "R" además de las traducciones no seleccionadas por ningún selector también se incluyen las determinadas por el selector de adverbios derivados.

5.1 EVALUACIÓN CUANTITATIVA DE LOS RESULTADOS DE LA FASE UNO

POS	Singulares		Monosémicos y no singulares		No monosémicos, no singulares y con factorización		Resto	
R	56,40 %	669	6,10 %	72	1,30 %	16	36,20 %	430
V	58,00 %	711	2,00 %	24	0,80 %	10	39,20 %	481
A	77,60 %	752	5,20 %	50	0,90 %	9	16,30 %	158
N	72,70 %	753	4,60 %	48	1,40 %	14	21,30 %	221
Todas	70,20 %	2885	4,70 %	194	1,20 %	49	23,90 %	984

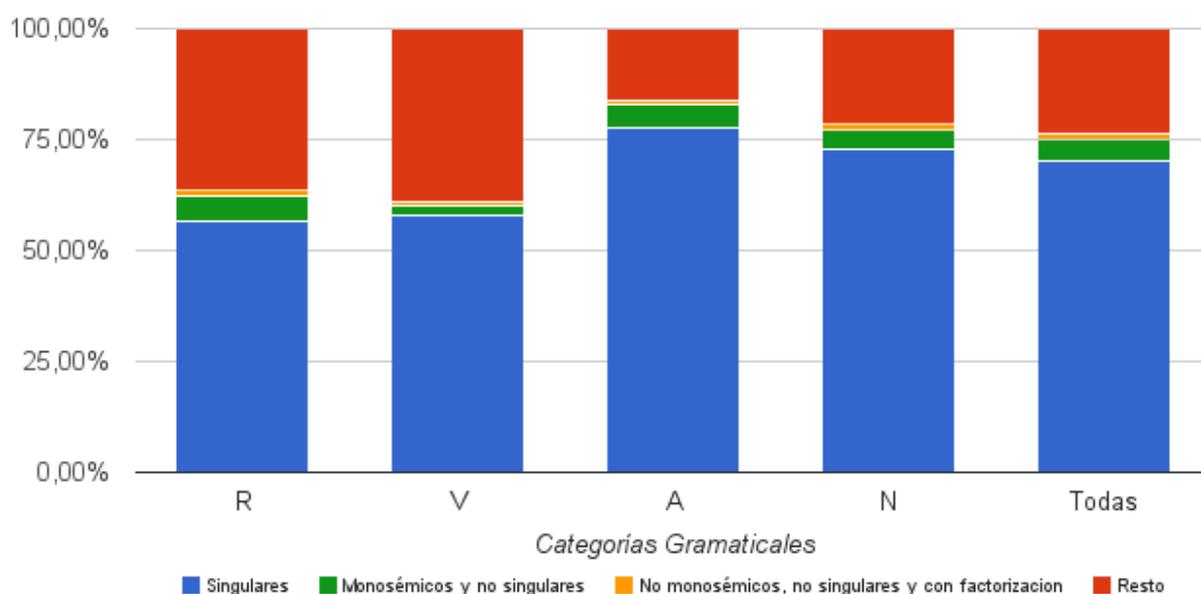


Gráfico 6: Distribución de traducciones por selector

Como se observa en el Gráfico 6, los verbos y adverbios fueron los que presentaron peor resultado, mientras que los adjetivos fueron los que tuvieron el mejor resultado con un 16,3 %. Se debe tener en cuenta que estos datos no contemplan los resultados del selector de adverbios derivados, que se excluyeron en esta comparación puesto que no podría ser comparado con el resto de las POS.

5.1 EVALUACIÓN CUANTITATIVA DE LOS RESULTADOS DE LA FASE UNO

POS	Si	No
R	73,90 % 739	26,10 % 261
V	52,80 % 528	47,20 % 472
A	59,90 % 599	40,10 % 401
N	63,70 % 637	36,30 % 363
Todas	62,60 % 2845	37,40 % 1155

5.1.4 Synsets para los cuales los selectores iniciales obtuvieron resultados

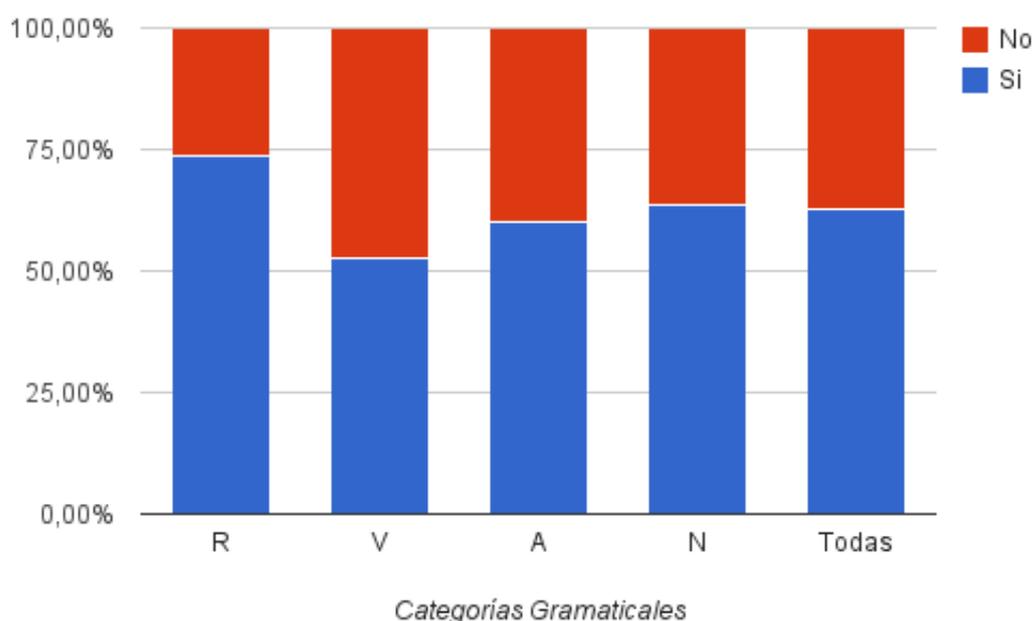


Gráfico 8: Synsets para los cuales los selectores iniciales obtuvieron resultado

Como se puede observar el POS con mayor cubrimiento por parte de los selectores iniciales fueron los adverbios con casi un 74 %, sin distinguir por POS el cubrimiento es de 62,60 %.

5.1.5 Comparación con WordNet actual

POS	Nuevos	Existentes
R	83,80 % 694	16,20 % 134
V	50,40 % 390	49,60 % 384
A	62,50 % 429	37,50 % 257
N	54,80 % 423	45,20 % 349
Todas	63,30 % 1936	36,70 % 1124

5.2 EVALUACIÓN DE LA PRECISIÓN DE LOS RESULTADOS DE LA FASE UNO

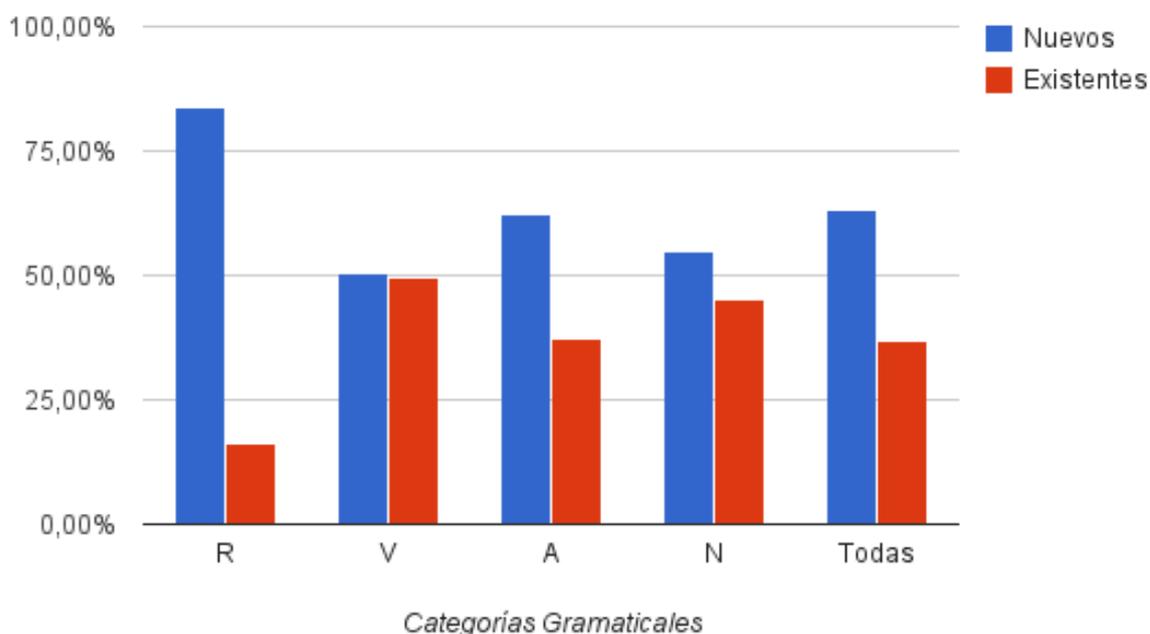


Gráfico 7: Comparación con WordNet actual

Como se puede apreciar en el Gráfico 7 para cada POS hubo un considerable porcentaje de synset para los que se obtuvo traducciones que no se encontraban en el WordNet español actual (MCR 3.0). Los adverbios son la categoría gramatical que más porcentaje obtuvieron, alrededor de un 83 %, en total serían un poco más de 63 % de nuevos synset. Este resultado con solo la aplicación de los selectores iniciales permitió pensar que se lograría una significativa mejora al terminar el proceso.

5.2 EVALUACIÓN DE LA PRECISIÓN DE LOS RESULTADOS DE LA FASE UNO

5.2.1 Procedimiento

Con el objetivo de medir la precisión de los resultados se realizó una evaluación cualitativa manual de los mismos. El procedimiento utilizado consistió en elegir aleatoriamente veinticinco synsets para cada POS (verbo, adverbio, nombre y adjetivo) de los agregados y verificar si el resultado era correcto o no. Para el caso de los selectores que trabajan a nivel de synset los siguientes datos son porcentajes de los synset resultantes correctos o incorrectos y para los selectores que trabajan a nivel de lema son porcentajes de los lemas correctos o incorrectos de los synset resultantes.

Por mas detalle de la evaluación se puede observar el anexo A.

5.2 EVALUACIÓN DE LA PRECISIÓN DE LOS RESULTADOS DE LA FASE UNO

5.2.2 Selector de monosemia

POS	Correctos	Incorrectos
V	93.48 %	6.52 %
A	96.08 %	3.92 %
N	93.48 %	6.52 %
R	97.14 %	2.86 %
Todos	95.04 %	4.96 %

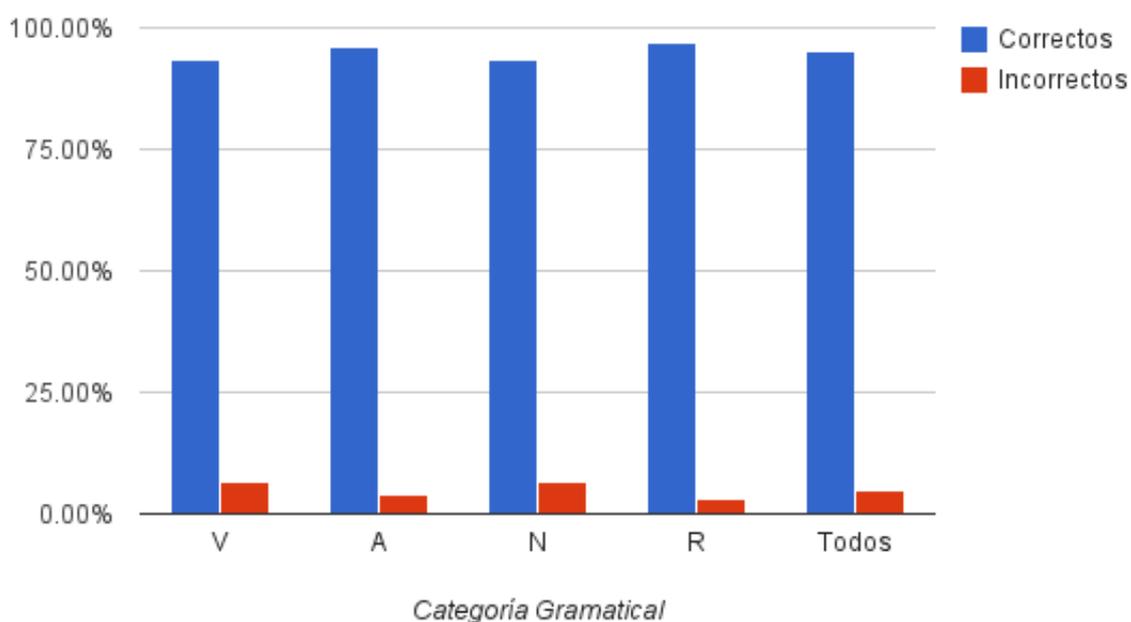


Gráfico 9: Resultados selector de monosemia

5.2.3 Selector de factorización de traducciones

POS	Correctos	Incorrectos
V	100.00 %	0.00 %
A	96.00 %	4.00 %
N	100.00 %	0.00 %
R	92.00 %	8.00 %
Todos	97.00 %	3.00 %

