

Neologismos con prefijos productivos

Silvana Elide Pierabella

Doctorado en Lingüística y Lenguas (UNR)

Rosario, Argentina

silvanapierabella@gmail.com

"Es increíble lo que ayuda un nuevo cajón en el lugar adecuado de nuestro archivo."

L. Wittgstein (Aforismos)

Abstract

The following report brings information about the method employed in automatic retrieval of neologisms within a corpus dealt with NooJ, a free NLP tool that is used all over the world in academic research, since it allows to work with linguistic data from any language.

Keywords: Computational Linguistics, Parser, Corpus, Natural Language Processing.

Resumen

El siguiente artículo brinda información sobre el método empleado en la detección automática de neologismos dentro de un corpus tratado con la herramienta NooJ, la cual es utilizada por investigaciones académicas en todo el mundo, dado que puede configurarse para cualquier idioma.

Palabras clave: Lingüística Computacional, Análisis automático, Corpus, Lenguaje artificial

1-Introducción

Este artículo se enmarca en el procesamiento automático del lenguaje natural (NLP) y aborda la detección de neologismos compuestos por prefijos. A fin de relevarlos, se utilizan como lista de exclusión los recursos disponibles gracias a los investigadores del Grupo Infosur (UNR), quienes implantaron el módulo de español de Argentina en la aplicación NooJ, creada por Max Silberztein, Profesor de la Universidad de Franche-Comté, junto a la comunidad INTEX desde la perspectiva léxico-grammatical de la LADL propuesta por Maurice Gross.

Los diccionarios de español con los que cuenta dicho módulo de NooJ reúnen 12969 lemas y se dividen según un criterio categorial: *nombres.dic*, *adjetivos.dic*, *verbos.dic*, *prepositions.dic*. Este último archivo no se compilará con el corpus, porque a pesar de que la preposición (P) interviene en la composición léxica, se trata de una clase funcional cerrada, por lo cual sus elementos no aceptan prefijos ni se incrementan como sucede con los nombres (N), adjetivos (A) y verbos (V)¹ que son sensibles a la innovación cada vez que un concepto o entidad es requerida por el vocabulario e incluso cuando una traducción se enfrenta a vacíos lingüísticos entre la lengua fuente y la lengua meta.

“Se sabe que el léxico no es finito, que la adquisición del léxico de parte del hablante no tiene límites; de igual modo, la incorporación de nuevos vocablos en las listas del diccionario -nuevos sustantivos, adjetivos, verbos- es constante (tanto en el diccionario mental como el escrito), con tal que la novedad se ajuste a los rasgos categoriales y a la morfología de la lengua de la que se trate.” (Múgica 2014:2)

¹ Las notaciones entre paréntesis corresponden a las denominadas categorías léxicas. Finalmente se aplica el comando “Unknown” a fin de comparar los diccionarios de español de NooJ con el corpus y Linguistic Analysis aporta las siguientes descripciones:

Corpus Language is "Spanish(sp)";
Original Text File format is "Default".
Corpus consists of 411 text files Text Delimiter is: "\n"
Corpus contains 82109 text units delimited by "\n" 45748
digits
382110 other delimiters
1267129 tokens
839271 word forms
Linguistic Resources applied to the text: tutorial.nod
Corpus contains 4790 different annotations

Aunque los neologismos han surgido en las lenguas romances desde su estadio temprano por medio de diversas operaciones morfológicas, la hipótesis de investigación sugiere que en la actualidad predominan novedosos procesos de formación con prefijos productivos.

2- Metodología de trabajo en corpus

Los textos que se trabajan fueron descargados de la web como CORLEC-TXT con formato oral-utf8 e importados a NooJ con el objeto de construir un corpus integrado por 411 documentos mediante el menú *File > New Corpus*. Así se obtuvo el archivo neologismos.noc que se compiló con los diccionarios anteriormente mencionados para permitir a NooJ localizar los neologismos.

Es importante destacar que los recursos lingüísticos a los que apela este método deben estar actualizados, de otra manera los resultados podrían mostrar omisiones en lugar de nuevas unidades lingüísticas. También es fundamental que en el archivo *properties.def* se haya declarado PREF y su clasificación.

Nooj pudo identificar 1267129 tokens entre los cuales 486 representan expresiones desconocidas, 272 elementos obedecen a la etiqueta PREF y 139 son reconocidos como intensificadores con la siguiente distribución: hipo (35), hiper (74), meta (13), macro (6), micro (11). Estos prefijos fueron almacenados en TAS (Textual Annotation Structure) de modo que estén accesibles en la siguiente etapa de investigación a fin de analizar su productividad.

3- Análisis y formulación de modelos

Los intensificadores forman una clase semántica, de ahí que comparten propiedades escalares a diferencia del resto de los prefijos, por lo tanto una hipótesis que justificaría su productividad² se asienta en el hecho de que la propiedad escalar tiene carácter transcategorial, o sea que sería pertinente tanto para sustantivos, adjetivos o verbos.

