

Inclusión de los papeles semánticos de FrameNet en DiCE

Sabela Prieto
Universidade da Coruña

The aim of our project is to enrich the actantial information of the Diccionario de Colocaciones del Español (DiCE) with labels about semantic roles. Since there are other projects which follow this line of research-such as FrameNet, we decided to include the existing information in the DiCE. Although our database focuses on collocations, it also identifies their actants and it compiles a wide-ranging corpus of predicates with their arguments. Therefore, the entry for each lexical unit contains the propositional form or argumental structure where a semantic description of the actants is given. For instance, in the entry for the noun ira we find: ira de individuo X contra individuo Y a causa del hecho Z. In this way, actants are also described semantically: ira is felt by a person against another person because of something. We are trying to add more semantical information to DiCE, by linking the actants of each lemma with the core elements of the relating frame, the same as FrameNet does. Taking up again the same example, the noun anger is set into the frame "Emotion_directed" and it presents four core elements: experiencer, expressor, stimulus, topic. These nuclear elements can be related to the actants that appear in the DiCE, giving the dictionary some detailed semantic information, regarding not only the lemma but also the elements relating to this lemma. This process of connection will allow us to label the corpus compiled in the DiCE to make the most of the data.

1. Introducción

El objetivo de nuestro trabajo es enriquecer la información actancial del *Diccionario de Colocaciones del Español* (DiCE, Alonso Ramos 2003a, 2003b) con las etiquetas de los papeles semánticos. Este etiquetado podría realizarse de forma manual pero, dado que existen proyectos que se dedican a la anotación semántica de corpus, como FrameNet (FN, Baker *et al.* 1998; Fillmore, Atkins 1998; Fillmore *et al.* 2001; Fillmore *et al.* 2003), aprovechamos la información existente para incorporarla al DiCE.

Por una parte, ambos proyectos estudian la valencia de las UL pero, para hacer referencia a los participantes de la situación lingüística, el DiCE habla de *actantes*, mientras que FN distingue *elementos del frame* (EFs).

Por otra, la base de datos del DiCE recoge un amplio corpus de predicados con sus argumentos, que será nuestro punto de partida para vincular actantes y EF.

Teniendo esto en cuenta, relacionamos los actantes de cada unidad léxica (UL) del DiCE con los elementos nucleares del *frame* evocado por la UL correspondiente en FN, con el propósito de ampliar la información semántica del DiCE. Este proceso de vinculación nos permitirá etiquetar el corpus colocacional recogido en el DiCE, para obtener un recurso léxico con una completa anotación semántica y sintáctica, que facilite el procesamiento semántico automático del lenguaje natural.

Tras explicar nuestro objetivo, exponemos el plan de trabajo. En primer lugar, presentamos cada proyecto y aclaramos sus conceptos fundamentales. En segundo lugar, analizamos e intentamos dar solución a los problemas que encontramos para vincular la información de ambos proyectos. Finalmente, anotamos de forma manual algunos ejemplos de nuestro corpus con los datos proporcionados por FN, para mostrar los resultados de dicho procedimiento.

2. Presentación del *Diccionario de Colocaciones del Español*

El DiCE¹ está especializado en el fenómeno léxico de las colocaciones, y presenta para cada lema todas las UULL con las que sostiene relaciones léxico-semánticas. Con *colocación* nos referimos a una construcción semi-idiomática en la que una palabra exige la presencia de otra para expresar un sentido dado² (Mel'čuk 1998: 30). Por ejemplo, entre *esperanza* y *abrigar* existe una relación con el significado “sentir”, al igual que entre *admiración* y *tributar*:

- (1) Abrigaba la esperanza *de que me salvaría*.
- (2) Cascales tributa admiración al valor y perseverancia de los moros sitiados.

Siguiendo la terminología de Hausmann (1979: 191) llamamos *base* al término de la colocación que es autónomo semánticamente y que selecciona, para un significado concreto, una UL específica que será denominada *colocativo*. Como ya adelantamos, el sustantivo *esperanza* selecciona el verbo *abrigar* para expresar el significado “sentir ~”, mientras que el sustantivo *admiración* selecciona el verbo *tributar* para expresar el mismo significado.

