

Prefijación en español: una propuesta de formalización computacional basada en modelos cognitivos

Isabel González

Mammut

Universidad de Los Andes

isabel@mammut.io

Génesis Montenegro

Mammut

Universidad de Los Andes

genesis@mammut.io

Rosángela Pulido

Mammut

Universidad de Los Andes

rosangela@mammut.io

Miguel Riveros

Mammut

Universidad de Los Andes

miguel@mammut.io

Hillary Romero

Mammut

Universidad de Los Andes

hillary@mammut.io

Lino Urdaneta

Mammut

Universidad de Los Andes

lino@mammut.io

Resumen

La interpretación sintáctica, morfológica y semántica de la formación de nuevas palabras presenta dificultades desde el punto de vista del procesamiento de lenguaje natural (Goldsmith, 2010), sobre todo si se realiza bajo el enfoque tradicional (Ungerer, 2002). En la siguiente ponencia se presenta un modelo computacional que formaliza prefijos del español utilizando algunos postulados de la lingüística cognitiva (Lakoff & Johnson, 1999; Langacker, 2008). El modelo se fundamenta en la hipótesis de que los prefijos tienen una estructura cognitiva de rasgos similar a la determinada para las preposiciones (Saint-Dizier, 2006; Waluch de la Torre, 2007; Zelinsky-Wibbelt, 1993), y que los primeros también pueden ser sometidos a operaciones metafóricas y metonímicas, además

de relaciones de recodificación (Ungerer 2002, 2007). Estas operaciones cognitivas son las que permiten diseñar un modelo computacional que elabora para cada prefijo un esquema que, a su vez, constituye una representación computacional de la palabra derivada, ya que permite elaborar una taxonomía basada en el uso metafórico de los prefijos en el proceso morfológico de formación de palabras. El esquema se amplía con información morfológica necesaria para relacionar el prefijo con la clase de palabra correspondiente. El modelo que se describe será utilizado en la anotación de un corpus de preguntas y respuestas que será procesado posteriormente utilizando algoritmos de aprendizaje de máquinas dentro de un *framework* desarrollado por la empresa Mammut.

Palabras clave: prefijos, procesamiento de lenguaje natural, morfología, semántica, gramática cognitiva.

1. INTRODUCCIÓN

Los procesos morfológicos de formación de nuevas palabras y, por extensión, el análisis morfológico de palabras compuestas y derivadas de las lenguas han sido tratados desde diversas perspectivas. Entre los mecanismos morfológicos de expansión léxica distribuidos en las distintas lenguas del mundo podemos encontrar prefijación, sufijación, compuestos *bahuvrihi*, composición, entre otros. A diferencia de la flexión y de la asignación de casos, dichas operaciones constituyen más un mecanismo eficiente de creación de nuevas formas léxicas que un procedimiento gramatical de asignación de funciones sintácticas dentro de la oración. Sin embargo, la delimitación estricta entre estas dos funciones es más una cuestión de grado que de categorización absoluta. En efecto, en muchas ocasiones un sufijo, además de posibilitar la creación de una nueva palabra, puede reorganizar sintácticamente y semánticamente el nuevo término con respecto al anterior. Por ejemplo, el sufijo *-dor*, además de derivar un sustantivo de un verbo (*comentar* → *comentador*), le proporciona a la palabra derivada una mayor probabilidad de funcionar en el rol semántico de agente de una oración que como paciente o benefactivo de la misma (Ungerer, 2002, 2007). Se trata, entonces, de un fenómeno morfológico complejo, con repercusiones en los niveles léxico, sintáctico, semántico y discursivo. La derivación léxica representa un reto, no solo teórico-analítico para la lingüística, sino también para el procesamiento de lenguaje natural (PLN), por cuanto dicho procedimiento y sus resultados deben ser modelados computacionalmente en diversos niveles, y a veces con distintas técnicas. Por ejemplo, el sufijo *-dor* puede ser entendido de las siguientes maneras:

- Interpretación morfosintáctica: la palabra derivada se reclasifica como sustantivo (*comentador*), con la posibilidad de adjetivación (*máquina clasificadora*). Admite variación en género y número, además de los prefijos y sufijos propios de los sustantivos. Al funcionar como sustantivo, puede tomar los roles sintácticos de sujeto, complemento, adjunto, atributo, etc. Como adjetivo, puede modificar a un sustantivo con el que establecerá lazos de concordancia.
- Interpretación semántica: el rol semántico de los derivados en *-dor* es diverso. Puede ser agente (*El vendedor vino a mi casa*), benefactivo (*Robaron a un contador*), paciente (*Compré un calentador*), atributo (*Esperar es agotador*), entre otros.
- Interpretación lexicográfica: el sufijo *-dor* suma a la base léxica de la palabra significados tales como ‘persona que realiza la acción *x*’, ‘artefacto capaz de realizar el

proceso y' o 'cualidad derivada o efecto de la acción z' '. Se trata de una función que relaciona el sentido general de la palabra de origen con la palabra derivada. Esta función se suspende o es inválida cuando un proceso diacrónico ha idiomatizado la forma léxica hasta el punto de ser irreconocible el procedimiento de derivación desde el punto de vista sincrónico.

- Interpretación pragmático-discursiva: hay sentidos pragmáticos relacionados con los usos sociales del sufijo *-dor*, tales como el sentido peyorativo que puede adquirir este afijo en ciertos contextos comunicativos (*los opinadores y comentadores de oficio son cada vez más frecuentes en las redes sociales*).

Los prefijos, por otra parte, actúan de forma algo diferente a la de los sufijos. La función semántica que reclasifica roles actanciales, por ejemplo, no está presente en la prefijación (Ungerer, 2002); en lugar de reclasificar, los prefijos re-sitúan la base con respecto a otros elementos que les sirven de referencia, como veremos más adelante.

Para la interpretación semántica y lexicográfica hemos optado por utilizar el marco cognitivo desarrollado principalmente por Langacker (2009, 2008, 2003) y Lakoff & Johnson (1991, 1999) porque consideramos que permite descomponer el prefijo en una serie de rasgos discretos organizados cognitivamente, que pueden facilitar su modelado.

El tratamiento computacional más simple que se le da a los afijos suele ser el reconocimiento de los mismos en una palabra (utilizando una gramática regular, por ejemplo), para luego asignar a dicha forma un código o rasgo morfológico o semántico (del tipo [+diminutivo] o "reiteración") que son útiles para extraer información de un corpus anotado, que no suelen ser interpretados por un sistema computacional (véase el tratamiento que hace, por ejemplo, Silberztein (2016)).

En la presente investigación se utilizarán algunos rasgos de base cognitiva para la clasificación semántica y léxica de algunos prefijos del español, con el propósito de desarrollar un modelo de anotación de prefijos que permita su modelado computacional por medio de esquemas compuestos por características diversas. Cabe destacar que el modelo que presentamos está en una primera fase de diseño, razón por la cual no ha sido puesto a prueba en una etapa de implementación. La parte computacional del modelo se formaliza como una serie de parámetros y valores asociados a ellos que se configuran para cada uso de un prefijo.

A continuación presentaremos el marco conceptual en el que se inscribe el modelo que está siendo desarrollado. Posteriormente presentaremos la visión tradicional de la prefijación y los antecedentes de este trabajo. Finalmente se presentarán las principales características del modelo cuyo valor principal (en esta etapa de la investigación) se encuentra en el uso de parámetros cognitivos para la formalización de los prefijos en español con aplicaciones en el PLN.

2. LA PERSPECTIVA GRAMATICAL COGNITIVA

La gramática cognitiva (GC) es una teoría lingüística basada en la premisa de que la gramática tiene un fundamento estrictamente simbólico (Langacker, 2008, 2009, 2013), es decir, que tanto gramática como léxico forman una gradación explicable a partir de

combinaciones estructurales semánticas que forman combinaciones estructurales fonológicas. Todo símbolo es, simultáneamente, semántico y fonológico, y las operaciones de combinación simbólica no son de ninguna manera autónomas o independientes del sentido. La GC forma parte de la lingüística cognitiva, y echa mano de varios de los recursos de esta última: teoría de la metáfora y la metonimia, semántica cognitiva, gramaticalización, investigaciones en funcionalismo tipológico universal, gramática de construcciones, entre otros (Langacker, 2003: 7).

