

Manual para la Anotación Discursiva de Corpus en Español

Laura Alonso i Alemany, Daniel Alonso i Alemany,
Ezequiel Andújar Hinojosa y Robert Sola Salvetierra

GRIAL Report
Grial-2004-2



GRIAL REPORT SERIES
Grup de Recerca Interuniversitari en Aplicacions Lingüístiques

GRIAL REPORT SERIES
Grup de Recerca Interuniversitari en Aplicacions Lingüístiques

Universitat Autònoma de Barcelona (UAB)
Universitat de Barcelona (UB)
Universitat de Lleida (UdL)

*Departament de Lingüística General Facultat de Filologia
Universitat de Barcelona Gran Via de les Corts Catalanes, 585
Barcelona*

Índice General

1	Objetivo	4
2	Etiquetación XML	4
3	Unidades Discursivas	4
4	Cómo determinar el punto de adjunción	5
5	Relaciones entre Unidades Discursivas	8

1 Objetivo

El punto de partida de este trabajo es un corpus en español en el que han sido identificadas las unidades discursivas básicas. El objetivo de la anotación de corpus es **determinar, para cada unidad discursiva, la relación con su contexto a nivel de sintaxis y semántica discursivas.**

2 Etiquetación XML

En el corpus, cada unidad discursiva está anotada en XML, de la siguiente forma:

```
<ud atributo="valor" atributo="valor">
  palabra-1
  palabra-2
  ...
  palabra-n
</ud>
```

El contenido de una unidad discursiva es el texto que se encuentra entre una etiqueta de inicio `<ud>` y una etiqueta de final `</ud>`. La etiqueta de inicio contiene una serie de pares atributo-valor que se aplican a la unidad discursiva y que pretenden describir la unidad en sí y sus relaciones con el contexto a nivel discursivo.

Los atributos que se encuentran en la etiqueta de inicio son los siguientes:

- **identificador.** Asigna un número único a cada unidad discursiva, por ejemplo, `<1.4.2>`.
- **tipo.** Describe el tipo de unidad discursiva de que se trata. Hemos listado los diferentes tipos en la sección 3. Por ejemplo, `<comp>`.
- **punto de adjunción.** Señala el identificador de la unidad discursiva con la que se relaciona la unidad discursiva actual, por ejemplo, `<adj="1.4.1">`. En la sección 4 se explica con detalle cómo determinar con qué unidad se relaciona cada unidad discursiva.
- **semántica de relaciones entre segmentos.** Se expresan como una lista de atributos a los cuales se asigna valor 1 si se aplican a la unidad discursiva actual, o se dejan vacíos si no se aplican, por ejemplo, `<cause="1" context="">`. Se describen en la sección 5.

De estos atributos, los dos primeros ya han sido asignados a cada unidad discursiva. La anotación de relaciones consistirá en asignar valores para el punto de adjunción y la semántica de las relaciones entre segmentos.

3 Unidades Discursivas

En el texto se han identificado las unidades discursivas básicas. Los tipos de unidad discursiva mínima que se han determinado son:

- **párrafo** `<par>`: fragmento de texto entre dos puntos y a parte.
- **oración** `<or>`: fragmento de texto entre dos puntos (incluyendo puntos de admiración, de interrogación, dos puntos y punto y coma).
- **título** `<tit>`: fragmento de texto que titula un texto o una sección.
- **complemento** `<comp>`: fragmento de texto con entidad discursiva propia pero gramaticalmente requerido por una unidad lingüística que lo contiene; por ejemplo, las frases subordinadas completivas (ver ejemplo 1).

- **segmento interior** <en>: fragmento de texto con entidad discursiva propia, no requerido para la buena formación gramatical o discursiva del texto (ver ejemplo 1).
- **marcador del discurso** <md>: fragmento de texto sin contenido proposicional, su semántica consiste en señalar relaciones entre unidades discursivas.

Cada una de estas unidades puede contener otras unidades en su interior, lo cual queda reflejado en la numeración del identificador de cada unidad, de la siguiente forma:

```

<1>
  palabra
  palabra
  palabra
  <1.1>
    palabra
    palabra
  </ud 1.1>
  palabra
  <1.2>
    palabra
    palabra
    palabra
    <1.2.1>
      palabra
    </ud 1.2.1>
    palabra
    palabra
  </ud 1.2>
  palabra
  palabra
</ud 1>

```

A medida que se establecen relaciones entre unidades discursivas, se crean estructuras que pueden tener la misma función que una unidad discursiva mínima, ya que otras unidades discursivas pueden establecer relaciones con estas estructuras de forma global, y no con uno de los elementos que las forman. Trataremos esta cuestión en la sección siguiente.

4 Cómo determinar el punto de adjunción

En un texto coherente, toda unidad discursiva se relaciona de alguna forma con su contexto, excepto la primera unidad discursiva con contenido de un texto, que se relaciona únicamente con ella misma.

Se podría considerar que la primera unidad de un nivel estructural inferior se relaciona con la unidad del nivel estructural inmediatamente superior, que la contiene, por ejemplo, la primera oración se relacionaría con el primer párrafo. Sin embargo, es deseable que las relaciones de tipo discursivo se establezcan entre unidades con contenido, y no puramente estructurales. Los párrafos se consideran como unidades estructurales antes de que todo su contenido haya sido analizado, por lo tanto, tan sólo son elegibles como puntos de adjunción cuando se ha completado su análisis.

El punto de adjunción únicamente identifica la unidad discursiva con la que la unidad discursiva actual establece una relación, independientemente del tipo de relación.

Las unidades con las que una unidad discursiva puede establecer relación son:

- unidades discursivas tipográficas: párrafos y oraciones. El valor para el atributo de punto de adjunción será el identificador de la unidad discursiva, por ejemplo, en el ejemplo 2, la relación entre la unidad discursiva "Sin embargo" y el párrafo anterior se expresa de la siguiente forma en la etiqueta de "Sin embargo": <md 2.1.1 adj="1">

No fueron pocos en la plaza quienes comentaron, con cierta hilaridad, que esta era la primera construcción "folrada" del siglo, aunque hubo quien replicó que el siglo no empieza hasta el próximo 1 de enero.

```
<or 1>
  No
  fueron
  pocos
  en
  la
  plaza
  quienes
  comentaron
  <en 1.1>
    , con
    cierta
    hilaridad,
  </ud en 1.1>
  <comp 1.2>
    que
    esta
    era
    la
    primera
    construcción
    "folrada"
    del
    siglo
  </ud comp 1.2>
  <en 1.3>
    , aunque
    hubo
    quien
    replicó
    <comp 1.3.1>
      que
      el
      siglo
      no
      empieza
      hasta
      el
      próximo
      1
      de
      enero.
    </ud comp 1.3.1>
  </ud en 1.3>
</ud 1>
```

Figura 1: Tipos de unidades discursivas y su notación en XML.

tipo de adjunción	definición	muestra
simple	una UD se adjunta a otra UD	<adj="id">
de conjunto	una UD se adjunta a varias UD adyacentes	<adj="id-id">
fragmentaria	una UD se adjunta a un fragmento no UD	<adj="id-cosa">
binario	una UD marcador se adjunta a dos o más UDs	<adj="id,id">
múltiple	una UD tiene más de una relación distinta	<adj="id adj="id">
combinado	cualquier combinación de los anteriores	<adj="id-id,id">

Tabla 1: Tipos de relación de adjunción de una unidad discursiva con su contexto.