5.2 EVALUACIÓN DE LA PRECISIÓN DE LOS RESULTADOS DE LA FASE UNO

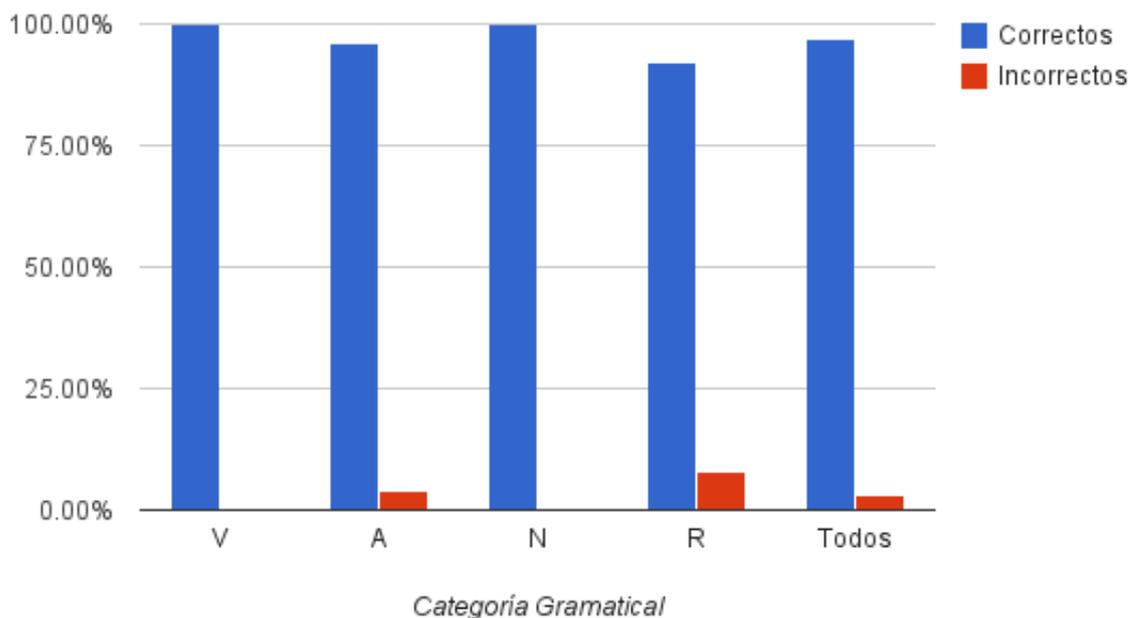


Gráfico 10: Resultados selector de factorización de traducciones

5.2.4 Selector de adverbios derivados

	Correctos	Incorrectos
Adverbios	92.00 %	8.00 %

El selector de adverbios derivados fue el menos preciso pero aún así alcanzó un muy buen resultado de un 92 %.

5.2.5 Selector de traducción singular

POS	Correctos	Incorrectos
V	98.39 %	1.61 %
A	100.00 %	0.00 %
N	100.00 %	0.00 %
R	94.59 %	5.41 %
Todos	98.25 %	1.75 %

5.2 EVALUACIÓN DE LA PRECISIÓN DE LOS RESULTADOS DE LA FASE UNO

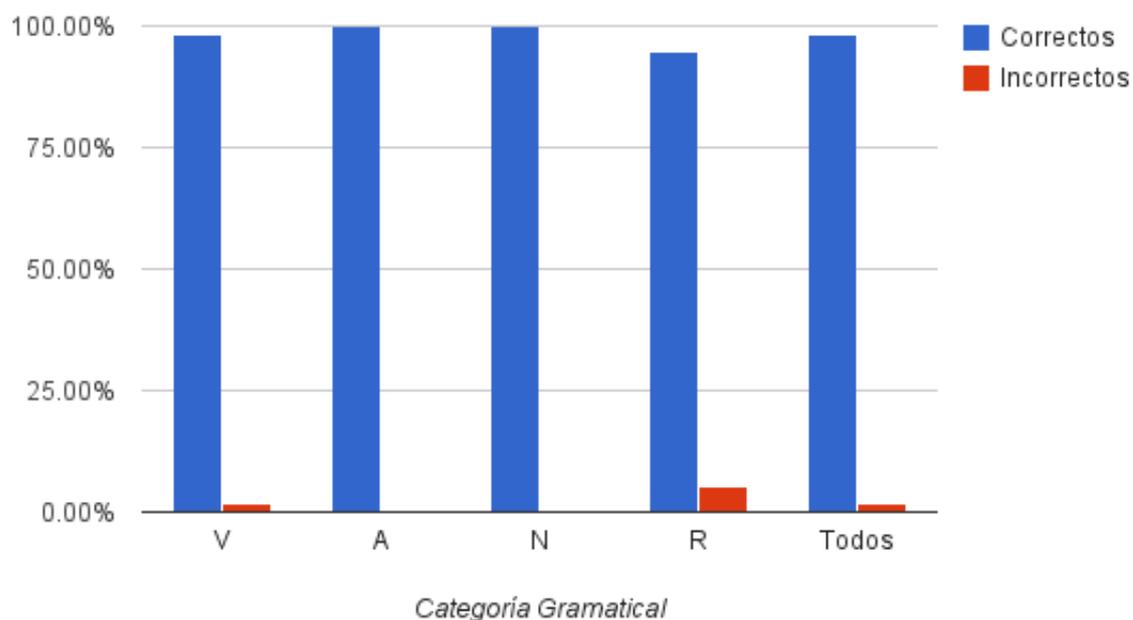


Gráfico 12: Resultados selector de traducción singular

5.2.6 Conclusión sobre la calidad de los selectores

Como se puede observar en las gráficas anteriores los resultados de los cuatro selectores fueron muy buenos, todos sobrepasan el 92 % de efectividad y algunos alcanzan el 100 % para algunas de las POS.

5.3 EVALUACIÓN DE LOS RESULTADOS DE LA FASE DOS

5.3 EVALUACIÓN DE LOS RESULTADOS DE LA FASE DOS

5.3.1 *Lemas procesados*

La aplicación y la evaluación de los resultados de la fase dos se realizó sobre los mil cuarenta synset que la fase 1 no fue capaz de traducir, por ser ambiguos. Dado que la fase dos puede fallar por varios motivos, en esta sección se desglosan los resultados obtenidos para identificar el porqué. Como la fase dos explota las asociaciones entre los synsets existentes en el WordNet construido hasta esta etapa, si los synsets no están relacionados con ningún otro, o lo están pero estos synsets están vacíos para el idioma español, este método no devolverá ningún resultado. Por este motivo se puede observar tres grupos diferentes en la siguiente tabla.

POS	Sin relaciones		Con relaciones y sin traducciones		Con relaciones y con traducciones		Total
R	83,10 %	118	10,56 %	15	6,34 %	9	142
V	1,20 %	6	10,40 %	52	88,40 %	442	500
A	1,79 %	3	34,52 %	58	63,69 %	107	168
N	0,00 %	0	22,17 %	51	77,83 %	179	230
Total	12,21 %	127	16,92 %	176	70,87 %	737	1040

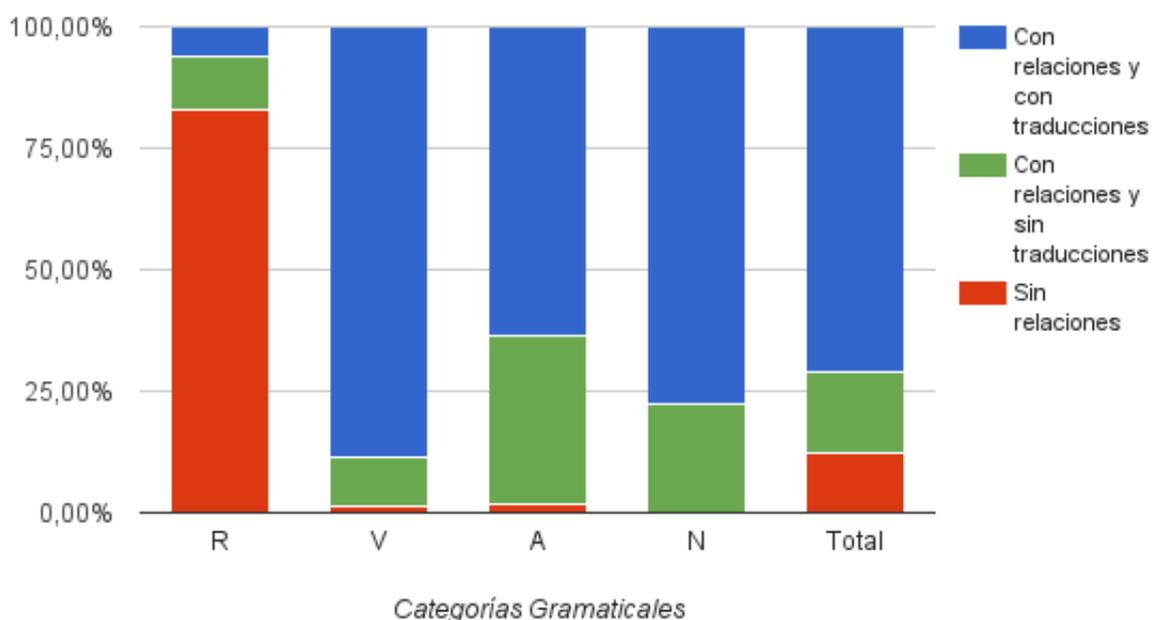


Gráfico 13: Synsets procesados

Como se puede apreciar en la gráfica anterior, los adverbios son la categoría gramatical que presenta los synsets menos relacionados, por lo que nuestro método no

5.3 EVALUACIÓN DE LOS RESULTADOS DE LA FASE DOS

consigue mucho para esta POS. El resto de las categorías gramaticales tienen bastantes relaciones y al mismo tiempo están suficientemente completas para que la fase dos obtenga resultados.

5.3.2 *Lemas procesados en la fase 2 con relaciones y con traducciones para estas relaciones*

Es importante observar que para los lemas que pertenecían a synsets que estaban relacionados con otros synsets ya completos, el método aplicado en la fase dos puede fallar si el número de ocurrencias en el corpus de los posibles candidatos fueran nulas para todos ellos. Por este motivo se presentan los siguientes resultados.

POS	Con resultado		Sin resultado	
R	33,33 %	3	66,67 %	6
V	63,80 %	282	36,20 %	160
A	60,75 %	65	39,25 %	42
N	70,95 %	127	29,05 %	52
Total	64,72 %	477	35,28 %	260

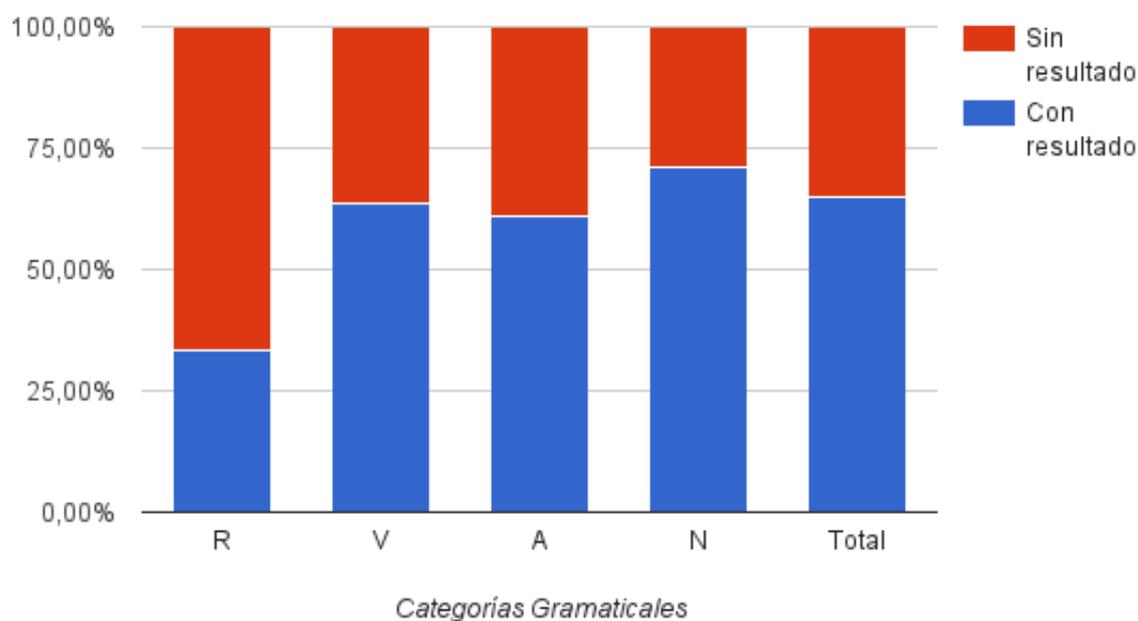


Gráfico 15: Lemas procesados en la fase 2 con relaciones y con traducciones para estas relaciones

Como puede observarse hay un gran margen de mejora, un 35 %, que además puede ser mejorado aumentando el tamaño del corpus de búsqueda.

5.3 EVALUACIÓN DE LOS RESULTADOS DE LA FASE DOS

5.3.3 Comparación con WordNet actual

En esta sección se comparan los resultados obtenidos en la fase dos con los del WordNet actual, ya que solo los resultados inexistentes en el WordNet actual significarán un verdadero aumento en la completitud del WordNet.