La GC es un paradigma experiencialista que pone como eje principal de la conceptualización la “naturaleza corpórea o corporeización del lenguaje” (Cuenca & Hilferty, 1999: 17). Esta “corporeización” se refiere a que el sujeto hablante es considerado uno de los puntos de referencia privilegiados a partir de los cuales se construyen y constituyen las configuraciones semánticas de una lengua en particular y del lenguaje en general (Lakoff & Johnson, 1999); la corporeización es uno de los pilares centrales que sostienen la configuración de características semánticas de los prefijos (que mostraremos más adelante), ya que en el modelo desarrollado se parte de la idea de que el sujeto hablante es el principal punto de referencia desde el cual se considera semánticamente el uso del prefijo. La suposición de un cuerpo físico que percibe fenómenos externos es muy útil para la esquematización de los prefijos en el sentido de que permite establecer relaciones como “el elemento se aleja de mí” o “con respecto a otro, este elemento está por encima de él”. En palabras de Langacker (2008: 524): “Ultimately, the world we construct is grounded in our experience as creatures with bodies who interact with their surroundings through physical processes involving sensory and motor activity. This is known in cognitive linguistics as **embodiment**.”

Además de la “mente corporeizada” de Lakoff y Johnson (1999), otro aspecto cognitivo que hemos utilizado en el modelado de prefijos es la noción de la metáfora conceptual como un mecanismo de conceptualización que permite concebir un dominio de experiencia (destino) a partir de otro dominio de experiencia (origen) –generalmente sensomotor– (Lakoff y Johnson, 1991, 1999). Esta relación entre el dominio de origen y el dominio de destino será aprovechada por el modelo para organizar los distintos usos metafóricos de los prefijos. Por ejemplo, el prefijo *ex-* se entiende de forma distinta en *excarcelar* y *exesposa* ya que el primero supone una ‘salida física’ de un lugar (la cárcel) y el segundo como una ‘salida abstracta’ de una categoría (la de esposa); a pesar de ser dos interpretaciones ligeramente diferentes, la noción de la metáfora conceptual permite establecer relaciones entre ellas.

Un tercer aspecto esencial en el diseño del modelo cognitivo de formalización semántica de prefijos fue el principio gestáltico de prominencia perceptual, en particular el fenómeno denominado segregación de figura / fondo (Ungerer, 2006), o modelo del trayector y el locus o *landmark* (Cuenca & Hilferty, 1999). Hemos utilizado esquemas de imagen de orientación (*orientational image schemas*) para predefinir las posibles relaciones entre trayectores (Tr) y el locus (LM). El Tr es un elemento que tiene prominencia perceptual, mientras que el LM funciona como el fondo sobre el que se aprecia el Tr.

Para la modelización de las relaciones entre Tr y LM han sido diseñados esquemas *ad hoc* que se basan en la distinción entre elementos móviles y fijos, y en relaciones espaciales factibles entre ellos (en una, dos y tres dimensiones). Estas relaciones espaciales, basadas

en la corporeización, se “extienden” a otras dimensiones conceptuales por medio de metáforas que explicitan los distintos nexos que pueden darse entre un Tr y un LM. Esto permite la elaboración de esquemas formales para cada uso prefijal, basados en las tres características señaladas (corporeización, metáfora conceptual y prominencia perceptual). Veremos más adelante algunas limitaciones derivadas de tal decisión, así como también algunas extensiones que hemos tenido que incorporar para subsanar dichas limitaciones.