Por el momento, el leuario del DiCE se limita al campo semántico de los nombres de sentimiento, pero cuenta aproximadamente con 300 lemas que son, a su vez, base de una colocación. Como la obra está orientada a la codificación, desde la base se accede a una información lexicográfica que propone más de 3600 colocativos. Además, la entrada de cada UL incluye la forma proposicional o estructura argumental en la que se da una somera descripción semántica de los actantes. Por ejemplo, en la entrada del sustantivo *enfado*, encontramos: “enfado de individuo X contra individuo Y a causa del hecho Z”. De este modo, también los actantes aparecen caracterizados semánticamente: el enfado es sentido por un “individuo” y va dirigida hacia otro “individuo” a causa de un “hecho”.

3. Presentación de FrameNet

FN es un proyecto de Lexicografía computacional que estudia las relaciones semánticas y sintácticas entre palabras de una misma lengua.³ Para ello, sus investigadores parten de la base de que el significado de cada palabra debe ser proporcionado teniendo en cuenta el *frame*, o marco semántico en el que se encuadra (Fillmore *et al.* 2003: 23). Un frame o marco semántico puede ser definido como la representación esquemática de una situación en la que están implicados varios participantes así como roles conceptuales, que pueden ser elementos o argumentos de dicho marco (Subirats Rüggeberg 2005).

En FN aparece una breve definición de cada UL, pero la verdadera descripción se hace a través de la correspondencia entre los participantes de una situación lingüística y sus realizaciones sintácticas. Por lo tanto, FN se ocupa tanto de la valencia semántica como de la valencia sintáctica de las UULL, y por esto explicita los papeles semánticos y las expresiones sintácticas de los participantes. Dichos papeles semánticos se subdividen en: nucleares (*core*), periféricos (*peripheral*) y extratemáticos (*extrathematic*). Los elementos periféricos forman parte de la situación, pero no son específicos de un frame en concreto, pues indican circunstancias que se pueden repetir en diversos contextos, como por ejemplo: *Time*, *Place*, *Manner* o *Degree*. Los elementos extratemáticos no se incluyen en la situación del frame, sino que sirven para relacionarlo con otro estado de cosas; por ejemplo: *Circumstances* o *Reason*. Sin embargo, los elementos nucleares son necesarios para la conceptualización de cada frame, pues lo

¹ Proyecto lexicográfico en marcha, que se puede consultar en la página web <http://www.dicesp.com/>.

² Visión del concepto compartida por otros investigadores como Hausmann (1979) o Cowie (1994).

³ Esta propuesta fue creada para la lengua inglesa pero, actualmente, está en marcha el proyecto *FrameNet Español*, desarrollado por la Universidad Autónoma de Barcelona (Subirats Rüggeberg 2005; 2007; Subirats Rüggeberg *et al.* 2003). El proyecto puede consultarse en la web <http://gemini.uab.es/SFN/>.

caracterizan frente a los demás y tienen una función sintáctica nuclear (como sujeto u objeto en una oración activa).

Además, FN ofrece un corpus etiquetado semánticamente y datos sobre la frecuencia de aparición de los distintos EFs, sobre la combinación de los mismos, y sobre sus distintas realizaciones sintácticas. Toda esta información resulta enriquecedora para el DiCE y nos lleva a vincular los actantes de sus bases con los EFs.

4. Desajustes iniciales en la vinculación de actantes y elementos del frame

Para comenzar la vinculación entre actantes y EFs, debemos tener en cuenta que los sustantivos recogidos en DiCE son principalmente nombres de sentimiento, y por tanto los frames evocados deben incluir unos rasgos semánticos concretos.

En primer lugar, buscamos en FN los equivalentes en lengua inglesa a las UL de sentimiento recogidas en el DiCE y, de este modo, podemos comprobar que no todas evocan un mismo frame. Aunque nuestra primera intuición nos hizo pensar que los nombres de sentimiento se encuadrarían en un único frame, lo cierto es que para cubrir todo el lecionario del DiCE, debemos recurrir a una veintena.⁴

En segundo lugar, recopilamos los frames evocados por un mayor número de sustantivos de sentimiento recogidos en el DiCE. Sin duda destaca el frame *Emotion_directed*, evocado por cuarenta y seis bases,⁵ y lo sigue de cerca *Experiencer_subj*, evocado por veintinueve sustantivos.