POS	Inexistentes		Existentes	
R	66,67 %	2	33,33 %	1
V	73,05 %	206	26,95 %	76
A	52,31 %	34	47,69 %	31
N	62,20 %	79	37,80 %	48
Total	67,30 %	321	32,70 %	156

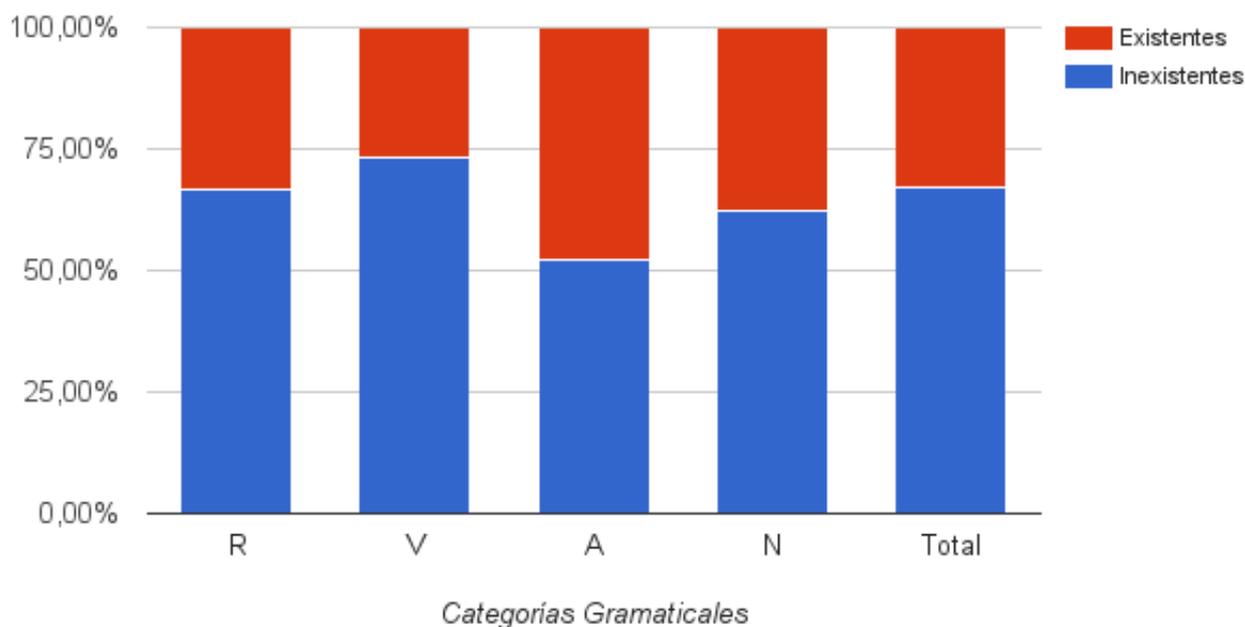


Gráfico 14: Lemas procesados con resultados

5.3.4 Evaluación manual de los synsets desambiguados

Con el objetivo de medir la precisión de los resultados se realizó una evaluación cualitativa manual de los mismos. El procedimiento utilizado consistió en elegir aleatoriamente veinticinco synsets para cada POS (verbo, adverbio, nombre y adjetivo) y verificar si el resultado era correcto o no. Cabe aclarar que para el caso de los adverbios solo teníamos dos resultados. Se debe destacar que la gran mayoría de los errores detectados en esta etapa se corresponden a lemas correctamente traducidos pero cuya traducción no era adecuada para incluir en el synset que estábamos tratando. Un

5.3 EVALUACIÓN DE LOS RESULTADOS DE LA FASE DOS

ejemplo de esta situación es el lema “cup” del synset “eng-30-03147901-n” que tiene el sentido de trofeo. Las traducciones obtenidas para el lema fueron “taza” y “copa” y al desambiguar el proceso seleccionó “taza” lo cual no es apropiado para este synset.

POS	Correctos	Incorrectos
R	100.00 %	0.00 %
V	68.00 %	32.00 %
A	84.00 %	16.00 %
N	68.00 %	32.00 %
Total	74.03 %	25.97 %

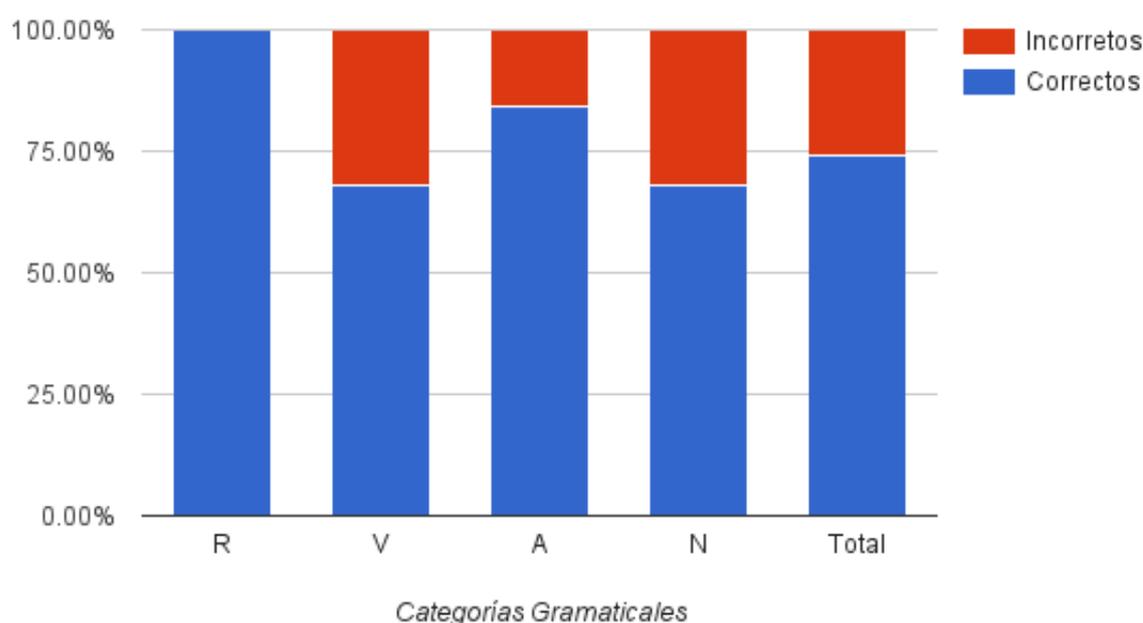


Gráfico 16: Evaluación manual en la fase 2

Se puede concluir de estas evaluaciones, que la fase dos no es tan precisa como la fase uno. Estos resultados podrían ser mejorados de dos formas: aumentando el tamaño del corpus o mejorando el método. Un corpus más grande contaría con más oraciones o sea con más contextos donde validar el significado de los candidatos. Para mejorar el método se podría descartar las traducciones que no coincidieran en género con los lemas en inglés. Ejemplificando esto descartaría casos como el del synset “spa-30-10129825-n” cuya glosa es “mujer joven”. Para el lema “girl” correspondiente a dicho synset inglés, se obtuvo que una traducción posible era “chico”. Este es un claro ejemplo en el cual el lema original en inglés y la traducción obtenida no coinciden en género. Otra forma para mejorar el método sería dar mayor preponderancia a ciertas relaciones.

5.4 EVALUACIÓN DE LOS RESULTADOS EN UNA TAREA DE SEMÁNTICA LÉXICA

5.4 EVALUACIÓN DE LOS RESULTADOS EN UNA TAREA DE SEMÁNTICA LÉXICA

Para poder evaluar los resultados obtenidos por ambas fases se implementó una tarea para medir el cubrimiento semántico sobre un corpus pequeño, en este caso el corpus Corin. Dicha tarea consistió en obtener todos los lemas del corpus, aplicarles Freeling para obtener la categoría gramatical y luego buscar en WordNet. Este proceso fue ejecutado primero contra el WordNet original del cual partimos y luego contra el WordNet resultado. Lo que se buscaba era medir la mejora en el cubrimiento de los lemas presentes en el corpus de estudio del WordNet resultado con respecto al WordNet actual. Cabe aclarar que como el proceso de mejora del WordNet fue ejecutado sobre un conjunto aleatorio de mil synset por POS, los resultados obtenidos deben ser ponderados teniendo en cuenta qué porcentaje representan estos mil synsets del total de synsets de cada POS. Estos porcentajes se muestran en la siguiente tabla.

POS	Total	Synsets sobre los que se ejecuto	
V	13845	1000	7.22 %
N	83090	1000	1.20 %
R	3621	1000	27.62 %
A	18156	1000	5.51 %

A continuación se muestra una tabla con los porcentajes de cubrimiento obtenidos según cada POS, para las dos versiones del WordNet: la original y la expandida por el método aquí presentado.

POS	WordNet Original	WordNet Expandido	Presentes en el Corpus		
V	75.35 %	1235	77.36 %	1268	1639
N	69.29 %	2780	70.09 %	2812	4012
R	32.79 %	121	62.87 %	232	369
A	51.00 %	840	54.34 %	895	1647

Podemos concluir que los adverbios fue la categoría que obtuvo los mejores resultados alcanzando un cubrimiento de casi el 63 % sobre el 33 % original. Esto se debe a dos razones, en primer lugar los adverbios era la categoría menos cubierta por el WordNet original, y además fue la POS a la que más se le aplicó la estrategia implementada, el cual se ejecutó sobre un poco más del 27 % de sus synsets.

Se observa además que para el resto de las POS también se logró mejorar el cubrimiento. Si bien es cierto que la mejora en estos casos fue relativamente pequeña (entre 1 % y 3 %), hay que considerar que en estos casos el método fue ejecutado para un porcentaje muy pequeño de los synsets presentes en WordNet.

CONCLUSIONES Y TRABAJO FUTURO

6.1 CONCLUSIONES

Como se mencionó en el capítulo 1, los principales objetivos del proyecto fueron diseñar e implementar distintas estrategias para, a partir del WordNet inglés, lograr enriquecer el WordNet español actual. Para lograr dicho objetivo se optó por seguir el modelo de expansión. En dicho modelo las variantes en inglés de los distintos synsets de WordNet son traducidas mediante la utilización de diferentes estrategias.

Como fuera mencionado en el capítulo 4 las estrategias elegidas fueron la utilización de una serie de selectores, a los cuales se les denominó “selectores iniciales” como paso inicial para luego aplicar lo que se llamó “fase de expansión”. Dicha fase tuvo como objetivo explotar las relaciones entre synsets para poder agregar variantes que los selectores iniciales no habían sido capaces de agregar. Se cree que estos objetivos fueron alcanzados y que los resultados fueron alentadores tal como fueron presentados en el capítulo 5.

Se diseñaron diferentes heurísticas para la transformación del WordNet inglés al español, que luego fueron implementadas y ejecutadas sobre una versión parcial del WordNet inglés y se pudo así expandir el WordNet español. La nueva versión parcial del WordNet fue utilizada luego en una tarea de semántica léxica, dicha tarea permitió medir cuantos lemas presentes en un corpus de estudio logramos agregar al WordNet español.

Observando los resultados obtenidos, se puede afirmar que la estrategia planteada es buena ya que alcanza una mejora considerable del WordNet español actual, cumpliendo así con las expectativas planteadas.

El punto más débil de la misma son los métodos y herramientas de traducción ya que ellos proporcionan los recursos sobre los cuales trabaja la heurística. Debido a esto condicionan enormemente los resultados finales. Con respecto a la estrategia implementada, los selectores iniciales ya son suficientes para realizar una mejora considerable del WordNet actual y su precisión de un 92 %, es casi inmejorable. En la fase 2, la precisión alcanzó el 74 %, que si bien presenta un margen de mejora, el cual se menciona en la sección siguiente, creemos que los resultados son satisfactorios.

6.2 TRABAJO FUTURO

Durante el desarrollo del enriquecimiento del WordNet español se encontraron varios posibles puntos de mejoras de la estrategia, que quedaron fuera del alcance del proyecto por razones de tiempo y recursos. En primer lugar se deberían mejorar los métodos de traducción utilizando más recursos que no solo enriquecerían las traducciones sino que además permitirían desarrollar un sistema de votación el cual permitiría tener mayor certeza al momento de la elección de las traducciones candidatas (antes de la desambiguación). También se podrían mejorar las traducciones de los lemas descartando aquellas en las que el par (lema, traducción) no coincida por ejemplo en género. Para ejemplificar esto se puede observar el synset "spa-30-10129825-n" cuya glosa es "mujer joven". Para el lema "girl" correspondiente al synset en inglés, se obtuvo que una traducción posible era "chico" la cual luego fue seleccionada por la fase 2 para agregar al WordNet construido. Este es un claro ejemplo en el cual el lema original en inglés y la traducción obtenida no coinciden en género.

Otra limitación encontrada al momento de implementar el método se dio en el selector de adverbios derivados. Este método, que como fue descrito en el capítulo 4, consistía en construir adverbios a partir de adjetivos, se valía fuertemente de un corpus contra el cual se verificaba la existencia del adverbio generado. Al tener un corpus de validación pequeño (3715 adverbios diferentes) lo que sucedía es que podíamos estar descartando muchos adverbios correctos pero que no se encontraban dentro de la lista de adverbios del corpus, es decir se podía generar un gran número de falsos positivos. Este inconveniente sería fácilmente superado contando con una lista más extensas de adverbios, la cual se podría obtener o bien de un corpus más grande o incluso uniendo diferentes corpus.

