

145 LA ANOTACIÓN DE PAPELES SEMÁNTICOS EN ADESSE.
EL PROBLEMA DE LAS CONSTRUCCIONES CON PARTICIPANTE
DESDOBLADO

Gael Vaamonde
Universidade de Vigo
gaelv@uvigo.es

1. Cuestiones generales

En la elaboración de corpus anotados con información lingüística relevante, uno de los niveles de análisis más demandados es el que concierne a la semántica y, muy especialmente, a la descripción de las relaciones semánticas entre el predicado y sus argumentos. La importancia de contar con recursos lingüísticos que ofrezcan información de este tipo se vuelve evidente si reparamos un momento en las limitaciones que conlleva una anotación exclusivamente sintáctica. Repárese en los ejemplos siguientes:

- (1) a. *La crisis* [0] *aumentó la venta de lotería* [1]
b. *La venta de lotería* [1] *aumentó*
- (2) a. *Del Bosque* [0] *sustituyó a Torres* [1] *por Llorente* [2]
b. *Del Bosque* [0] *sustituyó a Torres* [1]
c. *Llorente* [2] *sustituyó a Torres* [1]

El análisis sintáctico de los ejemplos de (1) reflejará que *la venta de lotería* funciona como ODIR (objeto directo) en (1a) y como SUJ (sujeto) en (1b), pero nada nos dirá sobre el hecho de que *la venta de lotería* mantiene la misma relación semántica con el verbo *aumentar* en ambas cláusulas. Por otro lado, nada diferencia sintácticamente a los ejemplos recogidos en (2b) y (2c), que responden al esquema transitivo SUJ-PRED-ODIR. Semánticamente, sin embargo, reconocemos un contraste entre los argumentos que funcionan como SUJ en cada construcción, diferencia que viene evidenciada por la posibilidad de expresar ambos argumentos en una única cláusula (2a).

Lo que demuestran ejemplos como los anteriores es que la sintaxis no permite dar cuenta de todo el significado implicado por una expresión lingüística. Debemos considerar igualmente las relaciones semánticas que se establecen entre el verbo y sus argumentos. En concreto, la anotación de papeles semánticos permite recoger las equivalencias de significado entre construcciones parcialmente diferentes (1), así como de divergencias de significado entre construcciones parcialmente similares (2).

Como un intento de responder a estas y otras cuestiones, han surgido en los últimos años varios proyectos de investigación que comparten entre sus intereses la descripción de las relaciones semánticas entre el verbo y sus argumentos. Para el caso del inglés, es indispensable citar los casos de PropBank (Palmer et. al., 2005) y de FrameNet (Ruppenhofer et al., 2006). En el ámbito hispánico, la anotación de papeles semánticos sobre datos de corpus se ha llevado a cabo al menos por cuatro proyectos de investigación: AnCora (Martí et al., 2007) SenSem (Castellón et al., 2006), Spanish FramNet (Subirats, 2004) y ADESSE (García-Miguel y Albertuz, 2005).

El proyecto de investigación ADESSE (Base de datos de verbos, Alternancias de Diátesis y Esquemas Sintáctico-Semánticos del Español), que se lleva realizando desde hace ya varios años en la Universidad de Vigo, pretende completar una base de datos relacional con información sintáctica y semántica relevante sobre los más de 3.000 verbos registrados en un corpus del español.

El objetivo principal de este proyecto es obtener una caracterización detallada de la valencia verbal en español a partir de datos de corpus, mediante la anotación de información sintáctico-semántica sobre la estructura argumental de cada verbo registrado. Además, esta base de datos, que cuenta con una versión web (<http://adesse.uvigo.es/data/>), pretende servir como un recurso lingüístico útil para el estudio empírico acerca de la interacción entre significado verbal y significado construccional.

2. La anotación de papeles semánticos en ADESSE

Un requisito previo al anotar argumentos verbales con papeles semánticos es el de delimitar el conjunto de roles que se van a aplicar en la anotación. Frente al consenso más o menos efectivo que suele existir en otros niveles de análisis, el reconocimiento de

un conjunto limitado y único de papeles semánticos parece un reto imposible, dada la larga lista de propuestas existentes (cf. Levin y Rappaport 2005: 38-44, para una revisión crítica de las propuestas más importantes en este sentido). Esta heterogeneidad se refleja en el número de roles contemplados por algunos de los proyectos de investigación ya comentados. En AnCora se adopta un inventario de 20 roles distintos, mientras que en SenSem el número de papeles anotados asciende a 32 opciones.