3. LA PREFIJACIÓN DESDE EL PUNTO DE VISTA DE LA GRAMÁTICA TRADICIONAL

Lang (1992) concibe los prefijos como un afijo que precede a la base del derivado. El rasgo más destacado del prefijo es su no integración con la base: no se funde con la raíz, se mantiene claramente delimitado respecto a ella (*pre/fabricado*, *anti/cuerpo*, *in/visible*, *contra/decir*). En cuanto a las funciones atribuidas a los prefijos, existen controversias. Martínez Amador (1987) comenta que algunos autores atribuyen al prefijo una función preposicional o adverbial; funciona como preposición cuando el prefijo rige al nombre, por ejemplo en *anteojo* (delante del ojo). Por otra parte, oficia de adverbio cuando modifica el nombre por razón del adverbio antepuesto; por ejemplo, *trasudor* (un sudor que sobreviene tras una congoja). Sin embargo, para Alarcos (1983) los prefijos no son funcionalmente ni lo uno, ni lo otro; históricamente pueden tener algún tipo de relación con preposiciones o adverbios. La función de los prefijos, como sucede a los morfemas ligados, se deriva de un significado abstracto básico; en otras palabras, la diferencia entre significado y función se desdibuja. Para Bajo (1997), a los prefijos pueden asignárseles valores semejantes a los que comportan las preposiciones, aunque, para Pottier (1970), la función es de adición: no subordinan, no coordinan, se limitan a añadir un valor sémico.

Morera (1997) opina, a su vez, que no hay mayores distinciones entre prefijos y preposiciones en cuanto a su determinación semántica. El modelo que presentamos en esta instancia parte, precisamente, de esta hipótesis.

4. ANTECEDENTES

Son relativamente escasos los estudios desarrollados en torno a la prefijación desde la perspectiva del PLN, sobre todo en español. Para el inglés, Byrd (1983) parte de los principios de la formación de las palabras y diseña un mecanismo de eliminación de afijos (*Affix Stripping*) para economizar el almacenamiento y también aumentar el tamaño aparente del corpus. Santana *et al* (2004) destina esfuerzos para el reconocimiento, producción y manipulación de relaciones morfológicas de cualquier palabra derivada por prefijación. Otro trabajo que es pertinente mencionar es el realizado por Bergsma *et al* (2010), quienes desarrollan un clasificador (utilizando *Support Vector Machines*) para predecir la composición derivativa de verbos formados por un prefijo y una raíz léxica. Por último, la investigación de Bernhard (2006) está especialmente dirigida a la extracción, y posterior clasificación dentro de familias semánticas, de términos científicos y técnicos generalmente compuestos por prefijos clásicos. En cambio, en la sufijación, los algoritmos e investigaciones PLN abundan, en particular por la necesidad de los sistemas de extracción y recuperación de información de encontrar similitudes entre las raíces de los elementos léxicos (procedimiento denominado *stemming*).

5. MODELO DE CARACTERIZACIÓN SEMÁNTICA DE PREFIJOS DE MAMMUT

Antes de continuar es necesario advertir que el modelo que presentamos está en una fase de diseño, razón por la cual muchos detalles de los esquemas que presentamos están bajo escrutinio. Sin embargo, consideramos que las bases teóricas del modelo son sólidas y que pueden contribuir tanto al estudio semántico detallado de los prefijos como a formalizar computacionalmente las funciones de estos elementos. En una siguiente fase, los esquemas de los prefijos que presentamos serán utilizados para construir automáticamente un subconjunto del lexicón a partir de un corpus de preguntas y respuestas.

El modelo computacional de prefijos que estamos construyendo en la actualidad se inspira en una modelización semántica previa hecha a partir de postulados cognitivos sobre el uso y significado de preposiciones en inglés y español (González *et al.*, en prensa). A continuación presentaremos algunos de los parámetros y valores que conforman los esquemas de interpretación de uso de los prefijos, y cuyo conjunto conforma el modelo. Luego mostraremos el modelado de algunos prefijos a manera de ejemplo.

Los parámetros que se presentan a continuación tienen diferentes valores; la distinta configuración de estos valores para cada prefijo permite formalizarlo semántica, léxica y morfológicamente.