- unidades discursivas básicas no tipográficas: complementos, segmentos interiores de oración, marcadores del discurso. El valor para el atributo de punto de adjunción será igualmente el identificador de la unidad discursiva.
- fragmentos de texto sin estatus discursivo: una unidad discursiva puede establecer una relación con un fragmento de texto que no se haya constituido como unidad discursiva. Por ejemplo, una frase de relativo suele establecer una relación con un sintagma nominal, y no con una unidad discursiva. Si el punto de adjunción de una unidad discursiva no se puede expresar mediante referencia a identificadores de otras unidades discursivas, se expresará de forma simbólica, uniendo el o los identificadores de las unidades discursivas donde se encuentra el fragmento de texto y el núcleo (sintáctico, semántico) el fragmento de texto en sí, de la siguiente manera: <en 1.3.2 adj="1.3-afluencia">, tal como se puede ver en los ejemplos 2 y 3.
- entidades contextuales sin presencia textual: en casos excepcionales, una unidad discursiva puede establecer una relación con un elemento del contexto que no se corresponda con un fragmento concreto de texto. En este caso, el valor del atributo *adj* será *ent*, y se deberá comentar el caso en grupo <en 5.4 adj="ent">.
- dos unidades: una unidad discursiva de tipo *marcador del discurso* puede establecer una relación con dos unidades discursivas. En este caso, el valor para el atributo de punto de adjunción serán dos números identificadores de unidad discursiva unidos mediante una coma, como se puede ver en los ejemplos 2 y 3. Si la relación que los une es simétrica (ver sección 5), el orden de los identificadores se corresponderá con el orden de las unidades discursivas. Si la relación que los une es asimétrica, el primer identificador corresponderá a la unidad discursiva que se encuentra en el nivel superior. Sin embargo, no es necesario que un marcador del discurso se relacione con dos unidades discursivas, es el caso, por ejemplo, de “Por suerte” en el ejemplo 3.
- múltiples unidades adyacentes entre ellas: una unidad discursiva puede establecer una relación con un fragmento continuo de texto que no se corresponda con una unidad discursiva mínima. En este caso, el valor para el atributo de punto de adjunción serán dos números identificadores de unidad discursiva, el de la primera unidad del fragmento y el de la segunda, unidos mediante un guión, de la siguiente manera: <md 2.1.1 adj="1.1-1.3">. En el caso de “Sin embargo” en el ejemplo 3 encontramos una adjunción a dos unidades, siendo la segunda una unidad múltiple.

Una unidad discursiva puede establecer más de una relación con su contexto. En este caso, se crearán tantos pares atributo-valor *adj=id* como relaciones se establezcan, de la siguiente forma: <en 3.2 adj="3.1 adj="3.3.1 adj="2">. En la Tabla 4 se puede ver un resumen de los tipos de relación de adjunción de una unidad con su contexto.

Para determinar cuáles de las unidades discursivas disponibles en el contexto es el punto de adjunción de una unidad discursiva, el conocimiento más fiable es la propia intuición del hablante. Sin embargo, suelen darse algunas regularidades a este respecto:

- las unidades discursivas de tipo *comp* suelen adjuntarse a la oración principal dentro de la cual se encuentran.
- las unidades discursivas con forma sintáctica de frase de relativo suelen adjuntarse a un elemento nominal de la oración principal dentro de la cual se encuentran.
- la unidad discursiva más importante de un texto, o la primera, se adjunta a sí misma.

5 Relaciones entre Unidades Discursivas

Las relaciones que una unidad discursiva puede establecer con su contexto pueden ser puramente estructurales o bien hacer una aportación semántica a la estructura del discurso. Para describir el posible matiz semántico de las relaciones, existen siete atributos, a los cuales se asigna un valor binario: "1" si la relación tiene el atributo en cuestión, o vacío si no lo tiene. El ejemplo 10 ilustra el funcionamiento de la notación, y las figuras 7 y 11 muestran una representación gráfica de la notación XML.

Los atributos descriptivos de la semántica de las relaciones entre unidades discursivas son:

- **contexto.** una unidad discursiva sitúa la unidad a la que se adjunta en un contexto determinado, por ejemplo, expresando el tiempo o lugar en que ocurre la unidad a la que se adjunta, véase ejemplo 5.
- **causa.** una unidad discursiva establece una relación de causa con la unidad a la que se adjunta, véase ejemplo 6.
- **paralelismo.** una unidad discursiva establece una relación de equivalencia con la unidad a que se adjunta, del tipo *A es (modalidad) B*, por ejemplo: *A es un tipo de B*, *A es un ejemplo de B*, *A es como B*, *A es lo contrario de B*, véase ejemplo 8.
- **revisión.** una unidad discursiva revisa la información contenida en la unidad a que se adjunta, por ejemplo, negando una implicación que se desprenda de ella, véase ejemplo 9.
- **progresión.** una unidad discursiva introduce un progreso en el desarrollo de la línea argumentativa del texto. Dado que la función primera de un texto es desarrollar una línea argumentativa de forma lineal, consideramos que todo fragmento de texto contribuye a su progresión argumentativa a no ser que se explicita lo contrario. Especialmente, las unidades discursivas que aportan información necesaria para el desarrollo del discurso establecen relaciones de progresión, como es el caso de las unidades de tipo *comp*.
- **elaboración.** una unidad discursiva introduce información redundante o sin continuidad en la línea argumentativa del texto. Por ejemplo, las frases de relativo suelen elaborar la información ya aportada por la unidad a la que se adjuntan.
- **simetría.** una unidad discursiva se encuentra en el mismo nivel estructural que la unidad a la que se adjunta. Se puede determinar si una unidad se encuentra en el mismo nivel estructural que otra si no es posible intercambiar su orden, ni siquiera sustituyendo las expresiones referenciales por sus referentes. Es el caso de las unidades unidas por "pero" en el ejemplo 10, que no pueden intercambiar su orden secuencial sin que el sentido de la oración original cambie también.
- **asimetría.** una unidad discursiva se encuentra en un nivel estructural distinto que la unidad a la que se adjunta. Estas relaciones se marcan siempre en la unidad que se sitúa en un nivel inferior. Por ejemplo, las unidades discursivas de tipo *comp* sólo tienen una relación de

Josep Ignasi Urenda, que fue comisionado por el alcalde Joan Clos para esta cuestión, dispuso un plazo orientativo para la presentación de posibles proyectos que acabó el pasado viernes, 31 de marzo.