En las relaciones semánticas que se dan entre los verbos y los argumentos verbales podemos observar diferentes grados de abstracción. En un extremo, cada verbo particular «defines a distinct set of participant roles that reflect its own unique semantic properties -e.g. the subject of *bite* is a slightly different kind of agent from the subject of *chew*-» (Langacker, 1991: 284). En el extremo contrario es en el que debemos enmarcar algunas propuestas que apelan a roles semánticos generalizados, resultado de un proceso de categorización máxima. Es el caso de los macroroles de *Actor* y *Undergoer* (Van Valin & LaPolla, 1997) o de los conceptos de *Proto-Agent* y *Proto-Patient* (Dowty, 1991). En un nivel intermedio nos encontramos con la lista de roles más usuales, conocidos frecuentemente como roles temáticos o relaciones temáticas: agente, paciente, tema, instrumento, lugar, etc.

Muchas veces, las discrepancias existentes entre unas propuestas y otras dependen precisamente del grado de generalización en el que se pretenda basar la anotación. Y este grado de generalización está subordinado a su vez a otros factores, como la finalidad del producto, el método de anotación, el tamaño del corpus, etc. Por lo que respecta a ADESSE, el interés fundamental estriba en la estructura argumental de los verbos y en su documentación empírica a partir de los datos de un corpus. De ahí que la anotación de papeles semánticos se muestre como una tarea básica para este proyecto, ya que permite recoger la relación entre los diferentes esquemas sintácticos de un verbo mediante la correspondencia entre funciones sintácticas y funciones semánticas. Además, es importante destacar que la finalidad de ADESSE es principalmente descriptiva y que la anotación de papeles semánticos se ha llevado a cabo de manera manual para cada esquema sintáctico de cada verbo registrado. Aspectos como estos pueden explicar por qué ADESSE aporta una granularidad de análisis bastante mayor que la reflejada en otros recursos lingüísticos similares, pero

que persiguen una utilidad más computacional o están basados en una metodología de anotación automática o semiautomática.

2.1. El proceso de anotación

Lo primero que debemos tener en cuenta es que en ADESSE no existe un único inventario de papeles semánticos que sirva como referencia para anotar argumentos. En realidad, tampoco se parte de un conjunto de papeles preestablecido. En ADESSE se adopta más bien una estrategia inductiva, tomando el significado verbal como punto de partida y describiendo participantes y tipos de participantes en un proceso de generalización cada vez mayor. A partir de esta idea, la definición de roles semánticos se ha llevado a cabo teniendo en cuenta hasta tres niveles de análisis diferentes:

2.1.1. Argumentos específicos del verbo

En primer lugar, se describe el conjunto de argumentos que son posibles con cada verbo (sentido verbal), esto es, el total de participantes previstos por ese verbo y cuya consideración conjunta permite dar cuenta de la escena evocada en su totalidad. Es lo que llamamos el potencial valencial del verbo (cf. Vater, 2003). Por ejemplo, el marco conceptual de *contar*, con el sentido de “narrar, relatar”, se puede describir mediante la consideración de cuatro argumentos: A1: “el que cuenta”, A2: “aquello que se cuenta”, A3: “aquel a quien se le cuenta” y A4: “aquello sobre lo que se cuenta”. El siguiente ejemplo recoge de manera explícita el total de argumentos considerados:

(3) Zambrini (A1) me (A3) **ha contado** tales hazañas (A2) de usted (A4)...[SON:229]

Siguiendo propuestas como la de PropBank, cada argumento recibe un número correlativo, desde A0 en adelante. Generalmente, no se postulan etiquetas específicas para cada argumento de cada verbo. De ser así, habría que reconocer tantas etiquetas de rol diferentes como huecos argumentales sean posibles para el total de verbos registrados en el corpus, lo que nos daría actualmente un total de 9,758 etiquetas de papel semántico para 4,016 sentidos verbales (una media de 2,4 argumentos por verbo). Se hace necesario un proceso de generalización que facilite la tarea de anotación. En

ADESSE, este proceso de generalización viene dado a partir de la clasificación semántica de cada verbo y de los roles semánticos asociadas a estas clases verbales.