En tercer lugar la fase 2 de la estrategia podría ser mejorada dándole una mayor preponderancia a ciertas relaciones semánticas, para esto se debería estudiar qué relaciones son más "fuertes" según la categoría gramatical. Esta "fortaleza" se debe a que en la fase 2 la estrategia se basa en contar las ocurrencias de pares de palabras en una misma oración, dentro de un corpus. Los pares de palabras están formados por diferentes candidatos de un mismo synset y los lemas de un synset asociado a través de las relaciones semánticas. Se da que algunas relaciones generan pares más frecuentes dentro de una oración que los generados por otras relaciones, por lo tanto naturalmente hay más contextos para validar la correctitud de un candidato, por lo decimos que estas relaciones son más "fuertes" que otras.

Finalmente sería de potencial interés mejorar los tiempos de ejecución del algoritmo. Como fue mencionado en el capítulo 4, la principal razón por la cual se decidió aplicar el método a mil synsets de casa POS se debió a los tiempos de ejecución obtenidos. La heurística planteada es totalmente paralelizable ya que la misma aplica el mismo procesamiento a cada synset y estos no son dependientes entre sí, estamos ante un

6.2 TRABAJO FUTURO

claro caso de *Single Instruction Multiple Data*¹. Este escenario permitiría paralelizar el procesamiento para poder aprovechar todos los recursos de cómputo disponibles en el ambiente en el que se lo ejecute (todos los procesadores de una computadora o todas las computadoras de un cluster).

¹ “Los repertorios SIMD consisten en instrucciones que aplican una misma operación sobre un conjunto más o menos grande de datos. Es una organización en donde una única unidad de control común despacha las instrucciones a diferentes unidades de procesamiento. Todas éstas reciben la misma instrucción, pero operan sobre diferentes conjuntos de datos. Es decir, la misma instrucción es ejecutada de manera sincronizada por todas las unidades de procesamiento.” (Foster, 1995)

REFERENCIAS

- Apertium*. (s.f.). <https://www.apertium.org/index.eng.html?dir=oci-cat#translation>. ([En línea; accedida 14-Mayo-2015])
- Devitt, A., y Vogel, C. (2004). The topology of wordnet: Some metrics. En *Proc. of the 2nd international wordnet conference (gwc)* (pp. 106–111).
- Dyvik, H. (2004). Translations as semantic mirrors: from parallel corpus to wordnet. *Language and computers*, 49(1), 311–326.
- Española, R. A. (s.f.). Sitio web. *Diccionario de la Lengua Española*.(22^a ed). Disponible en <http://buscon.rae.es/draeI/SrvltGUIBusUsual>. ([En línea; accedida 16-Mayo-2015])
- Estudios de lingüística del español*. (s.f.). <http://elies.rediris.es/elies2/index.htm>. ([En línea; accedida 14-Mayo-2015])
- Fellbaum, C. (1998). *Wordnet*. Wiley Online Library.
- Foster, I. (1995). *Designing and building parallel programs*. Addison Wesley Publishing Company.
- Freeling home page*. (s.f.). <http://nlp.lsi.upc.edu/freeling/>. ([En línea; accedida 14-Mayo-2015])
- Giménez, J., y Màrquez, L. (2005). Automatic translation of wordnet glosses.
- Grassi, M., Malcuori, M., Couto, J., Prada, J. J., y Wonsever, D. (2001). Corpus informatizado: textos del español del uruguay (corin). En *Spllt-2-second international workshop on spanish language processing and language technologies-jaén, españa*.
- Mcr mediawiki*. (s.f.). <http://adimen.si.ehu.es/web/MCR/>. ([En línea; accedida 14-Mayo-2015])
- Microsoft translator*. (s.f.). <http://datamarket.azure.com/dataset/bing/microsofttranslator>. ([En línea; accedida 16-Mayo-2015])
- Mouton, C., y de Chalendar, G. (2010). Jaws: Just another wordnet subset. *Proc. of TALN'10*.
- Nltk 3.0 documentation*. (s.f.). <http://www.nltk.org/api/nltk.tag.html>. ([En línea; accedida 14-Mayo-2015])
- Oliver, A. (2014). Wn-toolkit: Automatic generation of wordnets following the expand model. *Proceedings of the 7th Global WordNetConference, Tartu, Estonia*.
- Oliver, A., y Climent, S. (2011). Construcción de los wordnets 3.0 para castellano y catalán mediante traducción automática de corpus anotados semánticamente. *Procesamiento del Lenguaje Natural*, 47(0).
- Oliver, A., y Climent, S. (2012). Parallel corpora for wordnet construction: Machine translation vs. automatic sense tagging. , 110-121.
- Oliver, A., Climent, S., y Contreras, M. (2012). Revisión de técnicas para la construcción de wordnets mediante la estrategia de expansión. *Procesamiento del lenguaje natural*, 49, 119–128.

Referencias

- Pradet, Q., de Chalendar, G., y Desormeaux, J. B. (2014). Wonef, an improved, expanded and evaluated automatic french translation of wordnet.
- Vossen, P. (1998). Introduction to eurowordnet. En *Eurowordnet: A multilingual database with lexical semantic networks* (pp. 1–17). Springer.
- Web mcr interface meaning. (s.f.). <http://adimen.si.ehu.es/cgi-bin/wei/public/wei.consult.perl>. ([En línea; accedida 16-Mayo-2015])
- Wikimedia downloads. (s.f.). <http://dumps.wikimedia.org/backup-index.html>. ([En línea; accedida 14-Mayo-2015])
- Wiktionary. (s.f.). <https://www.wiktionary.org/>. ([En línea; accedida 16-Mayo-2015])
- Wn-mcr-transform. (s.f.). <https://github.com/pln-fing-udelar/wn-mcr-transform>. ([En línea; accedida 14-Mayo-2015])
- Wordnet. (s.f.). <http://wordnet.princeton.edu/>. ([En línea; accedida 14-Mayo-2015])
- Yablonsky, S. (2010). English-russian wordnet for multilingual mappings. En *Proceedings of 2010 workshop on cross-cultural and cross-lingual aspects of the semantic web*.

GLOSARIO

CORPUS Conjunto de datos, textos u otros materiales sobre determinada materia que pueden servir de base para una investigación o trabajo

POS Part of Speech, categorías gramaticales de las palabras del idioma

SELECTOR Mecanismo que al aplicarlo a un synset inglés se encargará de elegir la o las traducciones que tendrá el synset español generado a partir del original inglés

SYNSET Conjunto de sinónimos en los cuales se agrupan los lemas en WordNet

MONOSEMIA Fenómeno del lenguaje que consiste en que una palabra tiene un solo significado.

GLOSA Breve definición de un concepto

LEMA Palabra perteneciente al un synset

API Application Programming Interface, es un conjunto de rutinas protocolos y herramientas para la construcción de aplicaciones de software



EVALUACIÓN MANUAL DE LOS RESULTADOS

A.1 EVALUACIÓN MANUAL DE LOS RESULTADOS DE LA FASE 1

En este anexo se muestra la evaluación que se realizó para ambas etapas de la estrategia. En particular, para la primera fase se muestran por separado las evaluaciones de cada uno de los selectores.

A.1.1 Selector de traducción singular

IdLema	Apertium	Wiktionary	Bing	C	I
enigmatic#eng-30-00102930-a	enigmático#	enigmático#	enigmático#	1	
featured#eng-30-00235429-a		destacado#		1	
revised#eng-30-00353101-a	revisado#			1	
general#eng-30-00487198-a	general#	general#		1	
desperate#eng-30-00713351-a	desesperado#	desesperado#		1	
theatrical#eng-30-00796047-a	teatral#	teatral#	teatral#	1	
fresh#eng-30-01067694-a	fresco#		fresco#	1	
frigid#eng-30-01258264-a		frígido#	frígido#	1	
frozen#eng-30-01258264-a		congelado#		1	
glacial#eng-30-01258264-a	glacial#			1	
icy#eng-30-01258264-a		helado#		1	
wintry#eng-30-01258264-a	invernal#	invernal#	invernal#	1	
ignorant#eng-30-01308894-a	ignorante#	ignorante#	ignorante#	1	
unknowing#eng-30-01308894-a				1	
unknowledgeable#eng-30-01308894-a				1	
unwitting#eng-30-01308894-a			involuntario#	1	
trenchant#eng-30-01336371-a			mordaz#	1	
imperial#eng-30-01591394-a	imperial#			1	
majestic#eng-30-01591394-a		majestuoso#		1	
purple#eng-30-01591394-a	púrpura#morado#	morado#		1	
regal#eng-30-01591394-a	regio#	regio#real#		1	
royal#eng-30-01591394-a	real#	real#		1	
politic#eng-30-01813400-a			politico#	1	
dead#eng-30-01872265-a	muerto#	muerto#		1	
idle#eng-30-01872265-a		desocupado#		1	
cerebral#eng-30-01927455-a	cerebral#	cerebral#		1	
intellectual#eng-30-01927455-a	intelectual#		intelectual#	1	

A.1 EVALUACIÓN MANUAL DE LOS RESULTADOS DE LA FASE 1

IdLema	Apertium	Wiktionary	Bing	C	I
correlated#eng-30-01974298-a	correlativo#			1	
correlative#eng-30-01974298-a	correlativo#		correlativo#	1	
antisocial#eng-30-02258972-a	antisocial#	antisocial#		1	
piquant#eng-30-02306874-a		picante#	picante#	1	
salty#eng-30-02306874-a	salobre#salado#	salado#		1	
compact#eng-30-02387413-a	compacto#			1	
heavysset#eng-30-02387413-a			corpulento#	1	
stocky#eng-30-02387413-a	achaparrado#		robusto#	1	
thick#eng-30-02387413-a	grueso#	grueso#	grueso#	1	
thickset#eng-30-02387413-a			rechoncho#	1	
tolerant#eng-30-02436341-a		tolerante#	tolerante#	1	
binary#eng-30-02664700-a	binario#	binario#	binario#	1	
cosmologic#eng-30-02703001-a			cosmológico#	1	
cosmological#eng-30-02703001-a	cosmológico#	cosmológico#	cosmológico#	1	
generic#eng-30-02734016-a	genérico#	genérico#	genérico#	1	
legislative#eng-30-02829565-a	legislativo#	legislativo#	legislativo#	1	
rotary#eng-30-02998561-a	rotativo#			1	
processional#eng-30-03103520-a			procesional#	1	

A.1 EVALUACIÓN MANUAL DE LOS RESULTADOS DE LA FASE 1

IdLema	Apertium	Wiktionary	Bing	C	I
error#eng-30-00072068-n	error#	error#		1	
inflation#eng-30-00367280-n	inflación#	inflación#	inflación#	1	
gastronomy#eng-30-00936296-n	gastronomía#	gastronomía#	gastronomía#	1	
procedure#eng-30-01023820-n	procedimiento#	procedimiento#	procedimiento#	1	
process#eng-30-01023820-n	proceso#		proceso#	1	
altar#eng-30-02699629-n	altar#	altar#		1	
cat#eng-30-02983507-n	gato#		gato#	1	
defence#eng-30-03171356-n	defensa#	defensa#	defensa#	1	
defense#eng-30-03171356-n		defensa#	defensa#	1	
product#eng-30-04007894-n	producto#	producto#	producto#	1	
production#eng-30-04007894-n	producción#	producción#	producción#	1	
vigilance#eng-30-04664778-n		vigilancia#	vigilancia#	1	
watchfulness#eng-30-04664778-n		vigilia#	vigilancia#	1	
nonconformity#eng-30-04801030-n			inconformidad#	1	
continent#eng-30-04884177-n			continencia#	1	
parsimony#eng-30-04893525-n	parsimonia#	parsimonia#	parsimonia#	1	
thrift#eng-30-04893525-n		parsimonia#parquedad# economía#ahorro#	mano#	1	
slovenliness#eng-30-04896878-n			dejadez#	1	
experiment#eng-30-05798043-n	experimento#	experimento#	experimento#	1	
experimentation#eng-30-05798043-n	experimentación#	experimentación#	experimentación#	1	
algorism#eng-30-06004599-n			algoritmo#	1	
subscription#eng-30-07179820-n	suscripción#		suscripción#	1	
languor#eng-30-07484109-n		languor#languidez#	languidez#	1	
lassitude#eng-30-07484109-n			lasitud#	1	
listlessness#eng-30-07484109-n			apatía#	1	
cayenne#eng-30-07721942-n			cayena#	1	
producer#eng-30-10480018-n	productor#	productor#		1	
disseminator#eng-30-10483138-n			difusor#	1	
propagator#eng-30-10483138-n			propagador#	1	
bum#eng-30-10539715-n			vagabundo#	1	
crumb#eng-30-10539715-n	migaja#miga#	miga#		1	
puke#eng-30-10539715-n			vómito#	1	
rat#eng-30-10539715-n	rata#		rata#	1	
skunk#eng-30-10539715-n			zorro#	1	
stinker#eng-30-10539715-n			canalla#	1	
bracken#eng-30-13190060-n		helecho#		1	
atrophy#eng-30-14365741-n	atrofia#	atrofia#		1	
dehydration#eng-30-14536831-n	deshidratación#	deshidratación#	deshidratación#	1	
desiccation#eng-30-14536831-n	desección#		desección#	1	
log#eng-30-14942762-n	registro#		registro#	1	