Sin embargo, el proyecto presentado por Shuaibi no ofrece, según fuentes municipales, las precisiones requeridas para empezar a concretar la posible ubicación de un edificio que debe albergar, además de una mezquita, distintas dependencias (biblioteca, sala de conferencias, escuela coránica) de un centro cultural islámico.

```
<par 1 adj="1">
  <or 1.1 adj="1.1">
    Josep Ignasi Urenda
    <en 1.1.1 adj="1.1">
      ,que fue comisionado por el alcalde Joan Clos para esta cuestion,
    </en 1.1.1>
    dispuso un plazo orientativo para la presentacion de posibles proyectos
    <en 1.1.2 adj="1.1-plazo">
      que acabo el pasado viernes,
      <en 1.1.2.1 adj="1.1.2-viernes">
        31 de marzo.
      </en 1.1.2.1>
    </en 1.1.2>
  </or 1.1>
</par 1>
<par 2 adj="1">
  <or 2.1 adj="1">
    <md 2.1.1 adj="2,1">
      Sin embargo,
    </md 2.1.1>
    el proyecto presentado por Shuaibi no ofrece
    <en 2.1.2 adj="2.1">
      , segun fuentes municipales,
    </en 2.1.2>
    las precisiones requeridas
    <comp 2.1.3 adj="2.1">
      para empezar a concretar la posible ubicacion de un edificio
      <en 2.1.3.1 adj="2.1.3-edificio">
        que debe albergar
        <en 2.1.3.1.1 adj="2.1.3.1">
          <md 2.1.3.1.1.1 adj="2.1.3.1.1">
            , ademas de
          </md 2.1.3.1.1.1>
          una mezquita,
        </en 2.1.3.1.1>
        distintas dependencias
        <en 2.1.3.1.2 adj="2.1.3.1-dependencias">
          (biblioteca, sala de conferencias, escuela coranica)
        </en 2.1.3.1.2>
        de un centro cultural islamico.
      </en 2.1.3.1>
    </comp 2.1.3>
  </or 2.1>
</par 2>
```

Figura 2: Puntos de adjunción de unidades discursivas.

"Los inmigrantes no son conflictivos. Por suerte todavía no ha habido ningún problema con los vecinos", explica Martínez. Sin embargo, la afluencia masiva que no figura en ningún censo hace temer que el barrio se convierta en un gueto.

```

<par 1 adj="1">
  <or 1.1 adj="1.1">
    "Los inmigrantes no son conflictivos."
  </or 1.1>
  <or 1.2 adj="1.2">
    <comp 1.2.1 adj="1.2">
      <md 1.2.1.1 adj="1.2.1">
        Por suerte
      </md 1.2.1.1>
      todavía no ha habido ningún problema
      con los vecinos",
    </comp 1.2.1>
    explica Martínez.
  </or 1.2>
  <or 1.3 adj="1.2">
    <md 1.3.1 adj="1.3,1.1-1.2">
      Sin embargo,
    </md 1.3.1>
    la afluencia masiva
    <en 1.3.2 adj="1.3-afluencia">
      que no figura en ningún censo
    </en 1.3.2>
    hace temer
    <comp 1.3.3 adj="1.3">
      que el barrio se convierta
      en un gueto.
    </comp 1.3.3>
  </or 1.3>
</par 1>

```

Figura 3: Puntos de adjunción de unidades discursivas.

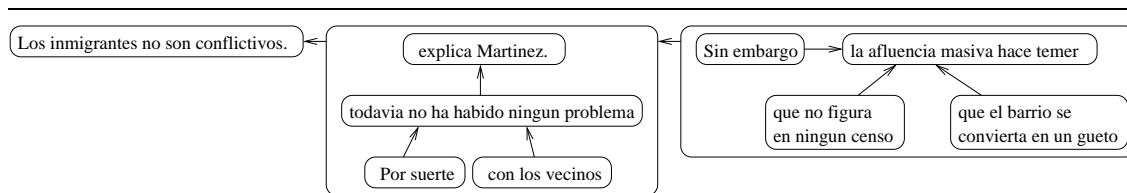


Figura 4: Representación gráfica del Ejemplo 3.

asimetría con la unidad a la que se adjuntan, sin ningún otro tipo de contenido. Una prueba fiable para determinar si una unidad tiene una relación de asimetría con otra es comprobar que la unidad que se sitúa en el nivel inferior puede cambiar su posición secuencial dentro del texto sin que éste cambie su sentido. Por ejemplo, en el ejemplo 6, las unidades introducidas por “para” y “porque” establecen una relación de asimetría con “hay que leer”, y por lo tanto las siguientes reordenaciones no cambian el sentido de la frase original:

para escribir hay que leer "porque, de Shakespeare a Dickens, la literatura es un gran plagio cuya belleza reside en que no se note".

hay que leer para escribir "porque, de Shakespeare a Dickens, la literatura es un gran plagio cuya belleza reside en que no se note".

"porque, de Shakespeare a Dickens, la literatura es un gran plagio cuya belleza reside en que no se note", para escribir hay que leer.

"porque, de Shakespeare a Dickens, la literatura es un gran plagio cuya belleza reside en que no se note", hay que leer para escribir.

En el caso de múltiples relaciones de una unidad discursiva con su contexto, se repetirán los pares atributo-valor que describen la semántica de las relaciones después del par atributo-valor que describe el punto de adjunción de la relación, tantas veces como sea necesario para describir todas las relaciones que establezca un segmento, de la siguiente manera:

```
<en 3.2 adj="3.1"
      cause="1"      context=""  parallel=""
      revision=""   elabora=""  progre="1"
      sym=' 1'      asym=""
adj=3.3.1
      cause="1"      context="1"  parallel=""
      revision=""   elabora="1"  progre=""
      sym=""        asym="1"
adj=2
      cause=""      context=""  parallel="1"
      revision=""   elabora=""  progre=""
      sym="1"      asym=""    >
```

Las relaciones retóricas más habituales, como *finalidad* o *consecuencia* se pueden describir mediante ciertas combinaciones de estas características, como se muestra en la Tabla 5. Por motivos de espacio, hemos agrupado las características anteriores en tres categorías: *de materia* (contexto, causa, paralalismo, revisión), *argumentativas* (progresión y elaboración) y *estructurales* (simetría y asimetría).

relación convencional	características			marcador de ejemplo
	de materia	argumentativa	estructural	
consecuencia	causa	progresión	simetría	<i>por eso</i>
finalidad	causa	progresión	asimetría	<i>para</i>
-	causa	elaboración	simetría	
razón	causa	elaboración	asimetría	<i>ya que</i>
lista	paralelismo	progresión	simetría	<i>primero</i>
nuevo argumento	paralelismo	-	asimetría	<i>además</i>
adición	paralelismo	elaboración	simetría	<i>además</i>
concreción	paralelismo	elaboración	asimetría	<i>por ejemplo</i>
contraste	revisión	progresión	simetría	<i>pero</i>
contraargument	revisión	progresión	asimetría	<i>no obstante</i>
concesión	revisión	elaboración	asimetría	<i>aunque</i>
-	revisión	elaboración	simetría	
secuencia	contexto	progresión	simetría	<i>entonces</i>
conclusión	contexto	progresión	asimetría	<i>en conclusión</i>
condición	contexto	elaboración	simetría	<i>si</i>
circunstancia	contexto	elaboración	asimetría	<i>cuando</i>

Tabla 2: Correspondencia entre características descriptivas y relaciones retóricas convencionales.