2.1.2. Roles específicos de la clase

Junto con la anotación de papeles semánticos, una de las tareas fundamentales en ADESSE ha consistido en la clasificación semántica de verbos. La clasificación semántica de ADESSE obedece a una estructura jerárquica en la que se reconocen diferentes niveles. En el nivel superior, se distinguen seis grandes grupos o macroclases, basados en gran medida en los tipos de procesos reconocidos en Halliday (1985). En un segundo nivel, cada macroclase se divide en varias clases, que representan dominios semánticos generales. La Tabla 1 recoge algunas de las divisiones propuestas en la clasificación verbal de ADESSE en relación con los dos niveles más generales¹¹⁵³:

Tabla 1: Clases y macroclases en ADESSE

MACROCLASE	CLASE	Verbos	Ejemplos
1 MENTAL	11 Sensación	274	<i>gustar</i>
	12 Percepción	124	<i>ver</i>
	13 Cognición	191	<i>saber</i>
2. RELACIONAL	21 Atribución	280	<i>ser</i>
	22 Posesión	198	<i>tener</i>
3 'MATERIAL'	31 Espacio	703	<i>ir</i>
	32 Cambio	951	<i>abrir</i>
	33 Otros hechos	448	<i>tocar</i>
	34 Comportamiento	379	<i>reír</i>
4 VERBAL	41 Comunicación	306	<i>decir</i>
	41 Valoración	82	<i>criticar</i>
5 EXISTENCIAL		204	<i>haber</i>
6 MODULACIÓN		165	<i>hacer</i>

Cada clase de verbos está asociada a un conjunto determinado de papeles semánticos, aquellos que se entienden como prototípicos o inherentes al dominio conceptual evocado por los verbos que integran la clase. Algunas de las etiquetas utilizadas para estos papeles semánticos pueden coincidir con los nombres de las relaciones temáticas más usuales. Sin embargo, la elección de estas etiquetas se ha

¹¹⁵³ Este trabajo se ha realizado gracias a la financiación recibida por el Ministerio de Educación y Ciencia "Ampliación de la BD léxica y el corpus sintáctico-semántico de semántica oracional del español SenSem" (HUM2007-65267).

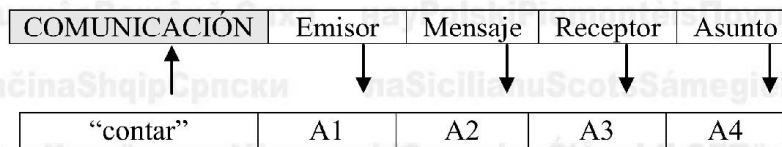
llevado a cabo teniendo en cuenta dos factores principales: especificidad (i.e. coherencia con la clase semántica correspondiente) y transparencia (i.e. adecuación descriptiva). En la Tabla 2 se recogen algunos de estos roles asociados a clases semánticas:

Tabla 2: Papeles semánticos asociados a clases semánticas en ADESSE

Clase	A0	A1	A2	A3	A4
Sensación		Experimentador	Estímulo		
Percepción	Iniciador	Perceptor	Percibido		
Cognición		Conocedor	Contenido		
Atribución	Asignador	Entidad	Atributo		
Posesión		Poseedor	Posesión		
Transferencia	Donante	Poseedor-final	Posesión		
Desplazamiento	Iniciador	Móvil	Origen	Dirección	Trayecto
Emisión		Emisión	Emisión		
Actividad		Agente	Actividad		
Comportamiento		Actor	Manera		
Comunicación		Emisor	Mensaje	Receptor	Asunto

Los primeros, los argumentos específicos del verbo, heredan por defecto las etiquetas correspondientes a los roles específicos de la clase a la que pertenezca ese verbo. Por ejemplo, el sentido de *contar*, ya comentado, está clasificado como un verbo de comunicación, por lo que sus argumentos heredan el conjunto de papeles semánticos que han sido propuestos para todos los verbos de comunicación, a saber:

Figura 1. Papeles semánticos heredados de la clase de comunicación



Actualmente, se han distinguido un total de 196 papeles asociados a clases semánticas. Puesto que la clasificación semántica de ADESSE es jerárquica, los roles específicos de la clase definen tipos de participantes en diferentes niveles de generalización. Por ejemplo, la clase de verbos de cambio se subdivide en tres subclases: verbos de creación (e.g. *crear*, *producir*), verbos de modificación (e.g. *romper*, *abrir*) y verbos de destrucción (e.g. *destruir*, *eliminar*). Como se observa en la figura siguiente, cada subclase verbal está asociada a diferentes roles semánticos. Sin embargo, la clase verbal de nivel más esquemático neutraliza estas oposiciones

semánticas, abstrayendo las propiedades comunes de los diferentes roles mediante la consideración de un Agente y un Paciente ('01' y '02' representan el primer y el segundo argumento del verbo, respectivamente):

Figura 2. Roles semánticos y niveles de generalización



2.1.3. Roles semánticos generalizados

Ya hemos comentado cómo los roles específicos del verbo reciben una numeración correlativa desde A0 en adelante y cómo las etiquetas utilizadas para describir los argumentos verbales suelen heredarse directamente de las clases semánticas en diferentes niveles de jerarquización. Un tercer paso que hemos contemplado en ADESSE se centra en conseguir una representación máximamente esquemática que agrupe los diferentes roles asociados a clases semánticas.