A.1 EVALUACIÓN MANUAL DE LOS RESULTADOS DE LA FASE 1

IdLema	Apertium	Wiktionary	Bing	C	I
automatically#eng-30-00005567-r	automáticamente#	automáticamente#	automáticamente#	1	
ahead#eng-30-00075442-r	de frente#adelante#	enfrente de#	adelante#	1	
forward#eng-30-00075442-r	adelante#		adelante#	1	
highly#eng-30-00089267-r	altamente#	altamente#	altamente#	1	
up#eng-30-00096921-r	arriba#	arriba#	arriba#	1	
more#eng-30-00099341-r	más#		más#	1	
most#eng-30-00111609-r	más#			1	
together#eng-30-001116510-r		junto#		1	
morally#eng-30-00134203-r	moralmente#	moralmente#	moralmente#	1	
gracelessly#eng-30-00179928-r			estrepitosamente#		1
accurately#eng-30-00204523-r	con exactitud#			1	
unreasonably#eng-30-00216100-r			irrazonablemente#		1
broadly#eng-30-00221583-r	en términos generales#			1	
generally#eng-30-00221583-r	por lo general# generalmente#	por lo general# generalmente#	generalmente#	1	
loosely#eng-30-00221583-r			libremente#	1	
roundly#eng-30-00228815-r	rotundamente#		rotundamente#	1	
aslant#eng-30-00274710-r			oblicuamente#	1	
athwart#eng-30-00274710-r		de lado a lado#		1	
obliquely#eng-30-00274710-r			oblicuamente#	1	
bureaucratically#eng-30-00280972-r			burocráticamente#	1	
artfully#eng-30-00293926-r			ingeniosamente#	1	
craftily#eng-30-00293926-r			astutamente#	1	
cunningly#eng-30-00293926-r			astutamente#	1	
slyly#eng-30-00293926-r			astutamente#	1	
densely#eng-30-00299753-r	densamente#		densamente#	1	
thickly#eng-30-00299753-r	apelmazadamente#			1	
autocratically#eng-30-00311430-r			autocráticamente#	1	
dictatorially#eng-30-00311430-r			dictatorialmente#	1	
magisterially#eng-30-00311430-r			magistralmente#	1	
mundanely#eng-30-00404879-r			prosaicamente#	1	
providentially#eng-30-00434213-r			providencialmente#	1	
sideways#eng-30-00454031-r	de lado#			1	
sidewise#eng-30-00454031-r			lateralmente#	1	
enthusiastically#eng-30-00456484-r		con entusiasmo#		1	
under#eng-30-00486296-r	debajo#			1	
softly#eng-30-00503945-r	suavemente#	suavemente#	suavemente#	1	
first#eng-30-00506928-r	primero#			1	

A.1 EVALUACIÓN MANUAL DE LOS RESULTADOS DE LA FASE 1

IdLema	Apertium	Wiktionary	Bing	C	I
dehydrate#eng-30-00211108-v	deshidratar#		deshidratar#	1	
desiccate#eng-30-00211108-v			desecar#	1	
cramp#eng-30-00236289-v		acalambrar#		1	
hamper#eng-30-00236289-v	dificultar#			1	
strangle#eng-30-00236289-v		estrangular#	estrangular#	1	
mature#eng-30-00249969-v	madurar#	madurar#		1	
ripen#eng-30-00249969-v		madurar#colorear#	madurar#	1	
freeze#eng-30-00363493-v	congelar#		congelar#	1	
suspend#eng-30-00363493-v	suspender#		suspender#	1	
deform#eng-30-00477665-v	deformar#	deformar#	deformar#	1	
find#eng-30-00522751-v	hallar#encontrar#	hallar#encontrar#	encontrar#	1	
get#eng-30-00522751-v	recibir#llegar#conseguir# coger#alcanzar#		obtener#	1	
incur#eng-30-00522751-v	incurrir#	incurrir#	incurrir#	1	
obtain#eng-30-00522751-v	obtener#agenciar#	obtener#	obtener#	1	
receive#eng-30-00522751-v	recibir#cobijar#acoger#	recibir#	recibir#	1	
commemorate#eng-30-00612042-v			conmemorar#	1	
immortalise#eng-30-00612042-v		inmortalizar#	inmortalizar#	1	
immortalize#eng-30-00612042-v			inmortalizar#	1	
memorialise#eng-30-00612042-v			conmemorar#	1	
memorialize#eng-30-00612042-v			conmemorar#	1	
record#eng-30-00612042-v	grabar#			1	
construct#eng-30-00634090-v	construir#	construir#	construir#	1	
reconstruct#eng-30-00634090-v	reconstruir#	reconstruir#	reconstruir#	1	
retrace#eng-30-00634090-v			desandar#		1
contemplate#eng-30-00704388-v	contemplar#	contemplar#	contemplar#	1	
meditate#eng-30-00704388-v		meditar#	meditar#	1	
study#eng-30-00704388-v	estudiar#	estudiar#		1	
align#eng-30-00733632-v	alinear#	alinear#	alinear#	1	
array#eng-30-00733632-v		arreglar#		1	
affect#eng-30-00838043-v	afectar#		afectar#	1	
dissemble#eng-30-00838043-v	disimular#	disimular#	desarmar#	1	
feign#eng-30-00838043-v	fingir#	fingir#	fingir#	1	
pretend#eng-30-00838043-v	pretender#fingir#	fingir#	fingir#	1	
sham#eng-30-00838043-v		fingir#		1	
vouch#eng-30-01015104-v			atestiguar#	1	
outpoint#eng-30-01101103-v			vencer#	1	
outscore#eng-30-01101103-v			superar#	1	
exceed#eng-30-01105639-v	superar#sobrepasar# exceder#	sobrepasar#	exceder#	1	
outdo#eng-30-01105639-v		sobrepasar#exceder#	superar#	1	
outmatch#eng-30-01105639-v			superar#	1	
outperform#eng-30-01105639-v			superar#	1	
outstrip#eng-30-01105639-v			superar#	1	
surmount#eng-30-01105639-v			superar#	1	
surpass#eng-30-01105639-v	superar#	sobrepasar#	superar#	1	
chain#eng-30-01288636-v	encadenar#			1	
berth#eng-30-01305361-v		ataque#		1	
liquidate#eng-30-01327301-v			liquidar#	1	
neutralise#eng-30-01327301-v	neutralizar#	neutralizar#	neutralizar#	1	
neutralize#eng-30-01327301-v			neutralizar#	1	
waste#eng-30-01327301-v	malgastar#desperdiciar#			1	
press#eng-30-01447632-v	pulsar#			1	
push#eng-30-01447632-v	pulsar#empujar#apechugar#			1	
care#eng-30-01766748-v	preocuparse#			1	
worry#eng-30-01766748-v	preocuparse#	preocupar#inquietar#	preocupar#	1	
dispatch#eng-30-01955127-v		despachar#		1	
sight#eng-30-02199307-v		ver#		1	
endure#eng-30-02618149-v	soportar#	perdurar#aguantar#	soportar#	1	
go#eng-30-02618149-v	rodar#ir#		ir#	1	
last#eng-30-02618149-v	durar#			1	
live#eng-30-02618149-v	vivir#		vivir#	1	
survive#eng-30-02618149-v	sobrevivir#	sobrevivir#	sobrevivir#	1	
wrangle#eng-30-02652376-v			disputar#	1	

A.1 EVALUACIÓN MANUAL DE LOS RESULTADOS DE LA FASE 1

A.1.2 Selector de factorización

IdLema	Apertium	Wiktionary	Bing	Factorización	C	I
capitalise#eng-30-00162167-v	capitalizar#		capitalizar#	capitalizar#	1	
capitalize#eng-30-00162167-v			capitalizar#	capitalizar#		
vitalise#eng-30-00164816-v			vitalizar#	vitalizar#	1	
vitalize#eng-30-00164816-v			vitalizar#	vitalizar#		
fortify#eng-30-00220869-v	fortalecer#	fortalecer#	fortificar#	fortalecer#	1	
strengthen#eng-30-00220869-v	fortalecer#afianzar#	fortalecer#	fortalecer#	fortalecer#		
oxygenise#eng-30-00308269-v			oxigenar#	oxigenar#	1	
oxygenize#eng-30-00308269-v			oxigenar#	oxigenar#		
socialise#eng-30-00406053-v			socializar#	socializar#	1	
socialize#eng-30-00406053-v			socializar#	socializar#		
nationalise#eng-30-00411547-v	nacionalizar#		nacionalizar#	nacionalizar#	1	
nationalize#eng-30-00411547-v			nacionalizar#	nacionalizar#		
privatise#eng-30-00412048-v	privatizar#		privatizar#	privatizar#	1	
privatize#eng-30-00412048-v		privatizar#	privatizar#	privatizar#		
homologise#eng-30-00417482-v			homologar#	homologar#	1	
homologize#eng-30-00417482-v			homologar#	homologar#		
neutralise#eng-30-00469904-v	neutralizar#	neutralizar#	neutralizar#	neutralizar#	1	
neutralize#eng-30-00469904-v			neutralizar#	neutralizar#		
professionalise#eng-30-00583242-v			profesionalizar#	profesionalizar#	1	
professionalize#eng-30-00583242-v		profesionalizar#	profesionalizar#	profesionalizar#		
catalog#eng-30-00652622-v	catalogar#			catalogar#	1	
catalogue#eng-30-00652622-v		catalogar#		catalogar#		
belittle#eng-30-00845909-v			menospreciar#	menospreciar#	1	
disparage#eng-30-00845909-v		denigrar#	menospreciar#	menospreciar#		
acquaint#eng-30-00874175-v	familiarizar#	familiarizar#	familiarizar#	familiarizar#	1	
familiarise#eng-30-00874175-v	familiarizar#		familiarizar#	familiarizar#		
familiarize#eng-30-00874175-v		familiarizar#	familiarizar#	familiarizar#		
synchronise#eng-30-00960734-v	sincronizar#		sincronizar#	sincronizar#	1	
synchronize#eng-30-00960734-v		sincronizar#	sincronizar#	sincronizar#		
groan#eng-30-01045419-v	gemir#	gruñir#gemir#	gemir#	gemir#	1	
moan#eng-30-01045419-v		gemir#		gemir#		
drink#eng-30-01170052-v	beber#		beber#	beber#	1	
imbibe#eng-30-01170052-v			beber#	beber#		
care#eng-30-01766748-v	preocuparse#			preocuparse#	1	
worry#eng-30-01766748-v	preocuparse#	preocupar# inquietar#	preocupar#	preocuparse#		
negociate#eng-30-02050688-v			negociar#	negociar#	1	
negotiate#eng-30-02050688-v	negociar#		negociar#	negociar#		
orient#eng-30-02159197-v	orientar#	orientar#		orientar#	1	
orientate#eng-30-02159197-v			orientar#	orientar#		
acquire#eng-30-02210855-v	adquirir#	obtener#	adquirir#	obtener#	1	
get#eng-30-02210855-v	recibir#llegar #conseguir#coger #alcanzar#		obtener#	obtener#		
offer#eng-30-02297142-v	ofrecer#ofertar# deparar#brindar#		ofertar#	ofertar#	1	
proffer#eng-30-02297142-v			ofertar#	ofertar#		
internationalise#eng-30-02441686-v			internacionalizar#	internacionalizar#	1	
internationalize#eng-30-02441686-v		internacionalizar#	internacionalizar#	internacionalizar#		
terrorise#eng-30-02506361-v			aterrorizar#	aterrorizar#	1	
terrorize#eng-30-02506361-v			aterrorizar#	aterrorizar#		
solemnise#eng-30-02579247-v			solemnizar#	solemnizar#	1	
solemnize#eng-30-02579247-v			solemnizar#	solemnizar#		
imply#eng-30-02636132-v	implicar#	implicar#	implicar#	implicar#	1	
involve#eng-30-02636132-v	implicar# acarrear#	envolver#	implicar#	implicar#		