“En términos de Kennedy y Mc Nally (2005), la gradabilidad es característica no solo de los A sino también de V y de N, y las propiedades escalares son compartidas por expresiones categorialmente distintas, pero derivacionalmente relacionadas.” (Nora

² Para Jackendoff la productividad consiste en regularidad más que en cantidad de ítems. Prefijos tales como *macro/micro/hipo/hiper* aparecieron unidos a gran variedad de lemas que ya existían como ítem lexical independiente y se convirtieron en neologismos mediante la prefijación.

Múgica, Ma. Eugenia Mangialavori Rasia, 2014:21)

Este procedimiento morfológico muy común en la actualidad, antes solamente se daba entre elementos que exclusivamente pertenecían al griego o al latín. Por lo tanto, hay abundante bibliografía que aborda los prefijos desde una perspectiva clásica y recuerda que antiguamente estuvieron vinculados a las preposiciones.

“Los especialistas en lingüística teórica y los expertos en lingüística computacional han tratado el léxico, por lo general, como un conjunto estático de sentidos de palabras, etiquetados con rasgos correspondientes a información sintáctica, morfológica y semántica. Desde este punto de vista, los significados de las palabras fueron asociados generalmente con ítems léxicos distintos.” (Pustejovsky, 1995:1)

Por ese motivo, el presente enfoque neológico se propone analizar la dinámica de los prefijos intensificadores y sus particularidades. Dado que es evidente su incidencia en la formación de neologismos recientes y la mayor frecuencia de *hiper*, se partió de este prefijo para formalizar una primera descripción en la se distinguen cuatro modelos:

- Modelo 1: prefijo vinculado a una base griega - Ex.: *hipérbole*, *hipertermia*, *hiperestesia*.
- Modelo 2: prefijo vinculado a un N del español- Ex.: *hipersensibilidad*, *hipertexto*.
- Modelo 3: prefijo vinculado a un A del español- Ex.: *hipercalórica* e *hiperestática*
- Modelo 4: prefijo libre + A del español - Ex.: *hiper famoso*³

Estos modelos contribuyeron a mostrar que el prefijo *hiper* cambia su procedimiento morfológico convencional y renueva sus propiedades semánticas e incluso sintágmáticas, lo cual brinda otra razón para explicar su productividad, además de la justificación transcategorial que es común a todos los intensificadores.

Según Cabré⁴, un neologismo, sea un nuevo elemento o una innovación en uno existente, debe tener un alto grado de integración en todos los niveles del lenguaje.

En tal sentido, otro motivo de productividad podría residir también la flexibilidad tanto semántica como morfológica para completar la grilla categorial: *hipertexto*, *hipertextual*, *hipertextualizar*, *hipertextualizado*.

³ En el modelo 4 se manifiesta la equivalencia semántica entre los intensificadores y los adverbios (*hipergoloso* = *muy goloso/extremadamente goloso*) a través de la paráfrasis, recurso que Koza utilizapara automatizar la explicación de términos médicos.

⁴ Cabré, Teresa; *La Terminología: Teoría, metodología, aplicaciones*, Barcelona, Editorial Antártida, 1992.

Sin embargo, no es obvio el modo en que se articula el significado de un nuevo elemento. Si pensamos en *hipertexto*, por ejemplo, podemos advertir que *hiper* no hace alusión a una escala como en *hipercongelado*, en donde se intensifica la propiedad representada en el *A congelado*. En el caso de *hipertexto* el prefijo altera la ontología, en otras palabras, introduce propiedades que no estaban en el *N texto*.

Este comportamiento podría interpretarse como consecuencia de que en el español moderno este intensificador se extendió gradualmente a las bases nominales haciendo referencia a la ponderación en lugar del aumento de los rasgos semánticos implicados. Por ejemplo, el *N hiperpiso* significa un piso muy grande o bien un piso con excelentes atributos, al igual que *hiperlíder* no remite al tamaño de la base nominal sino a su valoración.

En el neologismo *hipertexto* el intensificador pierde entonces su connotación de tamaño o cantidad porque no se aplica a un elemento compatible con la propiedad escalar, por lo que el nuevo elemento léxico redefine al prefijo y lo habilita para unirse a otros lemas sin restricciones semánticas.

Consecuentemente, gracias a la posibilidad de despojarse de la propiedad escalar, el prefijo no solo expande su alcance semántico sino también sus oportunidades en la selección de lemas.

4- Conclusiones

Este reporte muestra la metodología y los pasos que se siguen al tratar un corpus con NooJ para la detección de neologismos con un patrón morfológico determinado y al mismo tiempo demuestra que dicha aplicación informática es apta para procesar una relevante cantidad de textos en conjunto.