No obstante, en este nivel debemos afrontar un gran conjunto de variables heterogéneas y pertenecientes a diferentes dominios semánticos. Por tanto, si queremos buscar generalizaciones a partir de este conjunto debemos tener en cuenta otros aspectos que van más allá de la semántica. Es necesario considerar las semejanzas sintáctico-semánticas que permiten resumir en solo unas pocas opciones todo el conjunto de variables posibles.

En este sentido, la numeración correlativa de los argumentos verbales, además de describir el potencial valencial de cada verbo, nos permite representar generalizaciones basadas en posiciones argumentales básicas. En ADESSE, esta numeración se ha llevado a cabo teniendo en cuenta las siguientes correspondencias: A0: iniciador o causante; A1: primer argumento de un verbo; A2: segundo argumento de un verbo. La tabla siguiente recoge las correspondencias entre las posiciones argumentales A0-A1-A2 y algunos roles específicos de la clase postulados en ADESSE:

Tabla 3. Correspondencias entre roles semánticos y A0-A1-A2 en ADESSE

	A0	A1	A2
Sensación		Experimentador	Estímulo
Percepción	Iniciador	Perceptor	Percibido
Cognición		Conocedor	Contenido
Atribución	Asignador	Entidad	Atributo
Poseción		Poseedor	Poseción
Transferencia	Donante	Poseedor-final	Poseción
Desplazamiento	Iniciador	Móvil	
Localización	Iniciador	Móvil	
Modificación	Agente	Afectado	
Competición		Competidor	Antagonista
Existencia	Iniciador	Existente	
Tiempo	Iniciador	Existente	Duración

3. Las construcciones con participante desdoblado

3.1. Desdoblamiento de participantes

En la descripción de la estructura argumental de los verbos, la asignación de funciones sintácticas y funciones semánticas para cada argumento suele permitir normalmente una correspondencia de uno a uno. Esta correspondencia, además, se extiende al conjunto de participantes implicados en la situación evocada por el verbo, de tal forma que cada participante aparece codificado en la cláusula por una función sintáctica determinada, y está asociado a un papel semántico concreto. Por ejemplo:

(4) *David observaba los chopos* [JOV:173]

Argumentos	A1	A2
Func. sintácticas	SUJ	ODIR
Papeles semánticos	Perceptor	Percibido
Participantes	P1: David	P2: los chopos

(5) *El viejo entrega el niño a Renato* [SON:123]

Argumentos	A0	A1	A2
Func. sintácticas	SUJ	ODIR	OIND
Papeles semánticos	Donante	Poseción	Poseedor-final
Participantes	P1: el viejo	P2: el niño	P3: Renato

Existen algunos casos, no obstante, en los que esta correspondencia se rompe en uno u otro sentido. Por ejemplo, aunque muchos verbos que expresan algún tipo de competición entre dos entidades admiten una construcción biargumental, como en (6),

que permite la expresión de un competidor y un antagonista como SUJ y ODIR, respectivamente, también es posible la construcción con un solo argumento para ambos papeles semánticos, lo que produce casos de amalgama como los de (7):

(6) *La americana se enfrenta en cuartos a Gabriela Sabatini* [2VOZ:047]

Argumentos	A1	A2
Func. sintácticas	SUJ	CPREP(a)
Papeles semánticos	Competidor	Antagonista
Participantes	P1: la americana	P2: Gabriela Sabatini

(7) *Pronto se enfrentarían dos fuerzas opuestas* [TIE:074]

Argumentos	A1
Func. sintácticas	SUJ
Papeles semánticos	Competidor + Antagonista
Participantes	P1 + P2

Un caso hasta cierto punto opuesto al anterior se da con las que llamaremos construcciones de participante desdoblado (CPD). Si en los casos de amalgama debemos enfrentarnos al hecho de que un mismo argumento incluya dos participantes en la situación designada, las CPD se caracterizan por presentar argumentos adicionales, esto es, argumento no contemplados por el significado léxico del verbo, sin necesidad de incluir más participantes en la situación designada. Típicamente, esto sucede cuando diferentes partes o facetas de una sola entidad adquieren la suficiente prominencia como para ascender a constituyentes directos de la cláusula. Muchos verbos de contacto, que denotan básicamente una relación entre dos argumentos, un contactante y una entidad contactada (8a), admiten construcciones alternativas con un dativo posesivo (8b) o un locativo partitivo (8c), que expresan un NPC (nombre de parte del cuerpo) y su POS (poseedor) en constituyentes separados (cf. García-Miguel, 2007):