El modelo que proponemos organiza los esquemas en una tabla; cada caso (significado del prefijo) está ubicado en una fila distinta y los parámetros se encuentran en las columnas. El primer parámetro es el que hemos denominado **dimensión metafórica**. Este toma como valor, en cada fila, el dominio metafórico conceptual de destino del uso prototípico del prefijo. Estas dimensiones pueden ser ampliadas o reducidas según las necesidades descriptivas y analíticas de los investigadores; hasta el momento la dimensión metafórica incluye 8 valores: espacio, tiempo, espacio metafórico, transacción, circunstancia, medida, instrumento y conjunto. Las dimensiones son uno de los ejes centrales del modelo que proponemos, ya que contribuyen a vincular entre sí distintos usos de los prefijos. Por ejemplo, en *interponer*, el prefijo *inter-* se une a *poner* para indicar que se realiza la ‘acción de poner una entidad física entre otras dos distintas a la primera’, por ejemplo en *la pared se interpone entre nosotros* (obsérvese, de paso, que *interponer* requiere un argumento introducido por la preposición *entre*). El prefijo *inter-*, en este ejemplo, funciona como un indicador de interrelación física (equivalente en nuestro modelo a una dimensión espacial). En cambio, en *internacionalista*, el prefijo implica la interrelación de entidades no físicas, sino psíquicas o sociales (dimensión metafórica de circunstancia).

Los siguientes parámetros son el Tr (elemento prominente) y el LM (fondo), que pueden ser concebidos como **móviles** o **fijos**; esta distinción es útil cuando las relaciones connotadas por el prefijo implican dinamismo entre el Tr y LM. El prefijo *contra-* (*contradecir*, *contraposición*) implica un ‘ir en dirección contraria de algo’, lo que en nuestra propuesta se concibe como un Tr móvil que se aleja de un LM que consideramos, metafóricamente hablando, como una locación fija. La relación de distanciamiento entre Tr y LM que hemos utilizado en el ejemplo anterior se formaliza en el parámetro denominado **direccionalidad**: este atributo toma valores relacionados, en principio, con las posibles relaciones espaciales entre Tr y LM. Estas relaciones espaciales fueron construidas, en un

comienzo, a partir de las posibilidades experimentadas física y psíquicamente por un cuerpo, en una amalgama de las posibles ubicaciones relativas configuradas por una concepción tridimensional del espacio, y por el supuesto de que estas ubicaciones son percibidas siempre desde un punto de vista humano. Por ejemplo, un ser humano que observe dos entidades en una dimensión lineal (dos puntos en una recta, por ejemplo) sabe que estas pueden quedarse fijas en sus posiciones, pueden acercarse entre sí, alejarse entre ellas, u oscilar en sus posiciones relativas; si las consideramos dos puntos en una línea, sabemos que los puntos pueden quedarse equidistantes entre sí, o variar sus posiciones relativas para acercarse (acercarse tanto que puede sobrepasar en la recta al otro punto) o distanciarse (en una figurativización del espacio inspirado en la geometría euclidiana y en la mecánica newtoniana, estas son las únicas posibilidades). Este fue el principio organizador de lo que denominamos direccionalidad, y se vincula estrechamente con la concepción fija o móvil del Tr y LM ya que, en términos relativistas, entre los puntos de una recta no hay un punto de vista privilegiado y podemos tanto concebir que el punto A se mueve hacia B, o B hacia A; sin embargo, el punto de vista privilegiado del ser humano difumina la visión relativista y asigna, normalmente, a uno de las entidades un punto de referencia “absoluto”. Este punto de referencia se considera “fijo” y se le asigna la categoría “móvil” a la otra entidad, y es por ello que junto a la direccionalidad hay que considerar las características estáticas o dinámicas de las entidades involucradas. El proceso de la construcción de los valores posibles de la direccionalidad sigue esta analogía para dos o tres dimensiones, ampliando las posibilidades en cada caso. La metáfora del contenido / contenedor, tan fértil en gramática cognitiva, es formalizada en nuestra propuesta como un volumen en tres dimensiones que funge como LM y contiene otro volumen tridimensional concebido como el Tr. Nuestro modelo contempla 31 tipos de direccionalidad diferentes. Esta parametrización de relaciones espaciales físicas puede reinterpretarse metafóricamente cuando entra en conjunción con los distintos valores de la dimensión metafórica mencionados arriba.