Tras recopilar estos datos, encontramos como primer problema que algunas UULL del DiCE no encuentran equivalente en FN. Por ejemplo, no localizamos *conmiseration* (“conmiseración”) en FN pero, si acudimos al DiCE, vemos que *conmiseración* tiene como sustantivos cuasi-sinónimos *lástima* y *compasión*. Estos nombres, traducidos por *pity* y *compassion*, se incluyen en el frame *Experiencer_subj*. Por lo tanto, aunque *conmiseración* no esté representada en la base de datos, podremos afirmar que evoca al frame *Experiencer_subj*.

En otros casos, el sustantivo de sentimiento no aparece en FN, pero sí que lo hace el adjetivo o el verbo derivado. Cuando buscamos *anxiety* (*ansiedad*) en la base de datos, nos remite al adjetivo *anxious*, evocador del frame *Emotion_directed*. De este modo, relacionamos *anxiety* con *Emotion_directed*, aunque no se indique explícitamente en FN. Con respecto a este tema, destacamos que el frame *Subject_stimulus* remite a un gran número de adjetivos derivados de nombres de sentimiento, como *worried* (“preocupado”) o *irritated* (“irritado”). Algo similar ocurre con el frame *Experiencer_obj*, evocado por verbos como *frighten* (“asustar”) o *humiliate* (“humillar”).

A pesar de estas excepciones, en la mayoría de los casos, el equivalente inglés de las bases colocacionales está recogido en FN, y es posible llevar a cabo una enriquecedora vinculación entre actantes y EF.

⁴ *Attempt_suasion* (*aliento*), *Biological_urge* (*hastío*), *Certainty* (*confianza*), *Contrition* (*remordimiento*), *Costum* (*manía*), *Desiring* (*anhelo*), *Emotion_directed* (*abatimiento*), *Experiencer_subj* (*amor*), *Fame* (*fama*), *Hostile_encounter* (*hostilidad*), *Judgement* (*respeto*), *Medical_conditions* (*conmoción*), *Mental_property* (*desidia*), *Personal_relationship* (*amistad*), *Predicament* (*desdicha*), *Quarreling* (*disgusto*), *Regard* (*consideración*), *Remembering_experience* (*memoria*), *Station_of_entity* (*afección*), *Subject_stimulus* (*felicidad*).

⁵ El frame *Emotion_directed* es evocado por las siguientes bases del DiCE: abatimiento, aburrimiento, agonía, angustia, asombro, bochorno, chasco, cólera, consternación, contrariedad, coraje, decepción, delicia, desagrado, desánimo, desaliento, desconuelo, desengaño, desesperación, desilusión, dolor, encanto, enfado, enojo, estupor, exasperación, excitación, extrañeza, furia, furor, horror, ilusión, interés, ira, pena, perplejidad, pesadumbre, pesar, preocupación, rabia, regocijo, repulsión, sentimiento, terror, tristeza, vergüenza. Por su parte, el frame *Experiencer_subj* es evocado por: adoración, agrado, amor, antipatía, apego, arrepentimiento, bienestar, compasión, complacencia, desesperación, envidia, gozo, gusto, inquina, lástima, manía, miedo, misericordia, odio, pavor, placer, pudor, rabia, repugnancia, resentimiento, resquemor, satisfacción, temor, veneración.

5. Asignación de papeles semánticos a los actantes

Para ejemplificar el procedimiento de vinculación entre actantes y EFs, partimos del frame evocado por un mayor número de bases del DiCE, el frame *Emotion_directed*, que es definido con los siguientes términos:⁶

Los adjetivos y nombres pertenecientes a este frame describen a un “experimentador” (*Experiencer*) que siente o experimenta una reacción emocional a un “estímulo” (*Stimulus*) o a un “tema” (*Topic*). También pueden existir unas “circunstancias” (*Circumstances*) en las que tenga lugar esta reacción o una “razón” (*Reason*) que la provoque.