(8) a. *Golpea el pecho de su marido* [CAI:073]

b. *... sobre todo cuando me dijo golpeándome la espalda* [GLE:062]

c. *Dos desconocidos lo golpearon en la cabeza* [1VOZ:16]

A aquellas construcciones en las que se expresa un argumento adicional respecto del esquema conceptual básico del verbo sin que con ello aumente el número de

participantes en el evento es a lo que llamaremos CPD. Tomando el ejemplo (8b) como referencia, podemos representar esquemáticamente las correspondencias entre unos niveles y otros de la siguiente manera:

Argumentos	A1	A2	A3
Func. sintácticas	SUJ	ODIR	OIND
Papeles semánticos	Contactante	Contactado	Contactado'
Participantes	P1	P2	

Conviene señalar que este fenómeno no se circunscribe a verbos de contacto ni a NPCs, sino que podemos encontrar participantes desdoblados con verbos que no perfilan el contacto como componente semántico básico (9), o incluso con verbos que están claramente fuera de dicho dominio semántico y donde el argumento adicional producto del desdoblamiento remite más bien a una faceta o aspecto concreto de otra entidad (10):

(9) a. *El viejo le mira las manos* [SON:115]

b. *Se volvió mirándome en los ojos* [GLE:147]

(10)a. *...el hecho de que tales criterios se los imiten todos* [RAT:115]

b. *El mal sabor de ese recuerdo le cambió la vida* [CRO:68]

No obstante, en los párrafos que siguen nos ceñiremos únicamente a las CPD con verbos de contacto y, por lo tanto, a las diferentes posibilidades que se pueden contemplar para anotar alternancias construccionales como las ilustradas en (8)

3.2. Planteamiento del problema

Como hemos comentado, el caso más característico de CPD es el que se produce con verbos de contacto y con NPCs. Ello nos ofrece la posibilidad de desdoblar el participante que funciona como ODIR del esquema transitivo, es decir la entidad contactada, tal como queda reflejado en los ejemplos de (11):

(11)a. *Ana Besa a Teo que está algo rígido* [AYE:071]

b. *De repente, se arroja a sus pies, le besa las manos* [ZOR:029]

c. *Victor Joven se acerca a ella y sin más la besa en los labios* [HOM:030]

Además, hay que tener en cuenta que el participante que realiza la acción del contacto puede presentarse igualmente de forma desdoblada, generalmente mediante un CPREP (complemento preposicional) introducido por *con*:

(12)a. *El comisario golpeó la madera con los nudillos* [LAB:016]

b. *El comisario golpeó la madera*

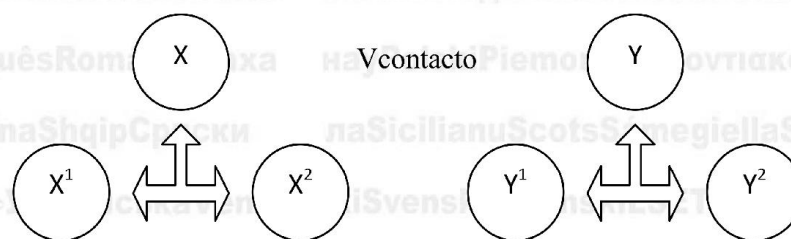
c. *Los nudillos del comisario golpearon la madera*

Esto nos lleva en última instancia a contemplar la posibilidad de tener hasta cuatro constituyentes diferentes para un verbo que rige necesariamente dos argumentos, simplemente con dos participantes desdoblados a partir del esquema transitivo básico del verbo:

(13) *Le golpea amistosamente en un biceps con el puño cerrado* [CIN:026]

Gráficamente, podríamos representar dicha posibilidad de acuerdo con el esquema de la figura siguiente:

Figura 3. Desdoblamiento de SUJ y ODIR



Conviene señalar que este tipo de construcciones no suponen realmente un problema de etiquetación para otros proyectos de investigación. De hecho, las etiquetas que encontramos en cualquier inventario o lista de papeles semánticos podrían servir para anotar las construcciones que venimos comentando. Así, en los ejemplos siguientes los roles de Beneficiario (14a), Locativo (14b) e Instrumento (14c) describen la relación semántica que se establece entre el verbo y los argumentos destacados en negrita,

mientras que los roles de Agente y Paciente completan las funciones de SUJ y ODIR en cada caso:

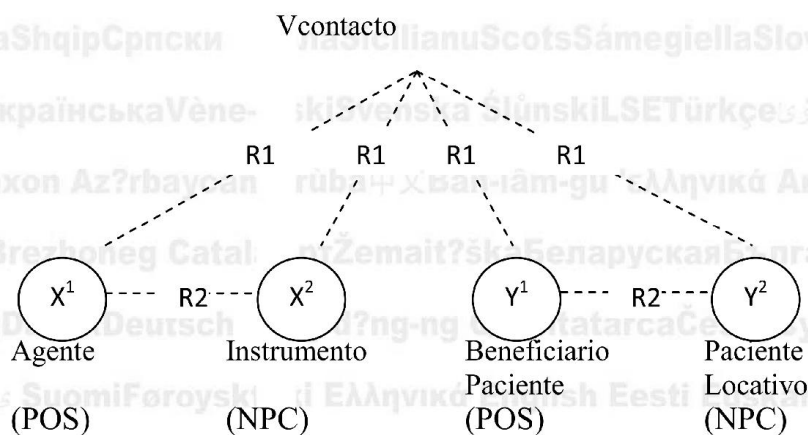
- (14)a. *Juan* (Agente) *le* (**Beneficiario**) *golpeó* *la cabeza* (Paciente)
 b. *Juan* (Agente) *lo* (Paciente) *golpeó en la cabeza* (**Locativo**)
 c. *Juan* (Agente) *lo* (Paciente) *golpeó con la mano* (**Instrumento**)

Por ejemplo, una solución similar a la expuesta en (14) es la que se adopta en varios ejemplos consultados en el corpus anotado de AnCorá, como los recogidos a continuación:

- (15)a. *Él* (A0: Agente) *llega a tirarla sobre la cama* y *le* (A2 Beneficiario) *toca la cara* (A1: Paciente) *con la mano* (AM: Adverbial)
 b. *Marcarles un número* (A1: Paciente) *en el brazo* (AM: Locativo)

No obstante, si lo que buscamos es un análisis más detallado de las relaciones semánticas entre el verbo y sus argumentos, las etiquetas que acabamos de comentar se muestran insuficientes para describir las CPD. Y esto es así porque roles como los de Beneficiario, Locativo o Instrumento no permiten dar cuenta de la relación posesiva que se establece entre el NPC y su POS (R2 en la Figura 4), sino únicamente de la relación semántica que mantienen estos últimos con el predicado verbal (R1 en la Figura 4):

Figura 4. Relaciones semánticas en las CPD



En consecuencia, nótese que en AnCora, por ejemplo, se utiliza la misma etiqueta de rol semántico para argumentos como los destacados en (16a) y (16b). Sin embargo, en el primero de ellos se establece una única relación semántica verbo-argumento (R1), mientras que en el segundo existe una relación semántica adicional argumento-argumento (R2), de naturaleza posesiva, que no aparece anotada de ninguna forma. Y lo mismo podemos decir de los argumentos destacados en (17):

(16) **Beneficiario**

- a. *El presidente también ofrece muchos regalos **a los dignatarios***
- b. *Él llega a tirarla sobre la cama y **le toca la cara***

(17) **Locativo**

- a. *La Fundació projecta construir 40 pisos **en la calle de Buxadé***
- b. *Por eso se molestaban en marcarles un número **en el brazo***

Por tanto, la anotación de papeles semánticos para este tipo de construcciones nos lleva necesariamente a contemplar tres cuestiones: a) definir el número de papeles semánticos necesarios para etiquetar el grupo de verbos de contacto, pues la sola consideración de un agente-contactante y un paciente-contactado se vuelve claramente insuficiente, b) establecer las correspondencias más apropiadas en la asignación de cada papel semántico diferenciado a las funciones sintácticas de los esquemas, tanto para la construcción transitiva básica como para las construcciones con participante desdoblado, y c) contemplar la doble relación semántica que implican este tipo de construcciones, la que se establece entre el verbo y los argumentos y la relación posesiva que se establece entre los dos argumentos implicados en las construcciones con participante desdoblado.

3.3. *La anotación de las CPD en ADESSE*

En respuesta a estas cuestiones, en ADESSE hemos abordado la anotación semántica de las CPD teniendo en cuenta los aspectos siguientes:

- I. Los verbos que están clasificados en la clase de contacto heredan necesariamente dos papeles semánticos nucleares, esto es, dos roles prototípicos que son inherentes a todo verbo cuyo significado implica el contacto entre dos entidades: un

contactante (CNTE) y un contactado (CTDO). Estos papeles se mantienen inalterados a lo largo de las diferentes alternativas construccionales de naturaleza transitiva que presentan los verbos de contacto, lo que viene a significar que las funciones semánticas de CNTE y CTDO se corresponden con las funciones sintácticas de SUJ y ODIR para dichas alternativas con esos verbos.

II. Como consecuencia de lo expuesto en el punto anterior, la animación de los participantes expresados mediante las funciones de SUJ y ODIR no altera su anotación semántica. Esto quiere decir que el CNTE incluye SUJs animados (18a) e inanimados (18b), y que el CTDO incluye OBJs animados (19a) e inanimados (19b).