Cuando se apliquen las modificaciones acordadas en el plan será posible acceder a Via Augusta en dirección a la ronda de Dalt desde la calle Vergòs.

```

<or 1.3>
  <en 1.3.1 adj="1.3"
    cause="" " context="1" parallel="" "
    revision="" " elabora="1" progre="" "
    sym="" " asym="1" >
    <md 1.3.1.1 adj="1.3,1.3.1"
      cause="" " context="1" parallel="" "
      revision="" " elabora="1" progre="" "
      sym="" " asym="1" >
      Cuando
    </ud>
    se apliquen las modificaciones
    <en 1.3.1.2 adj="1.3.1-modificaciones"
      cause="" " context="" " parallel="" "
      revision="" " elabora="1" progre="" "
      sym="" " asym="1" >
      acordadas en el plan
    </en 1.3.1.2>
  </ud>
  sera posible acceder a Via Augusta
  en direccion a la ronda de Dalt
  desde la calle Vergos.
</en 1.3.1>
</or 1.3>

```

Figura 5: Relación de contexto.

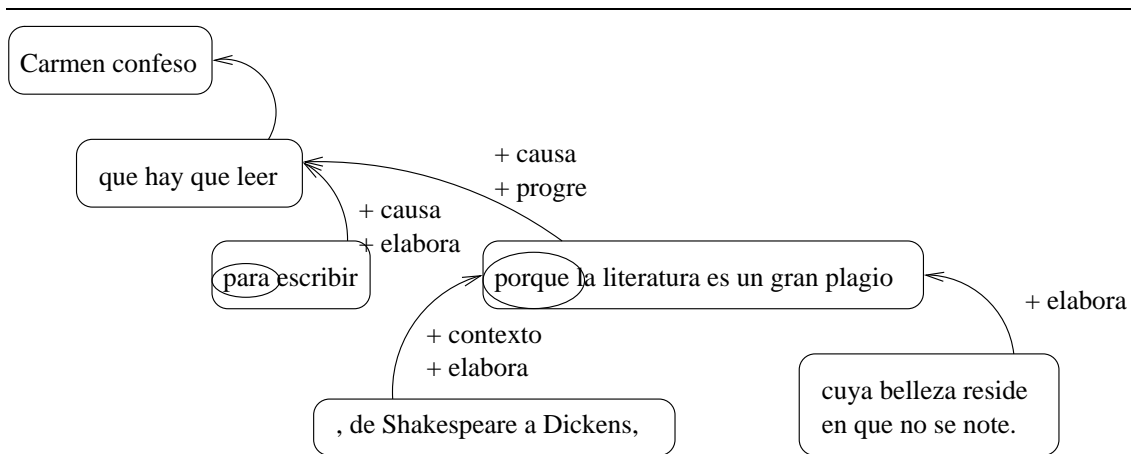


Figura 7: Representación gráfica del Ejemplo 6.

El nuevo plan propone algunos cambios sustanciales, como la decisión de dar doble sentido de circulación a la calle Sant Joan Bosco y la calle Vergòs.

```

<par 1 adj="1"
    cause=" " context=" " parallel=" "
    revision=" " elabora=" " progre="1"
    sym="1" asym=" " >
  <or 1.1 adj="1.1"
    cause=" " context=" " parallel=" "
    revision=" " elabora=" " progre="1"
    sym=" " asym=" " >
    El nuevo plan propone algunos cambios sustanciales
    <en 1.1.1 adj="1.1-cambios"
      cause=" " context=" " parallel="1"
      revision=" " elabora="1" progre=" "
      sym=" " asym="1" >
      <md 1.1.1.1 adj="1.1.1"
        cause=" " context=" " parallel=" "
        revision=" " elabora=" " progre=" "
        sym=" " asym=" " >
        , como
        </ud>
        la decision de
        <comp 1.1.1.1.1 adj="1.1.1-decision"
          cause=" " context=" " parallel=" "
          revision=" " elabora=" " progre="1"
          sym=" " asym="1" >
          dar doble sentido de circulacion a la
          calle Sant Joan Bosco y la calle Vergos.
          </ud>
        </ud>
      </ud>
    </or>
  </par>

```

Figura 8: Relación de paralelismo.

En adelante, el acceso a la zona restringida seguirá regulado por un pilón móvil, pero éste bajará con la simple proximidad de los automóviles.

```
<or 2.2">
  <en 2.2.1 adj="2.2"
    cause=" " context="1" parallel=" "
    revision=" " elabora="1" progre=" "
    sym=" " asym="1" >
    En adelante,
  </ud>
  el acceso a la zona restringida
  seguira regulado por un pilon movil
  <en 2.2.2 adj="2.2"
    cause=" " context=" " parallel=" "
    revision="1" elabora=" " progre="1"
    sym="1" asym=" " >
    <md 2.2.2.1 adj="2.2,2.2.2"
      cause=" " context=" " parallel=" "
      revision="1" elabora=" " progre="1"
      sym="1" asym=" " >
      , pero
    </ud>
    este bajara con la simple proximidad de los automoviles.
  </ud>
</or>
```

Figura 9: Relación de revisión.

El nuevo plan propone algunos cambios sustanciales, como la decisión de dar doble sentido de circulación a la calle Sant Joan Bosco y la calle Vergòs. De este modo, el distrito trata de hallar una vía de descongestión del tráfico en la zona sur de Tres Torres y, en lo fundamental, en la calle Doctor Roux. Cuando se apliquen las modificaciones acordadas en el plan será posible acceder a Via Augusta en dirección a la ronda de Dalt desde la calle Vergòs.

Otra decisión de cierta trascendencia incluida en el plan será la práctica eliminación del sistema de acceso restringido en Major de Sarrià, donde hasta ahora era precisa la utilización de una tarjeta que se distribuía exclusivamente al vecindario. En adelante, el acceso a la zona restringida seguirá regulado por un pilón móvil, pero éste bajará con la simple proximidad de los automóviles. En este mismo capítulo, se reforzará la capacidad de la calle Bonaplata para llegar al otro extremo del centro histórico. Con este objetivo se abrirá la calle Carme Karr, que ahora quedaba cortada antes del paseo de la Bonanova. Por último, el plan también incluye dos propuestas para modificar el recorrido de las líneas 34 y la línea 66 de los autobuses de TMB.

<par 1 adj="1"

cause=" " context=" " parallel=" "
 revision=" " elabora=" " progre="1"
 sym="1" asym=" " >

<or 1.1 adj="1.2"

cause=" " context="1" parallel=" "
 revision=" " elabora="1" progre=" "
 sym=" " asym="1" >

El nuevo plan propone algunos cambios sustanciales

<en 1.1.1 adj="1.1-cambios"

cause=" " context=" " parallel="1"
 revision=" " elabora="1" progre=" "
 sym=" " asym="1" >

<md 1.1.1.1 adj="1.1-cambios,1.1.1"

cause=" " context=" " parallel="1"
 revision=" " elabora="1" progre=" "
 sym=" " asym="1" >

, como

</ud>

la decision de

<comp 1.1.1.1.1 adj="1.1.1-decision"

cause=" " context=" " parallel=" "
 revision=" " elabora=" " progre="1"
 sym=" " asym="1" >

dar doble sentido de circulacion a la
 calle Sant Joan Bosco y la calle Vergos.