La formalización contempla dos parámetros más. El **volumen** se deriva no ya de la posición espacial relativa de dos entidades, sino de la relación entre el movimiento y las características físicas del Tr y LM: por ejemplo, el valor CONTACTO+ implica que una entidad delimitada se mueve hasta otra que también tiene un volumen definido, y que el movimiento los pone en ‘contacto’ (físico o metafórico). Las categorías de volumen relacionado con el movimiento pueden ser numerosas, así que solo son incluidas en el modelo cuando se requieren para distinguir dos casos diferentes que comparten el resto de los valores, es decir, son categorías que pueden considerarse *ad hoc*. La formalización del prefijo *re-*, al cual le asignamos un significado que podría glosarse como un movimiento de ‘volver atrás o al principio de algo’, requiere también incorporar una noción de reiteración posible, que implica que el límite del LM que marca el “principio” desde el cual se comienza, puede ser cruzado de manera recurrente; para indicar este reiterado traspaso del volumen hemos incorporado el valor REITERACIÓN+ al parámetro volumen.

El último parámetro lo hemos denominado **subjetividad**, y como el anterior, está conformado por valores que se presentan como necesarios para distinguir un caso de otro. A diferencia de volumen, en subjetividad incorporamos significados vinculados con la manera en la cual un observador evalúa la disposición de las entidades implicadas, sobre todo desde un punto de vista social o utilitario. Por ejemplo, el valor subjetivo

FINALIDAD+ se usa para indicar que el prefijo *pro-* además de un acercamiento, debe entenderse como algo que se realiza con una finalidad en mente (*procrear*).

Además de los parámetros mencionados, hemos incorporado otros 3 utilizados por Ungerer (2003): **clase conceptual**, **categoría de uso** del prefijo, y **clase de palabra** (POS) con la que el prefijo se utiliza. Se trata de parámetros conceptuales, funcionales y morfológicos respectivamente. En particular, el último es necesario porque enlaza los esquemas prefijales construidos con elementos del lexicón que pueden constituir la base de la palabra derivada. En la tabla 1 mostramos las categorías presentadas por Ungerer (2003).

Tabla 1. Categorización del prefijo según Ungerer (2003)

Clase conceptual	Categoría de uso	POS
ualidad	Antonimizador	Adjetivo
Dominio	Ubicador	Sustantivo
Acción	Graduador	Verbo
Persona		Adverbio
Objeto concreto		

Además de estos parámetros, estamos incorporando en la actualidad uno para especificar la posición relativa de un prefijo graduador con respecto a otro prefijo del mismo tipo; en este sentido, es importante determinar semánticamente que los graduadores forman una escala entre ellos: *super-* < *mega-* < *hiper-* < *ultra-*, entre otros. Esta gradación está por determinar, así que no será presentada en el ejemplo que sigue.

7. MODELADO DEL PREFIJO *RE-*

En esta sección presentaremos el esquema semántico del prefijo *re-* y comentaremos cómo deber ser interpretado. Los esquemas semánticos no se elaboran para todas las palabras derivadas que compartan un determinado prefijo, sino solo para casos prototípicos. Estos prototipos son luego utilizados para crear sugerencias automáticas en el reconocimiento de tokens durante un procedimiento de anotación / desambiguación.

En la tabla 2 mostramos la dimensión metafórica y la relación entre Tr y LM.

Tabla 2. Modelado de las formas derivadas a partir de *re-* en relación con la dimensión metafórica y la direccionalidad del Tr y LM

	Dimensión	Tr	Direccionalidad	LM
<i>rehacer</i>	tiempo	M	!>	F
<i>reapertura</i>	tiempo	M	!>	F