(18)a. *Le ha golpeado **un hombre** que está a mi servicio* [PAS:122]

b. *Las **palabras** lo golpeaban* [JOV:140]

(19)a. *Entonces zarandé, esposó y golpeó **a Antonio González*** [VOZ:25]

b. *Golpea **el pecho de su marido*** [CAI:073]

La anotación de papeles semánticos para las CPD podría basarse, de hecho, en la animación de los participantes, e informar sobre ella en función del rol semántico asociado a cada argumento. Por ejemplo, podríamos haber reservado el CTDO únicamente para aquellos participantes animados que reciben la acción verbal, bien sean ODIR (i.e. *Juan golpea **a Pedro***) o OIND (*Juan le golpea la cabeza **a Pedro***). Sin embargo, la animación es un rasgo que ya ha sido anotado para todos los argumentos del corpus, por lo que esta última estrategia resultaría redundante en ADESSE. Además, perderíamos con ello la correspondencia entre rol semántico y función sintáctica para las construcciones transitivas, de acuerdo a lo expuesto en el punto anterior.

III. Puesto que los papeles semánticos de CNTE Y CTDO permanecen inalterados a lo largo de todas las variantes transitivas, el resto de papeles semánticos necesarios para describir las CPD se reservan para anotar los argumentos adicionales: argumentos posibles pero que no son inherentes a los verbos de contacto, sino que surgen precisamente a partir del desdoblamiento de un participante inherente.

En concreto, se contemplan tres papeles semánticos para estos casos: un Poseedor de CTDO, un Lugar de Contacto y un Instrumento del CNTE. Los dos

primeros están vinculados al desdoblamiento del CTDO, mientras que el tercero se asocia al desdoblamiento del CNTE.

IV. Finalmente, la relación posesiva que se establece entre los argumentos implicados en las CPD se recoge mediante dos vías diferentes: a través de la propia nomenclatura utilizada para anotar dichos argumentos (i.e. *Poseedor de CTDO*, *Instrumento de CNTE*) y, sobre todo, a partir de la numeración correlativa de los argumentos que describen el potencial valencial del verbo (A0, A1, A2, ...). De esta forma, las entidades desdobladas mantienen la misma numeración que el argumento sobre el que se desdoblan: A1 en el caso del desdoblamiento del CNTE y A2 en el caso del desdoblamiento del CTDO. Sobre esta numeración no incrementada se especifica cada argumento desdoblado mediante una letra adicional que apunta al papel semántico correspondiente: A1-I en el caso del Instrumento del CTNE, A2-P en el caso del Poseedor del CTDO y A2-L en el caso del Lugar de Contacto.

La Tabla siguiente resume la anotación de papeles semánticos adoptada en ADESSE para cada esquema implicado en la anotación de las CPD:

Tabla 4. Anotación de las CPD con verbos de contacto en ADESSE

Esquema	SUJ A1	ODIR A2	OIND A2-P	CP(en) A2-L	CP(con) A1-I
<i>Alguien golpea a alguien</i>	CNTE	CTDO			
<i>Alguien golpea (su) NPC (de alguien)</i>	CNTE	CTDO			
<i>Alguien le golpea NPC a alguien</i>	CNTE	CTDO	Poseedor de CTDO		
<i>Alguien golpea a alguien en NPC</i>	CNTE	CTDO		Lugar de contacto	
<i>Alguien golpea con NPC</i>	CNTE				Instrum. de CNTE

4. Conclusiones

Enriquecer un corpus con información semántica supone un reto importante a efectos de anotación, ya que conlleva dificultades que no se suelen dar al trabajar en otros niveles de análisis.

El significado lingüístico, muchas veces escurridizo, suele remar en dirección contraria a las necesidades metodológicas que impone el proceso de anotación de un corpus, que se rige por la adopción de unidades discretas y de fácil aplicación. Para solucionarlo, nos vemos obligados a buscar un equilibrio entre la consistencia de los datos anotados y la granularidad del análisis realizado, concediendo mayor importancia a uno u otro factor en función de los objetivos que se persigan. El objetivo principal de ADESSE es el de describir la estructura argumental de los verbos de un corpus. Esta finalidad descriptiva explica por qué el citado proyecto inclina la balanza hacia la anotación detallada de papeles semánticos, anotación que se realizó de manera manual sobre cada esquema sintáctico de cada verbo. Sin embargo, la búsqueda de esta granularidad en la anotación no debe ser incompatible con la consideración de neutralizaciones de oposiciones semánticas en las relaciones que se establecen entre verbos y argumentos verbales. Es decir, debemos contemplar la posibilidad de encontrar similitudes semánticas entre argumentos de verbos diferentes, tanto si estos últimos pertenecen a un mismo dominio semántico como si pertenecen a dominios completamente distintos.