</ud>

</ud>

</ud>


```

<or 1.2 adj="1.2"
      cause=" " context=" " parallel=" "
      revision=" " elabora=" " progre="1"
      sym=" " asym=" " >
<md 1.2.1 adj="1.2,1.1"
      cause=" " context="1" parallel=" "
      revision=" " elabora=" " progre=" "
      sym=" " asym="1" >
  De este modo,
</ud>
el distrito trata de hallar una via de descongestion
del trafico
<en 1.2.2 adj="1.2.3"
      cause=" " context=" " parallel="1"
      revision=" " elabora="1" progre=" "
      sym=" " asym="1" >
en la zona sur de Tres Torres
<en 1.2.3 adj="1.2"
      cause=" " context="1" parallel=" "
      revision=" " elabora=" " progre="1"
      sym=" " asym="1" >
<md 1.2.3.1 adj="1.2.2,1.2.3"
      cause=" " context=" " parallel="1"
      revision=" " elabora=" " progre=" "
      sym="1" asym=" " >
  y
</ud>
<md 1.2.3.2 adj="1.2.3,1.2.2"
      cause=" " context=" " parallel="1"
      revision=" " elabora="1" progre=" "
      sym=" " asym="1" >
  en lo fundamental,
</ud>
en la calle Doctor Roux.
</ud>
</ud>

```

```

<or 1.3 adj="1.2"
      cause=" " context=" " parallel=" "
      revision=" " elabora="1" progre=" "
      sym="1" asym="1" >
<en 1.3.1 adj="1.3"
      cause=" " context="1" parallel=" "
      revision=" " elabora="1" progre=" "
      sym=" " asym="1" >
<md 1.3.1.1 adj="1.3,1.3.1"
      cause=" " context="1" parallel=" "
      revision=" " elabora="1" progre=" "
      sym=" " asym="1" >
      Quando
</ud>
      se apliquen las modificaciones
<en 1.3.1.2 adj="1.3.1-modificaciones"
      cause=" " context="1" parallel=" "
      revision=" " elabora="1" progre=" "
      sym=" " asym="1" >
      acordadas en el plan
</ud>
</ud>
      sera posible acceder a Via Augusta
      en direccion a la ronda de Dalt
      desde la calle Vergos.
</ud>
</ud>

```

<par 2 adj="1.3"

```
cause=" " context=" " parallel=" "
revision=" " elabora=" " progre="1"
sym="1" asym=" "
adj="1.2"
cause=" " context=" " parallel=" "
revision=" " elabora="1" progre=" "
sym="1" asym=" " >
```

<or 2.1 adj="1.3"

```
cause=" " context=" " parallel="1"
revision=" " elabora=" " progre="1"
sym="1" asym=" "
adj="1.2"
cause=" " context=" " parallel="1"
revision=" " elabora="1" progre=" "
sym=" " asym="1" >
```

Otra decision de cierta trascendencia

<en 2.1.1 adj="2.1-decision"

```
cause=" " context="1" parallel=" "
revision=" " elabora="1" progre=" "
sym=" " asym="1" >
```

incluida en el plan

</ud>

sera la practica eliminacion del sistema
de acceso restringido en Major de Sarria

<en 2.1.2 adj="2.1-en"

```
cause=" " context="1" parallel=" "
revision=" " elabora="1" progre=" "
sym=" " asym="1" >
```

<md 2.1.2.1 adj="2.1,2.1.2"

```
cause=" " context="1" parallel=" "
revision=" " elabora="1" progre=" "
sym=" " asym="1" >
```

donde

</ud>

<en 2.1.2.2 adj="2.1.2"

```
cause=" " context="1" parallel=" "
revision=" " elabora="1" progre=" "
sym=" " asym="1" >
```

hasta ahora

</ud>

era precisa la utilizacion de una tarjeta

<en 2.1.2.3 adj="2.1.2-tarjeta"

```
cause=" " context=" " parallel=" "
revision=" " elabora="1" progre=" "
sym=" " asym="1" >
```

que se distribuia exclusivamente al vecindario.

</ud>

</ud>

</ud>

```

<or 2.2 adj="2.1"
      cause=" " context=" " parallel=" "
      revision=" " elabora="1" progre=" "
      sym="1" asym=" "
<en 2.2.1 adj="2.2"
      cause=" " context="1" parallel=" "
      revision=" " elabora="1" progre=" "
      sym=" " asym="1" >
      En adelante,
</ud>
el acceso a la zona restringida
seguira regulado por un pilon movil
<en 2.2.2 adj="2.2"
      cause=" " context=" " parallel=" "
      revision="1" elabora=" " progre=" "
      sym="1" asym=" " >
<md 2.2.2.1 adj="2.2,2.2.2"
      cause=" " context=" " parallel=" "
      revision="1" elabora=" " progre="1"
      sym="1" asym=" " >
      , pero
</ud>
este bajara con la simple proximidad de los automoviles.
</ud>
</ud>

<or 2.3 adj="2.1"
      cause=" " context=" " parallel="1"
      revision=" " elabora=" " progre="1"
      sym="1" asym=" "
      adj="1.2"
      cause=" " context=" " parallel="1"
      revision=" " elabora="1" progre=" "
      sym=" " asym="1" >
<md 2.3.1 adj="2.1,2.3"
      cause=" " context=" " parallel="1"
      revision=" " elabora=" " progre="1"
      sym="1" asym=" " >
      En este mismo capitulo,
</ud>
se reforzara la capacidad de la calle Bonaplata
<en 2.3.2 adj="2.3"
      cause="1" context=" " parallel=" "
      revision=" " elabora=" " progre="1"
      sym=" " asym="1" >
<md 2.3.2.1 adj="2.3,2.3.2"
      cause="1" context=" " parallel=" "
      revision=" " elabora=" " progre="1"
      sym=" " asym="1" >
      para
</ud>
llegar al otro extremo del centro historico.
</ud>
</ud>

```

```

<or 2.4 adj="2.3"
      cause=" " context=" " parallel=" "
      revision=" " elabora="1" progre=" "
      sym=" " asym="1"
<md 2.4.1 adj="2.3,2.4"
      cause="1" context=" " parallel=" "
      revision=" " elabora="1" progre=" "
      sym=" " asym="1" >
      Con este objetivo
</ud>
se abrira la calle Carme Karr
<en 2.4.2 adj="2.4-calle"
      cause=" " context=" " parallel=" "
      revision=" " elabora="1" progre=" "
      sym=" " asym="1" >
      , que
<en 2.4.2.1 adj="2.4.2"
      cause=" " context="1" parallel=" "
      revision=" " elabora="1" progre=" "
      sym=" " asym="1" >
      ahora
</ud>
quedaba cortada
<en 2.4.2.2 adj="2.4.2"
      cause=" " context="1" parallel=" "
      revision=" " elabora="1" progre=" "
      sym=" " asym="1" >
      <md 2.4.2.2.1 adj="2.4.2,2.4.2.2"
      cause=" " context="1" parallel=" "
      revision=" " elabora="1" progre=" "
      sym=" " asym="1" >
      antes del
</ud>
paseo de la Bonanova.
</ud>
</ud>
</ud>