En primer lugar, hay que aclarar en este punto lo que vamos a entender como Tr y LM en los procesos de derivación léxica. La perspectiva que adoptaremos será la de hacer hincapié

en el proceso de derivación como una actividad de creación simbólica. La palabra derivada (*rehacer* por ejemplo) es una estructura simbólica compleja formada por un polo fonológico y un polo semántico unidos por un proceso de ensamblado simbólico (Langacker, 2008, 2013) que compone la palabra derivada a partir de la suma fonológica y la suma semántica del prefijo y su raíz. En términos de la creación simbólica, consideramos que el foco (Tr) está en la nueva estructura simbólica derivada, y el fondo (LM) es uno de los componentes de dicha estructura, en particular el sentido determinado por la base de la palabra. Se crea –y se usa– el vocablo *rehacer* para indicar que ‘se vuelve a *hacer* algo’. En este sentido, *rehacer* guarda una relación con *hacer*, y supondremos que esa relación también se puede ser vista a través del cristal del foco y el fondo: en términos de la derivación, *rehacer* se interpreta a partir de la base *hacer* como fondo. Es precisamente esta relación la que se formaliza en la tabla 2: *rehacer* se considera una palabra derivada que denota una acción (un movimiento, y por eso el Tr se considera móvil) que implica un nuevo comienzo (veremos unas líneas más adelante de dónde sale este sentido de ‘nuevo’) con respecto a otra acción (el LM *hacer*); esta acción se realiza en un plano temporal (dimensión metafórica de tiempo). Tenemos, entonces, que un *rehacer* (Tr móvil) supone comenzar desde el principio (desde atrás, indicado por el símbolo ¡>) un proceso que se había realizado tiempo atrás; el proceso anterior se toma como un punto de referencia fijo en el tiempo (LM fijo). En la notación que empleamos, M se entiende como móvil y F como fijo. Este análisis se corresponde, asimismo, con el sentido etimológico del prefijo *re-* (del proitálico *wre*, y este del protoindoeuropeo *ure*, que significaba ‘detrás’). Vale la pena destacar que hemos encontrado que los étimos de los prefijos han resultado de mucha utilidad para establecer la dimensión metafórica básica, así como también para determinar la direccionalidad de la relación.

El esquema semántico prefigurado en la tabla 2 es, sin embargo, insuficiente: hace falta incorporar el sentido de reiteración, de que se trata de un nuevo principio. En efecto, una *reapertura* no es la apertura desde el principio, sino que se trata de una nueva apertura, desde el comienzo. Esta idea de reiteración se consigue con el rasgo REITERACION+, que ya discutimos anteriormente. En la tabla 3 se muestran los parámetros volumen y subjetividad.

Tabla 3. Parámetros de volumen y subjetividad en la modelización de las palabras derivadas con el prefijo *re-*

	Volumen	Subjetividad
<i>rehacer</i>	REITERACIÓN+	NINGUNO
<i>reapertura</i>	REITERACIÓN+	NINGUNO

En la tabla 4 se exhiben los tres últimos parámetros. Según la clase conceptual, se trata de un prefijo de acción, que ubica a la palabra derivada con respecto a la raíz (la categoría de uso es la de ubicador). En lo único en lo que se distinguen los prefijos *re-* en *rehacer* de *reapertura* es que uno se utiliza con verbos y otro con sustantivos. Esto puede reinterpretarse, desde el punto de vista morfológico, como un solo prefijo, que se adhiere a dos clases de palabras diferentes. Como se dijo arriba, esta información es muy importante porque limita las combinaciones posibles de los prefijos con las raíces léxicas. Este último

parámetro, POS, no forma parte de la caracterización semántica llevada a cabo con las herramientas cognitivas, pero resulta fundamental para la esquematización del prefijo, sobre todo porque restringe morfológicamente sus combinaciones.