Aunque la definición destaque cinco EFs, en la base de datos se explicita que existen seis elementos nucleares: *Event*, *Experiencer*, *Expressor*, *State*, *Stimulus*, *Topic*. Teóricamente, las palabras que evocan un frame heredan sus elementos y, por lo tanto, todos los sustantivos relacionados con *Emotion_directed* deberían destacar estos seis elementos nucleares. Sin embargo, revisando datos hemos podido comprobar que los sustantivos de sentimiento sólo destacan cuatro de estos elementos: *Experiencer*, *Expressor*, *Stimulus*, *Topic*. El *Experiencer* es la persona o entidad que experimenta o siente las emociones. El elemento del frame *Expressor* se refiere a las expresiones que indican una parte del cuerpo, un gesto u otra expresión del *Experiencer* y que reflejan su estado emocional. El *Stimulus*⁷ es la persona, evento o estado que evoca la reacción emocional en el *Experiencer*. Y, finalmente, el *Topic* es el área general en donde ocurre la emoción, como una extensión del posible *Stimulus*.

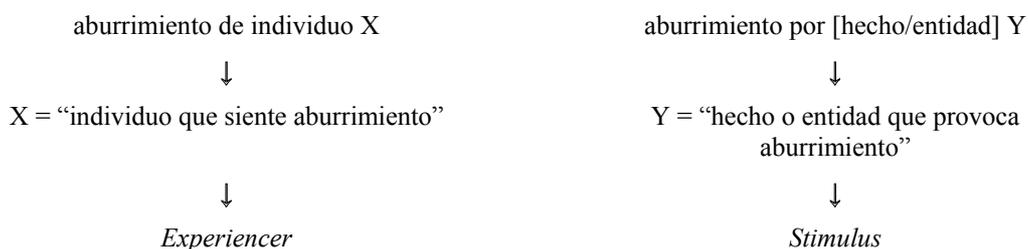
Para ilustrar dichos elementos, anotamos ejemplos del sustantivo *boredom*, tomados de FN:

- (3)
- a. Lack of employment for women and BOREDOM for men_{Experiencer} emerged as the main reasons people indulge in prostitution.
 - b. His eyes_{Expressor} glassy with BOREDOM, he greeted me like a drowning man clutching at a lifebelt.
 - c. BOREDOM with politics_{Stimulus} was a state of mind they could only comprehend intellectually.
 - d. So tell the world, and its BOREDOM about your troubles_{Topic} will heal you.

Sin embargo, según los datos del DiCE, el sustantivo *aburrimiento*, equivalente español de *boredom*, presenta la siguiente estructura actancial:

aburrimiento de individuo X por [hecho/entidad]Y

Teniendo en cuenta los significados aportados al describir los EFs, apelamos a nuestra intuición semántica para proponer la siguiente correspondencia:⁸



⁶ Definición traducida al español.

⁷ Esta definición de *Stimulus* como “persona, evento o estado que evoca un sentimiento” resulta un tanto confusa, por lo que la reformulamos como “persona, evento o estado hacia la que se dirige un sentimiento”.

⁸ Una vinculación similar, entre papeles semánticos y actantes de la unidad léxica aparece en Alonso Ramos 2003c.

De este modo, al actante X de *aburrimiento* le asignamos el papel semántico *Experiencer*, y al actante Y, el papel semántico *Stimulus*. Por lo tanto, sólo dos elementos nucleares del frame *Emotion_directed* resultan pertinentes para el sustantivo *aburrimiento*. Y, no sólo eso, pues tras revisar todas las bases colocacionales del DiCE, podemos afirmar que esta vinculación es apropiada para la mayoría de sustantivos que evocan el frame *Emotion_directed*.

Sin embargo, frente a la generalidad de bases biactanciales, el nombre *ira* supone una excepción, con una estructura argumental de tres actantes:

ira de individuo X contra individuo Y a causa del hecho Z

Como en el caso anterior, el actante X de *ira* se corresponde con el papel semántico *Experiencer*, y el actante Y con el papel semántico *Stimulus*. Pero no existe ningún EF nuclear que pueda ser asignado al tercer actante, Z de ira = “hecho que causa la ira”. Como ningún elemento nuclear expresa un sentido similar a Z, repasamos los EFs periféricos y extratemáticos, y en este último grupo encontramos el elemento *Reason*, definido como “causa que provoca una emoción”. Así, completamos el siguiente cuadro de equivalencias:

X	Y	Z
Experiencer	Stimulus	Reason

Esta primera vinculación nos muestra que la relación entre actantes y EFs no es tan sencilla como podría parecer a simple vista, y que tendremos que recurrir a nuestro conocimiento lingüístico para solventar las irregularidades existentes entre ambos proyectos.