A.1 EVALUACIÓN MANUAL DE LOS RESULTADOS DE LA FASE 1

IdLema	Apertium	Wiktionary	Bing	Factorización	C	I
inverse#eng-30-00202095-a		inverso#	inverso#	inverso#	1	
reverse#eng-30-00202095-a	inverso#	en sentido contrario #al revés#	inverso#	inverso#		
brave#eng-30-00262792-a	valiente#atrevido#	valiente#	valiente#	valiente#	1	
courageous#eng-30-00262792-a	brioso#	valiente#valeroso #corajudo#	valiente#	valiente#		
incomplete#eng-30-00523978-a	incompleto#	incompleto#	incompleto#	incompleto#	1	
uncomplete#eng-30-00523978-a			incompleto#	incompleto#		
nonconducting#eng-30-00558232-a			conductor#	conductor#		1
nonconductive#eng-30-00558232-a			conductor#	conductor#		
conterminous#eng-30-00566961-a			contiguo#	contiguo#	1	
contiguous#eng-30-00566961-a	contiguo#	contiguo#	contiguo#	contiguo#		
endogamic#eng-30-00871816-a			endogámico#	endogámico#	1	
endogamous#eng-30-00871816-a			endogámico#	endogámico#		
cheap#eng-30-00934199-a	barato#		barato#	barato#	1	
inexpensive#eng-30-00934199-a	económico#	económico# barato#	barato#	barato#		
fair#eng-30-00956131-a	justo#			justo#	1	
just#eng-30-00956131-a	justo#			justo#		
eerie#eng-30-00968522-a		misterioso# extraño#	espeluznante#	espeluznante#	1	
eery#eng-30-00968522-a			espeluznante#	espeluznante#		
artless#eng-30-01309991-a	inocente#	cándido#	ingenuo#	ingenuo#inocente#	1	
ingenuous#eng-30-01309991-a	ingenuo#	inocente#	ingenuo#	ingenuo#inocente#		
illogical#eng-30-01430847-a	ilógico#	ilógico#	ilógico#	ilógico#	1	
unlogical#eng-30-01430847-a			ilógico#	ilógico#		
rare#eng-30-01678112-a	raro#		raro#	raro#	1	
uncommon#eng-30-01678112-a		raro#	infrecuente#	raro#		
travelable#eng-30-01725129-a			transitable#	transitable#	1	
traversable#eng-30-01725129-a			transitable#	transitable#		
overpowering#eng-30-01989446-a			abrumador#	abrumador#	1	
overwhelming#eng-30-01989446-a	aplastante# agobiante# abrumador#		abrumador#	abrumador#		
satiric#eng-30-02080110-a			satírico#	satírico#	1	
satirical#eng-30-02080110-a	satírico#	satírico#	satírico#	satírico#		
evidential#eng-30-02162733-a	probatorio#		probatorio#	probatorio#	1	
evidentiary#eng-30-02162733-a			probatorio#	probatorio#		
thermoset#eng-30-02410156-a			termoestable#	termoestable#	1	
thermosetting#eng-30-02410156-a	termoendurecible#		termoestable#	termoestable#		
insecure#eng-30-02524443-a		inseguro#	inseguro#	inseguro#	1	
unsafe#eng-30-02524443-a			inseguro#	inseguro#		
alchemic#eng-30-02610603-a	alquímico#		alquímico#	alquímico#	1	
alchemical#eng-30-02610603-a			alquímico#	alquímico#		
broadband#eng-30-02671038-a	de banda ancha#		ancho#	ancho#	1	
wideband#eng-30-02671038-a			ancho#	ancho#		
juridic#eng-30-02753724-a			jurídico#	jurídico#	1	
juridical#eng-30-02753724-a	jurídico#		jurídico#	jurídico#		
oligarchic#eng-30-02770742-a		oligárquico#	oligárquico#	oligárquico#	1	
oligarchical#eng-30-02770742-a	oligárquico#		oligárquico#	oligárquico#		
typographic#eng-30-02910248-a			tipográfico#	tipográfico#	1	
typographical#eng-30-02910248-a		tipográfico#	tipográfico#	tipográfico#		
insurreccional#eng-30-02919100-a			insurreccional#	insurreccional#	1	
insurrectionary#eng-30-02919100-a			insurreccional#	insurreccional#		
dialectic#eng-30-03023644-a	dialéctico#		dialéctico#	dialéctico#	1	
dialectical#eng-30-03023644-a			dialéctico#	dialéctico#		
involve#eng-30-02636132-v	implicar# acarrear#	envolver#	implicar#	implicar#		

A.1 EVALUACIÓN MANUAL DE LOS RESULTADOS DE LA FASE 1

IdLema	Apertium	Wiktionary	Bing	Factorización	C	I
beginning#eng-30-00235435-n	principio#comienzo#	principio# inicio# comienzo#	inicio#	inicio#	1	
commencement#eng-30-00235435-n		inicio# incoación#	inicio#	inicio#		
start#eng-30-00235435-n	salida# inicio# arranque#		inicio#	inicio#		
carpentry#eng-30-00607775-n	carpintería#	carpintería#	carpintería#	carpintería#	1	
woodwork#eng-30-00607775-n			carpintería#	carpintería#		
woodworking#eng-30-00607775-n			carpintería#	carpintería#		
realisation#eng-30-01119620-n	realización#		realización#	realización#	1	
realization#eng-30-01119620-n			realización#	realización#		
acidophil#eng-30-01349333-n			acidófilo#	acidófilo#	1	
acidophile#eng-30-01349333-n			acidófilo#	acidófilo#		
beading#eng-30-02815389-n			abalorio#	abalorio#	1	
beadwork#eng-30-02815389-n			abalorio#	abalorio#		
bangle#eng-30-02887970-n	pulsera#		brazalete#	brazalete# pulsera#	1	
bracelet#eng-30-02887970-n	brazalete#	pulsera# brazalete#	pulsera#	brazalete# pulsera#		
bradawl#eng-30-02888898-n			punzón#	punzón#	1	
pricker#eng-30-02888898-n			punzón#	punzón#		
carburetor#eng-30-02962200-n		carburador#	carburador#	carburador#	1	
carburettor#eng-30-02962200-n	carburador#		carburador#	carburador#		
colors#eng-30-03072056-n			color#	color#	1	
colours#eng-30-03072056-n			color#	color#		
defence#eng-30-03171356-n	defensa#	defensa#	defensa#	defensa#	1	
defense#eng-30-03171356-n		defensa#	defensa#	defensa#		
disc#eng-30-03208556-n	disco#	disco#	disco#	disco#	1	
disk#eng-30-03208556-n	disco#	disco#	disco#	disco#		
suppressor#eng-30-04361641-n			supresor#	supresor#	1	
suppressor#eng-30-04361641-n			supresor#	supresor#		
vigilance#eng-30-04664778-n		vigilancia#	vigilancia#	vigilancia#	1	
watchfulness#eng-30-04664778-n		vigilia#	vigilancia#	vigilancia#		
diffusion#eng-30-05088056-n	difusión#	difusión#	difusión#	difusión#	1	
dissemination#eng-30-05088056-n	diseminación#		difusión#	difusión#		
cartilage#eng-30-05288091-n	cartílago#	cartílago#	cartílago#	cartílago#	1	
gristle#eng-30-05288091-n		cartílago#	cartílago#	cartílago#		
leucorrhoea#eng-30-05417821-n			leucorrea#	leucorrea#	1	
leukorrhoea#eng-30-05417821-n			leucorrea#	leucorrea#		
divergence#eng-30-05863302-n	divergencia#		divergencia#	divergencia#	1	
divergency#eng-30-05863302-n			divergencia#	divergencia#		
epiphora#eng-30-07100113-n			epífora#	epífora#	1	
epistrophe#eng-30-07100113-n			epífora#	epífora#		
placidity#eng-30-07515169-n		placidez#	placidez#	placidez#	1	
placidness#eng-30-07515169-n			placidez#	placidez#		
indignation#eng-30-07517737-n	indignación#	indignación#	indignación#	indignación#	1	
outrage#eng-30-07517737-n	ultraje#	atrocidad#	indignación#	indignación#		
prawn#eng-30-07794159-n	langostino#	gamba#camarón#	gamba#	gamba#	1	
shrimp#eng-30-07794159-n	gamba#		camarón#	gamba#		
episcopacy#eng-30-07946969-n			episcopado#	episcopado#	1	
episcopate#eng-30-07946969-n			episcopado#	episcopado#		
inside#eng-30-08588294-n	interior#	interior#		interior#	1	
interior#eng-30-08588294-n	interior#	interior#		interior#		
aluminium#eng-30-14627820-n	aluminio#	aluminio#	aluminio#	aluminio#	1	
aluminum#eng-30-14627820-n	aluminio#		aluminio#	aluminio#		
fiber#eng-30-14866889-n			fibra#	fibra#	1	
fibre#eng-30-14866889-n	fibra#	fibra#	fibra#	fibra#		

A.1 EVALUACIÓN MANUAL DE LOS RESULTADOS DE LA FASE 1

IdLema	Apertium	Wiktionary	Bing	Factorización	C	I
badly# eng-30-00011516-r	mal#	malamente# mal#	mal#	mal#	1	
ill#eng-30-00011516-r		mal#		mal#		
poorly# eng-30-00011516-r	pobrememente# mal#		mal#	mal#		
always# eng-30-00019339-r	siempre#	siempre#	siempre#	siempre#	1	
ever# eng-30-00019339-r	nunca#	siempre#		siempre#		
furthermore# eng-30-00029367-r	además#	más# es más# además#	además#	además#	1	
moreover# eng-30-00029367-r	además#	adicionalmente# además#		además#		
earliest# eng-30-00034641-r			más#	más#		1
soonest# eng-30-00034641-r			más#	más#		
rarely# eng-30-00035385-r	raramente#	raramente# rara vez# poco frecuentemente#	raramente#	raramente#	1	
seldom# eng-30-00035385-r	raramente#	raramente# rara vez#	rara vez#	raramente#		
eventually# eng-30-00048138-r	finalmente#	finalmente# al final#		finalmente#	1	
finally# eng-30-00048138-r	por último# finalmente# al fin#	por fin#	finalmente#	finalmente#		
currently# eng-30-00048268-r	actualmente#	hoy#ahora# actualmente#	actualmente#	actualmente#	1	
presently# eng-30-00048268-r	actualmente#			actualmente#		
alright# eng-30-00053004-r	bien#		muy#	bien#	1	
fine# eng-30-00053004-r	bien#			bien#		
chiefly# eng-30-00073897-r	principalmente#	principalmente#	principalmente#	principalmente#	1	
mainly# eng-30-00073897-r	principalmente#	principalmente#	principalmente#	principalmente#		
primarily# eng-30-00073897-r	principalmente#	primariamente#	principalmente#	principalmente#		
principally# eng-30-00073897-r	principalmente#		principalmente#	principalmente#		
ahead# eng-30-00075442-r	de frente# adelante#	enfrente de#	adelante#	adelante#	1	
forward# eng-30-00075442-r	adelante#		adelante#	adelante#		
below# eng-30-00094396-r	debajo#abajo#	abajo#		abajo#	1	
downstairs# eng-30-00094396-r	abajo#		abajo#	abajo#		
inside# eng-30-00110815-r	dentro#adentro#	dentro#		dentro#	1	
within# eng-30-00110815-r	dentro#			dentro#		
transversally# eng-30-00137770-r			transversalmente#	transversalmente#	1	
transversely# eng-30-00137770-r			transversalmente#	transversalmente#		
publically# eng-30-00161932-r			públicamente#	públicamente#	1	
publicly# eng-30-00161932-r	públicamente#	públicamente#	públicamente#	públicamente#		

A.1 EVALUACIÓN MANUAL DE LOS RESULTADOS DE LA FASE 1

IdLema	Apertium	Wiktionary	Bing	Factorización	C	I
apart# eng-30-00180756-r	aparte#	separadamente# aparte#	aparte#	aparte#	1	
aside# eng-30-00180756-r	aparte#	aparte# a un lado#		aparte#		
sagely# eng-30-00201570-r			sabiamente#	sabiamente#	1	
wisely# eng-30-00201570-r	sabiamente# avivadamente#	sabiamente#	sabiamente#	sabiamente#		
lawfully# eng-30-00251820-r			legalmente#	legalmente#	1	
legally# eng-30-00251820-r	legalmente#	legalmente#	legalmente#	legalmente#		
inward# eng-30-00258549-r		adentro#	adentro#	adentro#		1
inwards# eng-30-00258549-r		adentro#		adentro#		
amiss# eng-30-00271998-r			mal#	mal#	1	
awry# eng-30-00271998-r			mal#	mal#		
inconsequentially# eng-30-00294782-r			inconsecuentemente#	inconsecuentemente#	1	
inconsequently# eng-30-00294782-r			inconsecuentemente#	inconsecuentemente#		
skilfully# eng-30-00455933-r	consejeramente#		hábilmente#	hábilmente#	1	
skillfully# eng-30-00455933-r			hábilmente#	hábilmente#		
sweet# eng-30-00472323-r		dulcemente#		dulcemente#	1	
sweetly# eng-30-00472323-r		dulcemente# abemoladamente#	dulcemente#	dulcemente#		
tauntingly# eng-30-00474487-r			burlonamente#	burlonamente#	1	
teasingly# eng-30-00474487-r			burlonamente#	burlonamente#		
best# eng-30-00509846-r	más#		mejor#	mejor#	1	
better# eng-30-00509846-r			mejor#	mejor#		
coincidentally# eng-30-00510629-r			coincidentemente#	coincidentemente#	1	
coincidentally# eng-30-00510629-r			coincidentemente#	coincidentemente#		