Tabla 4. Clase conceptual, categoría de uso y POS de las palabras derivadas

	Clase conceptual	Uso	POS
<i>rehacer</i>	Action	Ubicador	verbo
<i>reapertura</i>	Action	Ubicador	sustantivo

8. CONSIDERACIONES FINALES

El modelo que esbozamos en esta ponencia apenas está en ciernes. Sin embargo, la visión cognitiva del modelo le da sustento teórico a nuestra propuesta, permitiendo observar que esta visión es válida para el análisis de los prefijos en español, así como útil para algunas aplicaciones en el procesamiento de lenguaje natural. Por otro lado, desde el punto de vista práctico, resuelve dos problemas fundamentales para el sistema que está siendo diseñado por la empresa Mammuth para la comprensión inteligente de datos lingüísticos. Primero, consigue una representación semántica formal de la prefijación, usando la composicionalidad de los elementos léxicos y morfológicos. Segundo, consideramos que va a permitir la extensión automática de un lexicón elaborado manualmente, con herramientas analíticas ya diseñadas para tratar con las preposiciones (González et al, en prensa), lo que incidirá en el posterior ahorro de recursos (tiempo y dinero) empleados en la anotación manual requerida para el desarrollo de algoritmos de clasificación automática.

AGRADECIMIENTOS

Agradecemos el apoyo institucional brindado por Mammuth.io durante la preparación de esta ponencia.

Referencias bibliográficas

- Alarcos Llorach, Emilio. 1974. *Gramática estructural*, Madrid: Gredos.
- Bajo Pérez, Elena. 1997. *La derivación nominal en español*. Madrid: Arco/Libros.
- Bergsma Shane, Bhargava Aditya, He Hua, and Kondrak Grzegorz. 2010. Predicting the semantic compositionality of prefix verbs. En *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, 293-303. Cambridge.
- Bernhard, Delphine. 2006. Multilingual term extraction from domain-specific corpora using morphological structure. En *Proceedings of the 11th European Chapter Meeting of the ACL*, 171-174. Trento, Italy.
- Byrd, Roy. 1983. Word Formation in Natural Language Processing Systems. En *Proceedings of the Eighth International Joint Conference on Artificial Intelligence*, 704-706. Karlsruhe, Germany.
- Cuenca, María & Joseph Hilferty. 1999. *Introducción a la lingüística cognitiva*. Barcelona, España: Ariel.

- Goldsmith, John. 2010. Segmentation and Morfology. En *The Handbook of Computational Linguistics and Natural Language Processing*. Sussex: Wiley-Blackwell
- Lakoff, George & Mark Johnson. 1991. *Metáforas de la vida cotidiana*. Madrid: Cátedra.
- Lakoff, George & Mark Johnson. 1999. *Philosophy in the flesh. The embodied mind and its challenge to western thought*. Nueva York: Basic Books.
- Lang, Mervyn Francis. 1992.: *Formación de palabras en español*. Madrid: Cátedra.
- Langacker, Ronald. 2008. *Cognitive grammar. An introduction*. Nueva York: Oxford University Press.
- Langacker, Ronald. 2009. *Investigations in cognitive grammar*. Berlín: Mouton de Gruyter.
- Langacker, Ronald. 2013. *Essentials of cognitive grammar*. Nueva York: Oxford University Press.
- Martínez Amador, Emilio. 1987. *Diccionario gramatical y de dudas del idioma*. Barcelona: Sopena.
- Morera Pérez, Marcial. 1997. Naturaleza semántica de los prefijos españoles. En *Actas del Congreso Internacional de Semántica*, (La Laguna, 27-31- octubre) (en prensa).
- Pottier, Bernard. 1970. Los infijos modificadores en portugués. Nota de morfología general. En *lingüística moderna y filología hispánica*, 99-133. Madrid: Gredos.
- Santana, Octavio, José Rafael Pérez, Francisco Carreras & Gustavo Rodríguez. 2004. Relaciones morfológicas prefijales del español. *Procesamiento de Lenguaje Natural* 32: 9-36.
- Silberztein, Max. 2016. *Formalizing Natural Languages. The Nooj Approach*. Londres: Wiley
- Ungerer, Friedrich. 2002. The conceptual function of derivational word-formation in English. *Anglia*, 120, 534-567.
- Ungerer, Friedrich & Hans-Jorg Schmid. 2006. *An introduction to cognitive linguistics* (Segunda edición.). Edinburgo: Pearson Education Limited.
- Ungerer, Friedrich. 2007. Word-Formation. En Geeraerts, Dirk & Hubert Cuyckens, *The Oxford Handbook of Cognitive Linguistics*, 650-675. Nueva York: Oxford University Press.