```

```

<or 2.5 adj="2.3"
      cause="1"      context=""  parallel="1"
      revision=""   elabora=""  progre="1"
      sym="1"       asym=""
adj="1.2"
      cause=""      context=""  parallel="1"
      revision=""   elabora="1"  progre=""
      sym=""        asym="1"    >
<md 2.5.1 adj="2.3,2.5"
      cause=""      context=""  parallel="1"
      revision=""   elabora=""  progre="1"
      sym="1"       asym=""    >
  Por ultimo,
</ud>
el plan tambien incluye dos propuestas
<en 2.5.2 adj="2.5"
      cause="1"      context=""  parallel=""
      revision=""   elabora=""  progre="1"
      sym=""        asym="1"    >
  <md 2.5.2.1 adj="2.5,2.5.2"
      cause="1"      context=""  parallel=""
      revision=""   elabora=""  progre="1"
      sym=""        asym="1"    >
    para
  </ud>
  modificar el recorrido de las lineas 34
  y la linea 66 de los autobuses de TMB.
</ud>
</ud>
</ud>

```

Figura 10: Descripción completa de los atributos semánticos de las relaciones entre unidades discursivas.

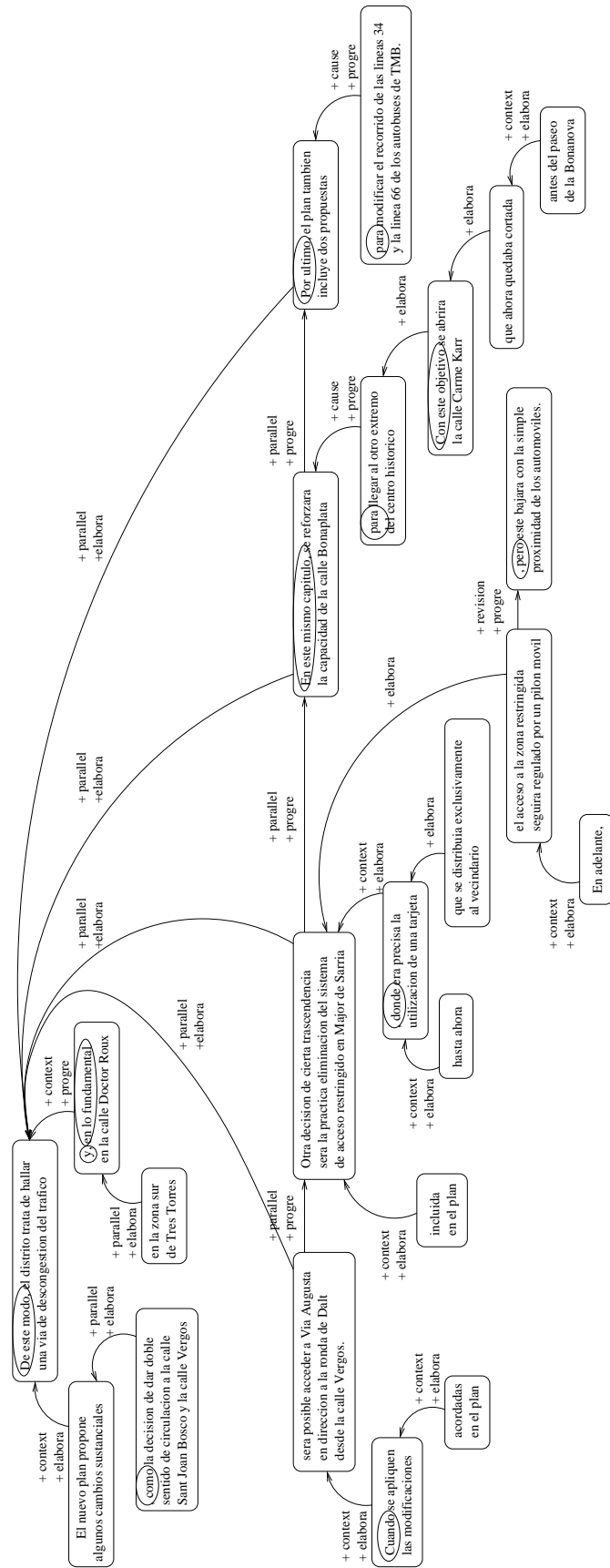


Figura 11: Representación gráfica del Ejemplo 10.

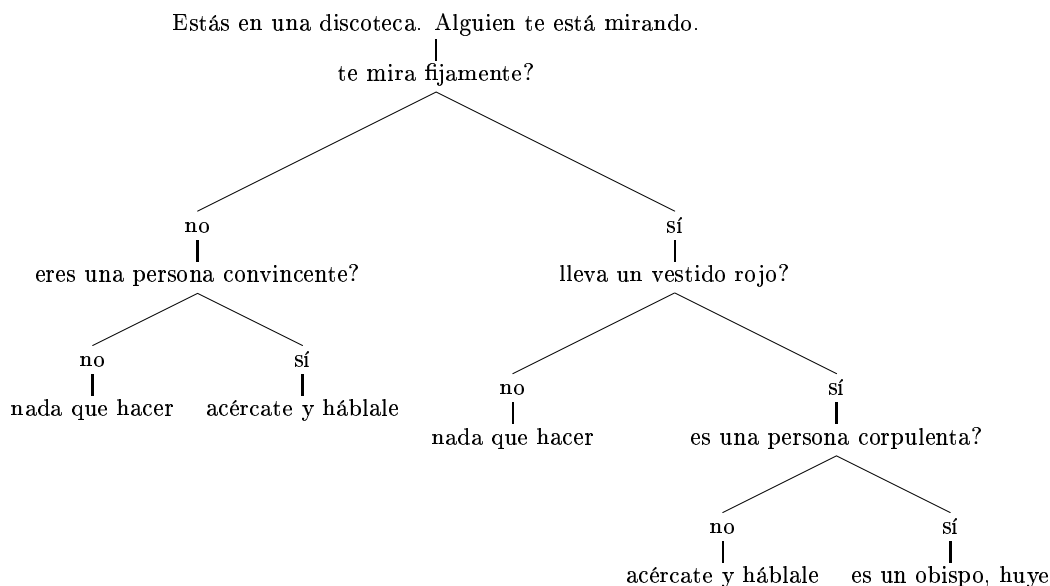


Figura 12: Ejemplo de árbol de decisión

Árboles de decisión para la anotación de textos

Este apéndice está motivado por la necesidad de mayor concreción en los criterios de anotación discursiva de corpus, dados los bajos valores de acuerdo κ entre jueces (nunca más altos de .65) en los trabajos preliminares de anotación.