Para profundizar en el proceso de asignación semántica, acudimos al segundo frame más evocado por los nombres del DiCE: *Experiencer_subj*. Este frame reconoce los mismos elementos nucleares que el anterior, y simplemente introduce en lugar del *Stimulus* el elemento *Content*, un tipo de estímulo que funciona como objeto directo de los verbos en los que el *Experiencer* es sujeto.

En cuanto a la definición del frame, las palabras recogidas en él son «emociones del *Experiencer* con respecto a un *Content*».⁹ Siguiendo esta descripción, entendemos que los dos frame son muy similares, pues comparten elementos nucleares y sus definiciones sólo se diferencian en la perspectiva.

Al igual que pasaba en la anterior vinculación de bases colocacionales con EFs, la estructura argumental de estos nombres presenta fundamentalmente dos actantes, que se corresponden semánticamente con el *Experiencer* y el *Content*. Como ejemplo, los argumentos del sustantivo *placer* en «placer de individuo X por hecho Y», se corresponden respectivamente con el *Experiencer* y el *Content*.

Para probar dicha vinculación, proponemos algunos ejemplos recogidos en el corpus del DiCE, anotados con papeles semánticos de FN:

- (4) [El silencio]_{Stimulus} *provoca el aburrimiento de* [los talleres honrados]_{Experiencer}
- (5) [Los ministros y personajes políticos]_{Stimulus} *le* _{Experiencer} *despertaban una irrefrenable antipatía y un profundo aburrimiento.*
- (6) [Un descuido]_{Reason} *desató la ira del jefe* _{Experiencer} *hacia mí* _{Stimulus}
- (7) *En Gran Bretaña* [la prensa]_{Stimulus} *es objeto de la ira popular* _{Experiencer} [debido al presunto papel de los paparazzi]_{Reason}
- (8) En la clase de Psicomotricidad [el niño]_{Experiencer} *vive el placer del juego* _{Content} *en todas sus expresiones.*
- (9) Luis_{Experiencer} *encuentra un gran placer* [en leer y estudiar]_{Content}

⁹ Definición traducida al español.

De este modo verificamos que hemos propuesto una asignación adecuada de papeles semánticos, pues ha servido para anotar diferentes nombres de sentimiento. En cuanto a los frames estudiados, observamos que sus significados son tan próximos que se podrían fusionar, ya que la única diferencia se presenta en la nomenclatura de los estímulos, porque el *Stimulus* etiqueta cualquier estímulo en general, mientras que el *Content* comparte su significado y difiere únicamente en su función sintáctica. Esta distinción de papeles semánticos basada en diferencias sintácticas resulta poco productiva para nuestro trabajo, pues no afecta directamente a la semántica.

Finalmente, cabe añadir que en nuestra vinculación utilizaremos en mayor medida los elementos nucleares del frame, pues suelen coincidir con los actantes de las colocaciones, aunque no debemos olvidar los elementos periféricos y extratemáticos, porque nos pueden ayudar a cubrir lagunas (como ocurría en el caso del tercer actante de *ira*).

Con estas apreciaciones concluimos que los papeles semánticos de FN pueden relacionarse perfectamente con los actantes del DiCE, simplemente aplicando unos mínimos reajustes ocasionados por la diferencia idiomática entre el inglés y el español, así como por los distintos marcos teóricos que dieron lugar a ambos proyectos.