Para abordar estas cuestiones, en ADESSE partimos de manera inductiva desde el significado léxico del verbo para reconocer diferentes niveles de generalización en la anotación de papeles semánticos, lo que nos lleva a establecer una distinción fundamental entre roles específicos del verbo, roles específicos de la clase y roles semánticos generalizados. Por otro lado, la granularidad en el análisis semántico del corpus que proponemos en ADESSE nos obliga a encarar algunos problemas de anotación que serían secundarios o irrelevantes si se partiese de una metodología diferente o si se persiguiese un objetivo menos descriptivo. Un buen ejemplo de esos problemas de anotación a los que debemos hacer frente lo constituyen las que hemos denominado construcciones de participante desdoblado (CPD), que se producen típicamente con verbos de contacto y con argumentos que designan nombres de partes

del cuerpo (NPC). La estrategia adoptada en ADESSE para estas CPD se basa en mantener una anotación única para el esquema transitivo básico (i.e. A1:Contactante - A2:Contactado), con independencia de la animación que presente cada argumento implicado, y en usar etiquetas específicas solo para los argumentos correspondientes a los participantes desdoblados.

La relación semántica adicional de carácter posesivo que presentan estos participantes se recoge mediante la numeración correlativa de los argumentos que describen el potencial valencial del verbo: $\Lambda 1$, $\Lambda 1$ -I, $\Lambda 2$, $\Lambda 2$ -P, $\Lambda 2$ -L.

BIBLIOGRAFÍA CITADA

- CASTELLÓN, Irene, Ana Fernández-Montraveta, Gloria Vázquez, Laura Alonso Alemany & Joan Antoni Capilla, 2006. «The SenSem Corpus: A Corpus Annotated at the Syntactic and Semantic Level», *V International Conference on LREC*, 355-359
- DOWTY, David, 1991. «Thematic Proto-roles and Argument Selection», *Language*, 67/3, 547-619
- GARCÍA-MIGUEL, José M^a & Francisco Albertuz, 2005. «Verbs, Semantic Classes and Semantic Roles in the ADESSE project», en Erk, Katrin, Alissa Melinger & Sabine Schulte im Walde (eds): *Proceedings of the Interdisciplinary Workshop on the Identification and Representation of Verb Features and Verb Classes*, Saarbrücken, 28 febrero - 1 marzo 2005.
- GARCÍA-MIGUEL, José M^a, 2007. «Potencial valencial y tipología de argumentos», en I. Castellón y Ana Fernández (eds.), *Perspectivas de análisis de la unidad verbal*, Barcelona, Universitat de Barcelona, 21-33
- HALLIDAY, M.A.K., 1985. *An Introduction to Functional Grammar*, London, Edward Arnold
- LANGACKER, Ronald W., 1991. *Foundations of Cognitive Grammar. Vol. II: Descriptive Application*, Standford, Standford University Press.
- LEVIN, Beth & Malka Rappaport-Hovav, 2005. *Argument Realization*, Cambridge, Cambridge University Press
- MARTI, María Antònia, Mariona Taulé, Manu Bertran & Lluís Márquez, 2007. *AnCora: Multilingual and multilevel Annotated Corpora*. Dispon. en <http://clic.ub.edu/ancora/>
- PALMER, Martha, Daniel Gildea & Paul Kingsbury, 2005. «The Proposition Bank: An Annotated Corpus of Semantic Roles», *Computational Linguistics*, 31/1.
- RUPPENHOFER, Josef, Michael Ellsworth, Miriam L. R. Petruck, Christopher L. Johnson & Jan Scheffczyk, 2006. *FrameNet II: Extended Theory and Practice*. Dispon. en <http://framenet.icsi.berkeley.edu/book/book.html>

Valladolid
21 al 23 de Junio de 2010 **IX Congreso Internacional
de Lingüística General**

- SUBIRATS, Carlos, 2004. «FrameNet español. Una red semántica de marcos conceptuales», *VI International Congress of Hispanic Linguistics*, Leipzig University (Germany), octubre 2003
- VAN VALIN, Robert D. & Randy J. LaPolla, 1997. *Syntax. Structure, meaning and function*, Cambridge, Cambridge University Press.
- VATER, H., 2003. «Valency and diathesis», en H. Cuyckens, T. Berg, R. Dirven y K. Panther (eds.), *Motivation in Language. Studies in honor of Günter Radden*, Amsterdam, John Benjamins.