Se basa en las observaciones realizadas durante el trabajo preliminar de anotación de corpus llevado a cabo en el último trimestre de 2003 por Ezequiel Andújar Hinojosa y Roberto Sola Salvetierra. Daniel Alonso i Alemany contribuyó a la delimitación del nivel de anotación, y al diseño final de la visualización en XML. También ha hecho aportaciones substanciales Henk Zeevat, sobretudo respecto a la característica de *revisión*.

Árboles de decisión

Se aplica una metodología de árboles de decisión que ha sido utilizada con frecuencia en anotación discursiva de corpus, para consolidar la consistencia y reproducibilidad de la anotación. Los árboles de decisión ofrecen un procedimiento a seguir para cada decisión a tomar, como se ve en el ejemplo 12.

En nuestro caso, se trata de árboles que permiten pasar del caso por defecto a casos más específicos. Es decir, en cada decisión que nos encontremos existen varias posibilidades, una de las cuales es el caso por defecto, y lo aplicaremos si no se cumple ninguna de las condiciones para aplicar los demás casos.

Árboles de decisión para características discursivas

Los casos por defecto se aplicarán siempre que el ejemplo no contradiga la definición de la característica.

Simétrico, Asimétrico

El caso por defecto es *simétrico*.

simétrico una unidad discursiva se encuentra en el mismo nivel en una estructura arbórea del texto que la unidad a la que se adjunta.

Millones de personas están consumiendo estas drogas en Estados Unidos y en Europa y no sabemos realmente nada sobre los efectos que producen.

asimétrico una unidad discursiva se encuentra en un nivel estructural distinto que la unidad a la que se adjunta.

Hay que leer porque la literatura es un gran plagio.

Progresión, Elaboración

El caso por defecto es *progresión*.

Atención!!

Para cada segmento que introduce una progresión, hay que añadir una glosa a la anotación, que resuma en pocas palabras qué contenido u objetivo introduce. Los segmentos que introducen elaboración toman su glosa del elemento que ha introducido la progresión que están elaborando.

progresión una unidad discursiva aporta un elemento nuevo a la línea argumentativa del texto.

Para freír un huevo hay que seguir los siguientes pasos: primero, calentar el aceite en una sartén. Luego, romper el huevo encima de la sartén.
glosa → “romper el huevo”

elaboración una unidad discursiva aporta información adicional sobre un elemento ya introducido en la línea argumentativa del texto.

amplía información

Luego, romper el huevo encima de la sartén. Esto puede ser muy peligroso.
glosa → “romper el huevo”

desarrolla un objetivo

Luego, romper el huevo encima de la sartén. Para ello, deberéis golpear el huevo contra alguna superficie dura con contundencia pero delicadamente.
glosa → “romper el huevo”

Revisión, Causalidad, Paralelismo, Contexto

El caso por defecto es *ninguno*.

Atención!!

Para cada segmento que introduce una relación de este tipo, hay que añadir una glosa a la anotación:

revisión la información que se niega y la forma como se niega, distinguiendo cuatro tipos de revisión: negación de expectativas, concesiva, oposición semántica y contraste de focalización.

causalidad la conjunción de causa con que podemos relacionar los segmentos, distinguiendo cuatro tipos:

Conjunciones de causa: porque, para que, por eso, para eso

paralelismo la clase a la que pertenecen los dos elementos relacionados, relevante en el contexto

contexto la expresión circunstancial con que podemos preguntar por el segmento

revisión una unidad discursiva niega una información que se desprende de la unidad a que se adjunta, ya sea porque la información está contenida en ella o porque está implicada por ella.

negación de expectativas (Lagerwerf 1998)

Hoy llueve pero saldré a dar una vuelta.

glosa → “cuando llueve me quedo en casa”

Pepe ha pedido un visado para los EEUU pero no es ningún terrorista.

glosa → “si pides un visado para los EEUU, eres un terrorista”

concesiva (Lagerwerf 1998)

Es muy trabajador, pero le cuesta entender las cosas.

glosa → “queremos hacer el trabajo con él”

oposición semántica (Lagerwerf 1998)

Pepe es simpático, pero Lolo es un imbécil.

glosa → “simpático - imbécil”

No es blanco, sino negro.

glosa → “blanco - negro”

contraste de foco (Umbach pear)

Hemos limpiado toda la casa, pero es que también hemos arreglado el jardín.

sin glosa

causa una unidad discursiva establece una relación de causa con la unidad a la que se adjunta.

Me gustan las manzanas, me voy a comer una.

(*Me gustan las manzanas, y por eso me voy a comer una.*)

glosa → “por eso”

paralelismo una unidad discursiva establece una relación de equivalencia con la unidad a que se adjunta, del tipo *A es (modalidad) B*, por ejemplo: *A es un tipo de B*, *A es un ejemplo de B*, *A es como B*, *A es lo contrario de B*.

Nos compraremos un piso. Pediremos una hipoteca, viviremos con el agua al cuello toda la vida y cuando no nos podamos ya mover, el piso será nuestro.

glosa → “lista”

No me gustan los hombres. Tú, menos.

glosa → “hombres”

contexto una unidad discursiva sitúa la unidad a la que se adjunta en un contexto determinado.

No se dirigieron la palabra durante todo el viaje.

glosa → “cuando”

La relación entre el segmento B y el segmento A es...

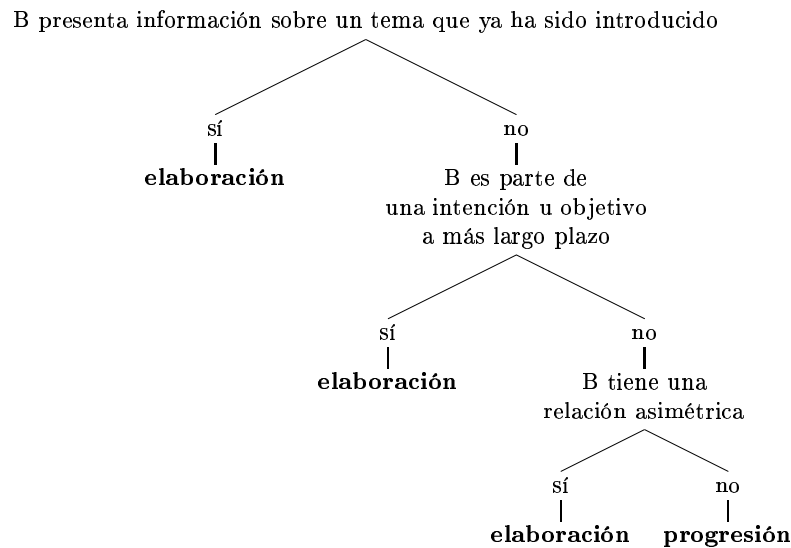


Figura 13: Árbol de decisión para las características progresión y elaboración, donde *progresión* es el caso por defecto.

La relación entre el segmento B y el segmento A es...