La etapa de investigación expuesta ha proporcionado registros sobre la incidencia que tienen los prefijos en la creación de neologismos, y especialmente los denominados intensificadores. Dado que con el correr del tiempo han dejado de unirse únicamente a elementos de origen griego renovando sus rasgos semánticos, pudieron participar en lo sucesivo no solo de nuevos procesos morfológicos (Modelo 2) sino también sintácticos (Modelo 4).

De este modo se pone de manifiesto que la formalización de estos fenómenos demanda integrar los distintos niveles de descripción lingüística aprovechando el potencial que ofrece NooJ, ya que dispone de recursos destinados a abordar tanto la flexión como la derivación morfológica y el léxico, sean palabras simples, compuestas o expresiones fijas, en la medida

en que comprende las cuatro clases de gramáticas de la jerarquía Chomsky Schützenberger: gramática de estado finito, gramáticas libres de contexto, gramáticas contextuales y gramáticas no restrictivas.

La conclusión hasta ahora es que los neologismos principalmente requieren de la flexibilización de las reglas morfológicas y la adaptación de los elementos que los componen, inclusive en caso de una clase cerrada como son los prefijos, de suerte que aumenta la productividad entendida más desde un punto de vista dinámico que cuantitativo.

La neología es un área de la lexicografía que desde los años ochenta provee una clasificación estable desde la perspectiva lingüística y se ha transformado en una disciplina muy abarcadora que ofrece numerosas utilidades.

Esta investigación obviamente se ampliará a fin de identificar otros tipos de neologismos en base a las descripciones formales realizadas, de modo que los resultados permitan completar las entradas lexicográficas en NooJ o en otros diccionarios electrónicos, glosarios, tesauros y aplicaciones NLP.

Además, el relevamiento de neologismos puede dar indicios de nuevos campos de conocimiento a través de un observatorio lexicográfico, como también facilitar la optimización del diseño de bases de datos y motores de búsqueda.

“El desarrollo sin precedentes de las tecnologías de la comunicación ha permitido, principalmente a partir de Internet, la producción, el acceso y el intercambio de un enorme flujo de información y conocimiento científico a usuarios de todo el mundo. No obstante, para acceder a esa gran masa de datos, se hace necesario disponer de herramientas que puedan procesarlos y que cuenten con sistemas de almacenamiento y de recuperación de la información.” (Koza, 2015:124)

En síntesis, actualizar los recursos lingüísticos informatizados es fundamental dado el auge de las humanidades digitales, la inteligencia artificial y la organización automática de contenidos.

Referencias

[1] Múgica, Nora; “*Qué decimos y hacemos cuando trabajamos el léxico*”, Rosario, Signos ELE, ISSN 1851-4863, 2014.

[2] Jackendoff, Ray; *Semantic Structures*, Cambridge, The MIT Press, 1999.

[3] Koza, Walter; *Propuesta de extracción automática de candidatos a término del dominio médico procesando información lingüística. Descripción y evaluación de resultados*, San Andrés do Río Preto, Alfa, Revista de Lingüística, (59)1, 113-127, 2015.

[4] Cabré, Teresa; *La Terminología: Teoría, metodología, aplicaciones*, Barcelona, Editorial Antártida, 1992.

Bibliografía

Biderman, María Tereza Camargo; *Teoría lingüística: teoría lexical y lingüística computacional*, São Paulo, Martins Fontes, 2001.

Bonino, Rodolfo y Rodrigo, Andrea; “Tutorial NooJ para español”, Rosario, UNR, 2015.

Cabré, Teresa; *La Terminología: Teoría, metodología, aplicaciones*, Barcelona, Editorial Antártida, 1992.

García Platero, J. M. (1998): “Los seudoprefijos. Su importancia para la creación léxica”. En Delgado León, F., M. L.

Koza, Walter; *Propuesta de extracción automática de candidatos a término del dominio médico procesando información lingüística. Descripción y evaluación de resultados*, San Andrés do Río Preto, Alfa, Revista de Lingüística, (59)1, 113-127, 2015.

Mangialavori Rasia, Ma. Eugenia & Nora Múgica, Estructuración semántica del Trayecto y visibilidad en la interfaz syn-sem, en “*En torno a las interfaces del lenguaje en la lingüística generativa*”, Vol. Temático de la SAL ed., 2016.

Múgica, Nora; “*Qué decimos y hacemos cuando trabajamos el léxico*”, Rosario, Signos ELE, ISSN 1851-4863, 2014.

Paryzek, Piotr; “Comparison of selected methods for retrieval of neologisms”, en revista *Investigationes Linguisticae*, vol. XVI, Wawrzynczyk, Adam Mickiewicz University, 2008.

Pustejovsky, James; *The Generative Lexicon*, Cambridge, MA: MIT, 1995.

Radimsky, Jan; *Position of the head in Italian n-n compounds: the case of “mirror*

compounds”, České Budějovice, *Linguistica Pragensia* 1, 2013.

Silberztein, Max; *Analyse automatique de corpus avec NooJ*, in LINX #34-35: *Hommage à Jean Dubois*, Nanterre, Université Paris X, 2014