A.1 EVALUACIÓN MANUAL DE LOS RESULTADOS DE LA FASE 1

A.1.3 Selector de monoesmia

IdLema	Apertium	Wiktionary	Bing	C	I
league#eng-30-00369064-v			ligar#	1	
reintegrate#eng-30-00398349-v			reintegrar#	1	
hone#eng-30-00473572-v			afilar#	1	
perfect#eng-30-00473572-v	perfeccionar#			1	
spelunk#eng-30-00649887-v		hacer espeleología#	explorar#	1	
polemicise#eng-30-00774932-v			polemizar#	1	
polemicize#eng-30-00774932-v			polemizar#	1	
reschedule#eng-30-00795475-v			reprogramar#	1	
annul#eng-30-00799798-v	anular#	revocar# invalidar# anular#	anular#	1	
countermand#eng-30-00799798-v			contradecir#	1	
repeal#eng-30-00799798-v		derogar#		1	
rescind#eng-30-00799798-v		rescindir# cancelar# anular#	rescindir#	1	
reverse#eng-30-00799798-v	revocar#invertir#			1	
revoke#eng-30-00799798-v	revocar#		revocar#	1	
vacate#eng-30-00799798-v		desalojar#	desocupar#	1	
thank#eng-30-00892315-v	dar las gracias a#				1
groan#eng-30-01045419-v	gemir#	gruñir#gemir#	gemir#	1	
moan#eng-30-01045419-v		gemir#		1	
feel#eng-30-01210352-v	sentir#		sentir#	1	
palpate#eng-30-01210352-v			palpar#	1	
besmear#eng-30-01251828-v			estar#		1
dissatisfy#eng-30-01816645-v			descontentar#	1	
dishearten#eng-30-01819387-v			descorazonar#	1	
cavort#eng-30-01883716-v		retozar# juguetear#	retozar#	1	
frisk#eng-30-01883716-v		cachear#		1	
frolic#eng-30-01883716-v		retozar# juguetear#		1	
gambol#eng-30-01883716-v			preguntar#		1
rollick#eng-30-01883716-v			divertir#	1	
sport#eng-30-01883716-v		jugar#burlarse#		1	
scamper#eng-30-01902405-v		corretear#		1	
queue#eng-30-02036755-v		poner en fila#		1	
catch#eng-30-02189168-v	coger#	capturar#	capturar#	1	
overhear#eng-30-02189168-v			escuchar#	1	
offer#eng-30-02297142-v	ofrecer# ofertar# deparar# brindar#		ofertar#	1	
proffer#eng-30-02297142-v			ofertar#	1	
mortgage#eng-30-02351737-v	hipotecar#	hipotecar#	hipotecar#	1	
enrol#eng-30-02471327-v	matricular#		matricular#	1	
enroll#eng-30-02471327-v		enrolar#	inscribir#	1	
enter#eng-30-02471327-v	introducir#	entrar#	entrar#	1	
inscribe#eng-30-02471327-v	inscribir#	inscribir#	inscribir#	1	
recruit#eng-30-02471327-v		reclutar#	reclutar#	1	
disenfranchise#eng-30-02476846-v			privar#	1	
disfranchise#eng-30-02476846-v			afectar#	1	
inhere#eng-30-02653564-v		ser inherente#		1	
blight#eng-30-02771320-v	infestar#			1	
plague#eng-30-02771320-v		plagar# molestar# atormentar#	plagar#	1	

A.1 EVALUACIÓN MANUAL DE LOS RESULTADOS DE LA FASE 1

IdLema	Apertium	Wiktionary	Bing	C	I
bewitching#eng-30-00166753-a			embruajador#		1
captivating#eng-30-00166753-a		arrebatador#	cautivante#	1	
enchanting#eng-30-00166753-a		encantador#	encantador#	1	
enthraling#eng-30-00166753-a			fascinante#	1	
entrancing#eng-30-00166753-a			fascinante#	1	
fascinating#eng-30-00166753-a		fascinante#	fascinante#	1	
propitious#eng-30-00177186-a			propicio#	1	
etched#eng-30-00317310-a			fuerte#	1	
incised#eng-30-00317310-a			inciso#		1
incontinent#eng-30-00591953-a			incontinente#	1	
crumbly#eng-30-00708738-a	quebradizo#		desmenuzable#	1	
friable#eng-30-00708738-a	friable#	quebradizo#frágil# friable# desmenuzable#		1	
conjectural#eng-30-00861216-a			conjetural#	1	
divinatory#eng-30-00861216-a			adivinatorio#	1	
hypothetic#eng-30-00861216-a			hipotético#	1	
hypothetical#eng-30-00861216-a	hipotético#	hipotético#	hipotético#	1	
supposed#eng-30-00861216-a	supuesto#	supuesto#		1	
discriminatory#eng-30-00997394-a	discriminatorio#		discriminatorio#	1	
invidious#eng-30-00997394-a			ingrato#	1	
haploid#eng-30-01147433-a	haploide#	haploide#	haploide#	1	
monoploid#eng-30-01147433-a	monoploide#	monoploide#	monoploide#	1	
infective#eng-30-01168434-a			infeccioso#	1	
morbific#eng-30-01168434-a			morboso#	1	
pathogenic#eng-30-01168434-a	patógeno#	patógeno#	patógeno#	1	
raftered#eng-30-01210174-a			experto#	1	
deafening#eng-30-01453467-a	estrepitoso#	ensordecedor#	ensordecedor#	1	
thunderous#eng-30-01453467-a			estruendoso#	1	
disgraceful#eng-30-01549964-a			vergonzoso#	1	
scandalous#eng-30-01549964-a	escandaloso#	escandaloso#	escandaloso#	1	
shameful#eng-30-01549964-a	vergonzoso#	vergonzoso#	vergonzoso#	1	
shocking#eng-30-01549964-a	impresionante#		impactante#	1	
psychosomatic#eng-30-01585781-a	psicosomático#	psicosomático#	psicosomático#	1	
fatherly#eng-30-01734607-a			paternal#	1	
correlated#eng-30-01974298-a	correlativo#			1	
correlative#eng-30-01974298-a	correlativo#		correlativo#	1	
deferent#eng-30-01994180-a			deferente#	1	
deferential#eng-30-01994180-a	deferente#		deferente#	1	
whatever#eng-30-02267686-a		cualquier#		1	
whatsoever#eng-30-02267686-a			absoluto#	1	
convincible#eng-30-02361848-a			convencible#	1	
persuadable#eng-30-02361848-a			asequible#	1	
persuasive#eng-30-02361848-a			persuasibles#	1	
bittersweet#eng-30-02396354-a		agridulce#	agridulce#	1	
semisweet#eng-30-02396354-a			semidulce#	1	
deliberative#eng-30-02420215-a			deliberativo#	1	
alkalotic#eng-30-02612058-a			alcalóticos#	1	
atactic#eng-30-02646952-a			atáctico#	1	
ataxic#eng-30-02646952-a			atáxico#	1	
carcinomatous#eng-30-02680017-a			carcinomatoso#	1	
recombinant#eng-30-02787740-a			recombinante#	1	
sacrificial#eng-30-03113271-a		sacrificial#		1	

A.1 EVALUACIÓN MANUAL DE LOS RESULTADOS DE LA FASE 1

IdLema	Apertium	Wiktionary	Bing	C	I
jog#eng-30-00112997-n		trote cochiner#		1	
nudge#eng-30-00112997-n		pequeño empujón#		1	
piecwork#eng-30-00581517-n			trabajo#	1	
sensitisation#eng-30-00829170-n	sensibilización#		sensibilización#	1	
sensitization#eng-30-00829170-n			sensibilización#	1	
sensitizing#eng-30-00829170-n			sensibilización#		1
phytoplankton#eng-30-01383896-n	fitoplancton#		fitoplancton#	1	
ormer#eng-30-01943087-n			tuberculata#	1	
blunderbuss#eng-30-02856237-n		trabuco#	trabuco#	1	
dihydrostreptomycin#eng-30-03198028-n			dihidroestreptomicina#	1	
doodad#eng-30-03218545-n			adorno#	1	
doohickey#eng-30-03218545-n			chisme#		1
gimmick#eng-30-03218545-n		truco#	truco#	1	
gizmo#eng-30-03218545-n		trasto# cuestión# coso#cosa# chisme#aparato#		1	
thingamabob#eng-30-03218545-n		chingadera#	cosa#	1	
thingamajig#eng-30-03218545-n			chisme#	1	
whatchamacallit#eng-30-03218545-n		vaina#coso# chunche# chisme#chingadera# caribbean#		1	
whatsis#eng-30-03218545-n			chucho#	1	
doyley#eng-30-03219010-n			servilleta#	1	
headband#eng-30-03502042-n	cinta#	cintillo#	diadema#	1	
lumberyard#eng-30-03697265-n			aserradero#	1	
rerebrace#eng-30-04077594-n			guardabrazo#	1	
playlist#eng-30-06494016-n		lista de reproducción#	reproducción#	1	
argot#eng-30-07157273-n		argot#		1	
jargon#eng-30-07157273-n	jerga#		jerga#	1	
lingo#eng-30-07157273-n			jerga#	1	
slang#eng-30-07157273-n	jerga#argot#		argot#	1	
vernacular#eng-30-07157273-n	lengua vernacular#	vernáculo#		1	
drogue#eng-30-07257582-n			paracaídas#	1	
sock#eng-30-07257582-n	calcetín#		calcetín#	1	
windsock#eng-30-07257582-n			manga#	1	
atmospherics#eng-30-07431683-n			condición#		1
static#eng-30-07431683-n		estática#		1	
lemonade#eng-30-07926250-n	limonada#	limonada#	limonada#	1	
pyre#eng-30-07963494-n		pira#		1	
neurasthenic#eng-30-10353813-n			neurastenia#	1	
revolutionary#eng-30-10527334-n	revolucionario#			1	
subversive#eng-30-10527334-n	subversor#			1	
moneylender#eng-30-10742384-n		prestamista#		1	
usurer#eng-30-10742384-n		usurero#usurera#	usurero#	1	
mesquit#eng-30-11765277-n			mezquita#	1	
mesquite#eng-30-11765277-n		mezquite#		1	
cucurbit#eng-30-12157677-n			cucurbitácea#	1	
humification#eng-30-13494192-n	humificación#		humificación#	1	
torticollis#eng-30-14557719-n			tortícolis#	1	
wryneck#eng-30-14557719-n		torcecuello#	torcecuello#	1	

A.1 EVALUACIÓN MANUAL DE LOS RESULTADOS DE LA FASE 1

IdLema	Apertium	Wiktionary	Bing	C	I
straightforwardly#eng-30-00051017-r			directamente#	1	
openly#eng-30-00053394-r	abiertamente#	abiertamente#	abiertamente#	1	
abruptly#eng-30-00061528-r	abruptamente#	abruptamente#	abruptamente#	1	
suddenly#eng-30-00061528-r	de repente#de pronto# de golpe y porrazo# de buenas a primeras# arrebataadamente#	súbitamente# repentinamente# de repente# de pronto#	de repente#	1	
slower#eng-30-00086621-r			más#		1
first#eng-30-00102736-r	primero#			1	
firstly#eng-30-00102736-r	primeramente# en primer lugar#	primeramente#		1	
foremost#eng-30-00102736-r			primero#	1	
initially#eng-30-00103194-r	inicialmente#	inicialmente#	inicialmente#	1	
afresh#eng-30-00112843-r		otra vez#de nuevo#	desde cero#	1	
anew#eng-30-00112843-r	de nuevo#	nuevamente# de nuevo#		1	
academically#eng-30-00121286-r	académicamente#	académicamente#	académicamente#	1	
unconstitutionally#eng-30-00122273-r			inconstitucionalmente#	1	
genetically#eng-30-00123924-r	genéticamente#	genéticamente#	genéticamente#	1	
atypically#eng-30-00128290-r			anormalmente#	1	
religiously#eng-30-00178460-r		religiosamente#		1	
sacredly#eng-30-00178460-r			sagradamente#	1	
apart#eng-30-00180944-r	aparte#	separadamente# aparte#	aparte#	1	
asunder#eng-30-00180944-r		en pedazos#en dos#		1	
exaggeratedly#eng-30-00189514-r			exageradamente#	1	
admirably#eng-30-00218886-r	admirablemente#	admirablemente#	admirablemente#	1	
laudably#eng-30-00218886-r		loablemente#		1	
knowingly#eng-30-00237636-r	deliberadamente# a sabiendas#	sabiamente# conscientemente# a sabiendas#	a sabiendas#	1	
wittingly#eng-30-00237636-r			deliberadamente#	1	
hereby#eng-30-00255614-r	de ahora en adelante#	por la presente# por este medio# por este#		1	
adorably#eng-30-00265298-r			adorablemente#	1	
endearingly#eng-30-00265298-r			cariñosamente#	1	
bounteously#eng-30-00279867-r			generosamente#	1	
bountifully#eng-30-00279867-r			generosamente#	1	
functionally#eng-30-00345659-r	funcionalmente#		funcionalmente#	1	
vertically#eng-30-00358516-r	verticalmente#	verticalmente#	verticalmente#	1	
sidesplittingly#eng-30-00385216-r			cómicamente#	1	
majestically#eng-30-00394462-r	majestuosamente#	majestuosamente#	majestuosamente#	1	
venomously#eng-30-00426761-r			venenosamente#	1	
transiently#eng-30-00480195-r			transitoriamente#	1	