6. Conclusión

Este proceso de vinculación que nos ha permitido asignar papeles semánticos a los actantes del DiCE proporciona una caracterización semántica muy precisa de los participantes de la situación. Recordemos que el DiCE tan sólo aporta los rasgos semánticos inherentes de las palabras que pueden ocupar el lugar de cada actante. Por ejemplo, un nombre que designa a una persona poseerá la etiqueta “individuo” en todos los casos, tanto si realiza una acción como si la padece:

(10) Juan siente ira hacia Manuel

En el actual etiquetado del DiCE tanto *Juan* como *Manuel* están definidos como “Individuo”, mientras que, con los nuevos papeles semánticos, adquieren rasgos opuestos: *Juan* será el *Experienter* o experimentador y *Manuel* será el *Stimulus* o la víctima.

Con este sencillo ejemplo comprobamos la importancia de la precisión semántica a la hora de recuperar la información. Por este motivo, nuestro próximo proyecto se centrará en la anotación manual de un corpus colocacional con papeles semánticos. La intención de este futuro trabajo será conseguir unas pautas de anotación tales que nos permitan llegar a un tratamiento semiautomático de los datos. Una vez alcanzado este objetivo, obtendremos un corpus colocacional anotado semánticamente con el que podremos realizar consultas de gran utilidad en el procesamiento semántico automático del lenguaje natural.

Referencias bibliográficas

- Alonso Ramos, M. (2003a). "Hacia un Diccionario de Colocaciones del español y su codificación". En Martí, M. A. *et al.* (eds.). *Lexicografía computacional y semántica*. Barcelona: Edicions de l'Universitat de Barcelona. 11-34.
- Alonso Ramos, M. (2003b). "Presentación del *Diccionario de colocaciones y marcadores del español: estructura y objetivos*". *Actas del IV Congreso de Lingüística General* (2000). Cádiz: Universidad de Cádiz.
- Alonso Ramos, M. (2003c). "Éléments du frame vs. Actants de l'unité lexicale". *The First International Conference on Meaning-Text Theory*. Paris.
- Baker, C.; Fillmore, C. J.; Lowe, J. B. (1998). *The Berkeley FrameNet Project, Proceedings of COLING-ACL*. Montréal. 86-90.
- Cowie, A. P. (1994). "Phraseology". En Asher, R. E.; Simpson, J. M. Y. (eds.). *The Encyclopedia of Languages and Linguistics* 6. Oxford: Pergamon Press. 3168-3171.
- Fillmore, C. J.; Atkins, B. T. S. (1998). "FrameNet and Lexicographic Relevance". *Proceedings of the First International Conference on Linguistic Resources and Evaluation*. Granada. 417-423.
- Fillmore, C. J.; Wooters, C.; Baker, C. (2001). "Building a Large Lexical Database Which Provides Deep Semantics". En Tsou, B.; Kwong, O. (eds.). *Proceedings of the 15th Pacific Asia Conference on Language, Information and Computation*. Hong Kong.
- Fillmore, C. J.; Johnson, C. H.; Petruck, M. (2003). "Background to FrameNet". *International Journal of Lexicography* 16 (3). 235-250.
- Hausmann, F. J. (1979). "Un dictionnaire des collocations est-il possible?". *Travaux de littérature et de linguistique de l'Université de Strasbourg* 17 (1). 187-195.
- Mel'čuk, I. (1998). "Collocations and Lexical Functions". En Cowie, A. P. (ed.). *Phraseology. Theory, Analysis, and Applications*. Oxford: Clarendon Press. 23-53.
- Subirats Rüggeberg, C.; Petruck, M. R. L. (2003). "Surprise: Spanish FrameNet!". *Proceedings of the International Congress of Linguists*, Praga. [Disponible en <http://gemini.uab.es/SFNpub/papers/subirats-petruck.pdf>].
- Subirats Rüggeberg, C. (2005). "FrameNet español. Una red semántica de marcos conceptuales". En Serra, E.; Wotjak, G. (eds.). *Cognición y percepción lingüísticas*. Valencia: Universidad de Valencia y Universidad de Leipzig. 182-196.
- Subirats Rüggeberg, C. (2007). "Relaciones semánticas entre marcos en FrameNet Español". En Cuartero, J.; Emsel, M. (eds.). *Vernetzungen. Bedeutung in Wort, Satz und Text. Festschrift für Gerd Wotjak zum 65. Geburtstag*. Frankfurt: Peter Lang. 357-366.