A.1 EVALUACIÓN MANUAL DE LOS RESULTADOS DE LA FASE 1

A.1.4 Selector de adverbios derivados

ILI	Glosa	Adverbio derivado	Has Derived	C	I
ili-30-00004288-r	in a harmful manner	negativamente#	negative	1	
ili-30-00004567-r	in an unkind manner or with unkindness	cruelmente#	unkind	1	
ili-30-00005343-r	as can be shown by argument	discutiblemente#	arguable	1	
ili-30-00005567-r	in a reflex manner	reflexivamente# automáticamente#	automatic	1	
			reflex		
			reflexive		
ili-30-00006259-r	in a significant manner	significativamente#	significant	1	
ili-30-00006858-r	in a stylish manner	enérgicamente# gallardamente# alegremente#	chic	1	1
			smart		
			voguish		
			a_la_mode		
			in_style		
			in_vogue		
			latest		
			modish		
			dapper		
			dashing		
			jaunty		
			natty		
			raffish		
			rakish		
			snappy		
			spiffy		
			spruce		
ili-30-00011516-r	(‘ill’ is often used as a combining form) in a poor or improper or unsatisfactory manner; not well	miserablemente# lastimeramente# pobremente#	hapless	1	
			miserable		
			misfortunate		
			pathetic		
			piteous		
			pitiable		
			pitiful		
			poor		
			wretched		
ili-30-00013793-r	in a manner affording benefit or advantage	ventajosamente# provechosamente#	advantageous	1	
ili-30-00014285-r	to a great extent or degree	sustancialmente# substancialmente# jugosamente#	considerable	1	
			significant		
			substantial		
ili-30-00015007-r	with great or especially intimate knowledge	íntimamente#	intimate	1	
ili-30-00015860-r	in the manner of prosperous people	cómodamente# ricamente# fácilmente# confortablem	comfortable	1	
			easy		
			prosperous		
			well-fixed		
			well-heeled		
			well-off		
			well-situated		
			well-to-do		

A.1 EVALUACIÓN MANUAL DE LOS RESULTADOS DE LA FASE 1

ILI	Glosa	Adverbio derivado	Has.Derived	C	I
ili-30-00016458-r	with great intensity (‘bad’ is a nonstandard variant for ‘badly’)	grandemente# malamente#	bad	1	
			big		
ili-30-00016678-r	in a disobedient or naughty way	traviesamente#	naughty	1	
ili-30-00016950-r	evilly or wickedly	malamente#	bad		
ili-30-00023493-r	in a conformable manner	conformemente#	conformable	1	
ili-30-00023721-r	in an unconventional manner	convencionalmente#	unconventional		1
ili-30-00028923-r	with indifference or without dejection	livianamente# ligeramente# claramente#	light	1	
ili-30-00033562-r	to a moderate degree	suavemente# levemente#	mild	1	
ili-30-00034862-r	in a merry sportive manner	deportivamente# alegremente#	coltish	1	
			frolicky		
			frolicsome		
			rollicking		
			sportive		
ili-30-00035385-r	not often	raramente#	rare	1	
ili-30-00036762-r	with respect to its inherent nature	intrínsecamente#	intrinsic	1	
			intrinsic		
ili-30-00046863-r	extremely	extraordinariamente# desproporcionadamente#	excessive	1	
			inordinate		
			undue		
			unreasonable		
			extraordinary		
			over-the-top		
			sinful		
ili-30-00048138-r	after an unspecified period of time or an especially long delay	eventualmente#	eventual	1	
ili-30-00048268-r	at this time or period; now	presentemente# actualmente#	present	1	
			current		
ili-30-00050075-r	in a shrill voice	agudamente# afiladamente# filosamente#	sharp	1	
			shrill		

A.2 EVALUACIÓN MANUAL DE LOS RESULTADOS DE LA FASE 2

A.2 EVALUACIÓN MANUAL DE LOS RESULTADOS DE LA FASE 2

Synset	Variante	C	I	Observación
spa-30-00662589-v	ver	1		
spa-30-01654628-v	hacer	1		
spa-30-02030764-v	romper	1		
spa-30-02725286-v	estancar	1		
spa-30-00589904-v	llegar		1	
spa-30-00434374-v	morir	1		
spa-30-01528069-v	alojar		1	
spa-30-02587532-v	dirigir		1	
spa-30-00075021-v	cansarse	1		
spa-30-01759021-v	volar	1		
spa-30-00358431-v	pasar		1	el lema en ingles es pass pero el synset refiere a morir
spa-30-00387680-v	retroceder	1		
spa-30-00891038-v	hacer		1	el lema en ingles es make el synset se refiere a "dar exito"
spa-30-02326355-v	explorar	1		
spa-30-00393398-v	afilar		1	el lema es sharpen el synset hace referencia a notas musicales debería ser . ^a gudizar"
spa-30-01104624-v	hacer trampa	1		
spa-30-00522751-v	incurrir en	1		
spa-30-02339171-v	permitir	1		
spa-30-02086458-v	sacar	1		
spa-30-02370131-v	utilizar		1	el lema es use el synset se refiere a un habito en pasado (She used to call her mother every week)
spa-30-00754156-v	llamar	1		es llamar en el sentido de ordenar llamar a una huelga general
spa-30-01510082-v	bordar		1	debería ser algo como "tirar por la borda"
spa-30-02534761-v	cortejar	1		
spa-30-01593937-v	apretar	1		el lema es squeeze
spa-30-01560034-v	entreverar	1		

Synset	Variante	C	I	Observación
spa-30-00100082-r	temprano	1		
spa-30-00013793-r	bien	1		

A.2 EVALUACIÓN MANUAL DE LOS RESULTADOS DE LA FASE 2

Synset	Variante	C	I	Observación
spa-30-08663354-n	cima	1		
spa-30-04893525-n	parquedad	1		
spa-30-07551052-n	genio	1		
spa-30-05603160-n	mandíbula	1		
spa-30-04981658-n	cerco		1	
spa-30-00458641-n	traba		1	El synset refiere al juego de "la pasadita"
spa-30-00806902-n	regulación	1		
spa-30-11786843-n	macal	1		
spa-30-08352303-n	oficina	1		
spa-30-08544593-n	esquina		1	El synset hace referencia a "un lugar remoto" (in many corners of the world)
spa-30-03147901-n	taza		1	El synset hace referencia a copa como trofeo
spa-30-09776346-n	afiliado	1		
spa-30-10129825-n	chico		1	Horrible!! El lema es girl!!
spa-30-02849051-n	espacio		1	
spa-30-03596787-n	joya	1		
spa-30-10196965-n	soñador	1		
spa-30-00737070-n	incorrección	1		
spa-30-02396014-n	verraco	1		
spa-30-13993842-n	juego		1	El synset refiere a dejar libre
spa-30-08594286-n	línea	1		
spa-30-05703429-n	comunicación		1	El synset refiere a el acto de notar o prestar atención (el lema es notice)
spa-30-08653873-n	cero	1		
spa-30-07883031-n	relleno	1		
spa-30-14813182-n	barro	1		
spa-30-00235435-n	arranque	1		

Synset	Variante	C	I	Observación
spa-30-01140514-a	incómodo		1	Si bien incomodo es traducción de awkward, en el synset este la traducción mas adecuada seria torpe
spa-30-00645982-a	extraordinario	1		
spa-30-00386392-a	neutro	1		
spa-30-01997748-a	confiable	1		
spa-30-02306763-a	emocionante	1		
spa-30-02502163-a	inútil	1		
spa-30-01170243-a	saludable	1		
spa-30-01336587-a	estúpido	1		
spa-30-01407267-a	cierto		1	El lema es "true"pero en el contexto del synset (legítimo), la traducción "verdadero"seria mas apropiada
spa-30-02414031-a	fino	1		
spa-30-02469407-a	representante		1	La traducción correcta para el synset debería ser representativo"
spa-30-00403385-a	profundo		1	El synset refiere a "intenso"
spa-30-00493460-a	único	1		
spa-30-01655538-a	ajustado	1		
spa-30-01247240-a	caliente	1		
spa-30-00431447-a	transparente	1		
spa-30-00039394-a	proactivo	1		
spa-30-00805566-a	fresco	1		
spa-30-01687965-a	nuevo	1		
spa-30-00988232-a	magro	1		
spa-30-00487653-a	raro	1		
spa-30-01407267-a	legal	1		
spa-30-01044730-a	casual	1		
spa-30-00430756-a	confuso	1		
spa-30-02580449-a	gastado	1		

B

PSEUDOCÓDIGO

En este anexo se muestra el pseudocódigo de la estrategia de traducción implementada. Se puede observar el esqueleto del programa principal y además se incluyen los pseudocódigos con los algoritmos de cada uno de los selectores iniciales y de la fase de expansión.

```
procedure EXECUTE
  ExecuteFase1()
  AgregarResultadosFase1()
  ExecuteFase2()
end procedure

procedure EXECUTEFASE1
  synsets = Synsets()
  for synset in synsets do
    pos = ObtenerPos(synset)
    lemas = Lemas(synset)
    for lema in lemas do
      traducciones = PalabraObtenerTodasSusTraducciones(lema, pos)
      resultadoS = SelectorTraduccionSingular(lema, pos)
      resultadoM = SelectorMonosemia(lema, pos)
      InsertarEnBaseAuxiliar(synset, lema, pos, traducciones, resultadoS, resultadoM)
    end for
    resultadoF = SelectorFactorizacionTraduccion(synset, pos)
    resultadoAD = SelectorAdvivioDerivado(synset, pos)
    InsertarEnBaseAuxiliar(synset, lema, pos, resultadoF, resultadoAD)
  end for
end procedure

procedure EXECUTEFASE2
  synsets = ObtenerSynsetsAmbiguosNoMonosemicos()
  for synset in synsets do
    resultado = ExecuteOne(synset)
    if resultado not in WordNetActual then
      InsertarEnWordNetActual(resultado)
    end if
  end for
end procedure

function SELECTORTRADUCCIONSINGULAR((lema, pos))
  traducciones = Traducir(lema, pos)
  return #traducciones
end function

function SELECTORMONOSEMIA((lema, pos))
  synsets = Synsets(lema, pos)
  return 1 == #synset
```

PSEUDOCÓDIGO

end function

function SELECTORFACTORIZACIONTRADUCCION((synset, pos))

lemas = Lemas(synset, pos)

p = Primero(lemas)

result = Traducir(p, pos)

for lema **in** lemas **do**

traducciones = Traducir(lema, pos)

result = result \cap traducciones

end for

return result

end function

function SELECTORADVERBIODERIVADO((synset, pos))

diccionarioAdverbios = CrearDiccionarioAdverbios()

synsetAdjetivos = HasDerived(synset)

derivados = ListaVacía()

for adjetivo **in** synsetAdjetivos **do**

lemas = Lemas(adjetivo, pos)

for lema **in** lemas **do**

traducciones = Traducir(lema, pos)

for traduccion **in** traducciones **do**

derivado = DerivarAdverbio(traduccion)

if derivado **in** diccionarioAdverbios **then**

AgregarElementoALista(derivados, derivados)

end if

end for

end for

return derivados

end function

function DERIVARADVERBIO((palabra))

if TerminaEn("o", palabra) **then**

palabra = palabra - "o"

palabra = palabra + "amente"

else if TerminaEn("r", palabra) **or** TerminaEn("n", palabra) **then**

palabra = palabra + "amente"

else

palabra = palabra + "mente"

end if

end function

function EXECUTEONE((synset))

mejor = vacío

sinonimos = NearSynonym(synset)

for sinonimo **in** sinonimos **do**

lemas = Lemas(sinonimo)

for lema **in** lemas **do**

candidatos = TraduccionesCandidatas(synset)

for candidato **in** candidatos **do**

resultado = ConsultarCorpus(candidato, lema)

if resultado > mejor **then**

mejor = candidato

end if

end for

end for

end function

antonimos = NearAntonym(synset)

for antonimo **in** antonimos **do**

lemas = Lemas(antonimo)

for lema **in** lemas **do**

candidatos = TraduccionesCandidatas(synset)

PSEUDOCÓDIGO

```

    for candidato in candidatos do
        resultado = ConsultarCorpus(candidato, lema)
        if resultado > mejor then
            mejor = candidato
        end if
    end for
end for

hiperonimos = HasHypernym(synset)
for hiperonimo in hiperonimos do
    lemas = Lemas(hiperonimo)
    for lema in lemas do
        candidatos = TraduccionesCandidatas(synset)
        for candidato in candidatos do
            resultado = ConsultarCorpus(candidato, lema)
            if resultado > mejor then
                mejor = candidato
            end if
        end for
    end for
end for

meronimos = HasMeronym(synset)
for meronimo in meronimos do
    lemas = Lemas(meronimo)
    for lema in lemas do
        candidatos = TraduccionesCandidatas(synset)
        for candidato in candidatos do
            resultado = ConsultarCorpus(candidato, lema)
            if resultado > mejor then
                mejor = candidato
            end if
        end for
    end for
end for

hiponimos = HasHyponym(synset)
for hiponimo in hiponimos do
    lemas = Lemas(hiponimo)
    for lema in lemas do
        candidatos = TraduccionesCandidatas(synset)
        for candidato in candidatos do
            resultado = ConsultarCorpus(candidato, lema)
            if resultado > mejor then
                mejor = candidato
            end if
        end for
    end for
end for
return mejor
end function

function CONSULTARCORPUS((candidato, lema))
    cantCandidato = ObtenerOcurrencias(candidato)
    cantLema = ObtenerOcurrencias(lema)
    cantAmbas = ObtenerCoocurrencias(candidato, lema)
    return cantAmbas / (cantLema + cantCandidato)
end function
```

PSEUDOCÓDIGO

(*WordNet*, s.f.) (*Estudios de Lingüística del Español*, s.f.) (Dyvik, 2004) (Devitt y Vogel, 2004) (Oliver, 2014) (*WN-MCR-Transform*, s.f.) (Oliver, Climent, y Contreras, 2012) (*Wikimedia Downloads*, s.f.)
(Giménez y Márquez, 2005) (Yablonsky, 2010) (Española, s.f.)