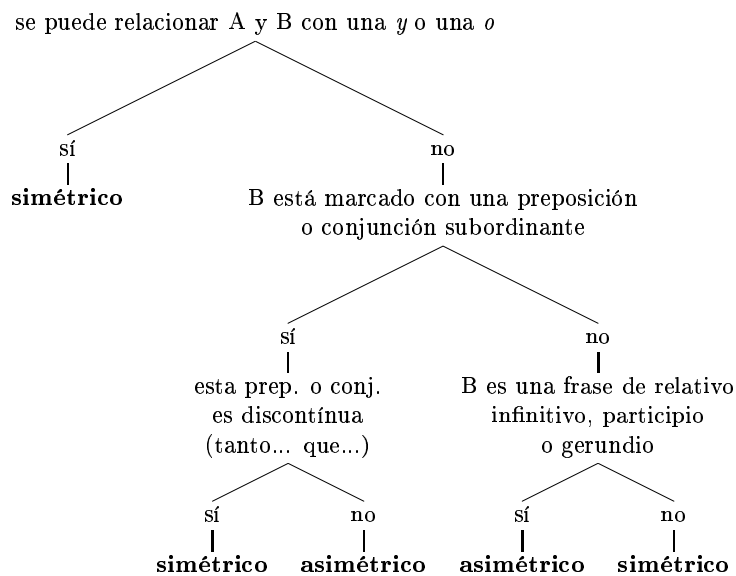


Figura 14: Árbol de decisión para las características simétrico y asimétrico, donde *simétrico* es el caso por defecto.

La relación entre el segmento B y el segmento A es...

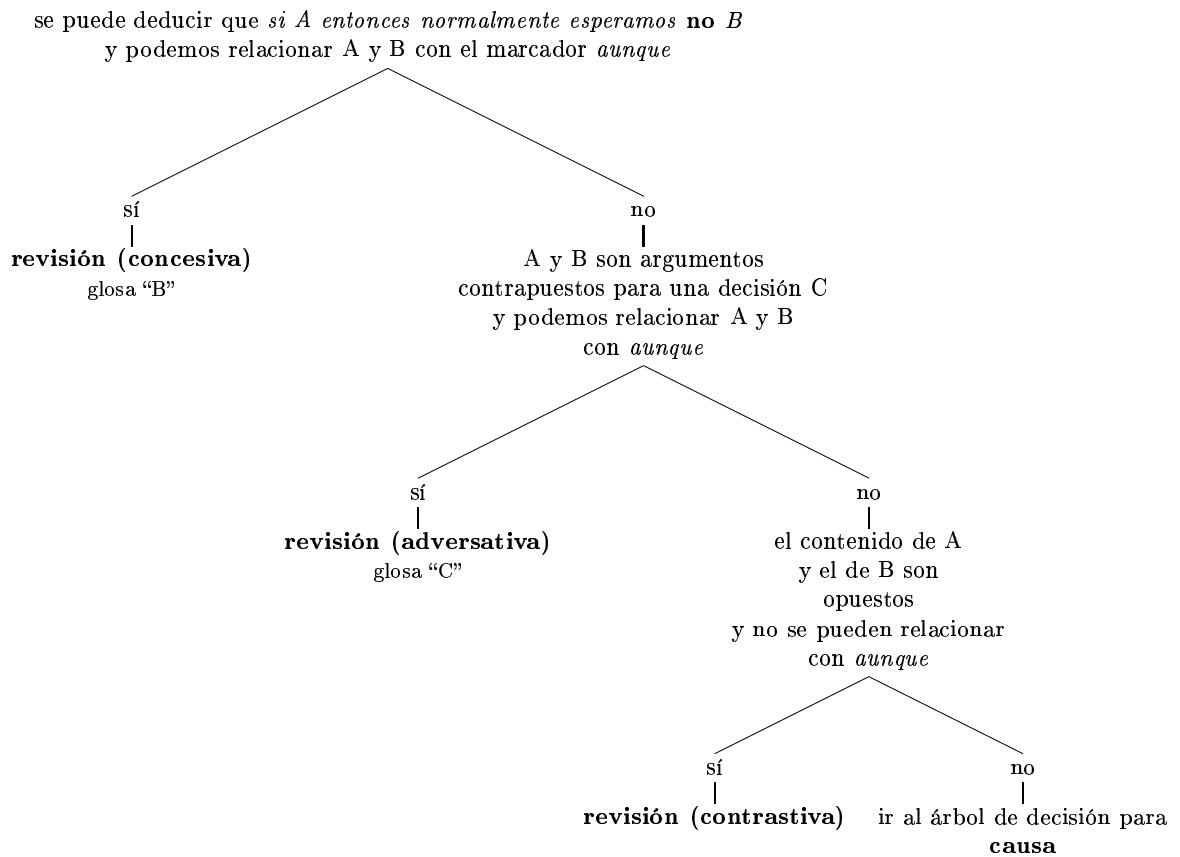


Figura 15: Árbol de decisión para la característica de revisión, donde no hay caso por defecto.

A y B se pueden relacionar mediante un marcador de causa

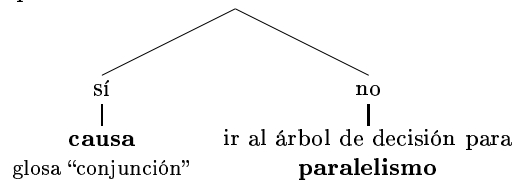


Figura 16: Árbol de decisión para la característica de causa, donde no hay caso por defecto.

A y B pertenecen a la misma clase de cosas,
y esta clase es deducible del contexto

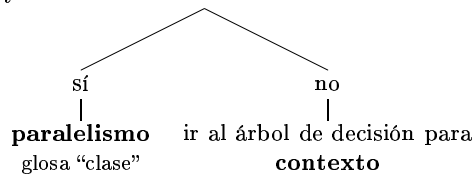


Figura 17: Árbol de decisión para la característica de paralelismo, donde no hay caso por defecto.

B responde a *cómo A?*, *cuándo A?*, *dónde A?*, etc.

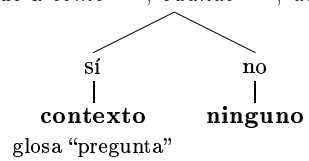


Figura 18: Árbol de decisión para la característica de contexto, donde no hay caso por defecto.

Referencias

- Carletta, J., A. Isard, S. Isard, J. Kowtko, G. Doherty-Sneddon, and A. Anderson (1996). Hcr dialog structure coding manual. Technical Report HCRC/TR-82, HCRC.
- Carlson, L. and D. Marcu (2001). Discourse tagging manual. Technical Report ISI-TR-545, ISI.
- Cooper, R., S. Larsson, C. Matheson, M. Poesio, and D. Traum (1999). Coding instructional dialogue for information states. Technical Report D1.1, The TRINDI Consortium.
- Core, M. G. and J. F. Allen (1997). Coding dialogs with the damsl annotation scheme. In Working Notes: AAI Fall Symposium on Communicative Action in Humans and Machines.
- Discourse Resource Initiative (1997). Standards for dialogue coding in natural language processing. Technical Report 167, Dagstuhl-Seminar.
- Lagerwerf, L. (1998). Causal Connectives Have Presuppositions; Effects on Coherence and Discourse Structure. Den Haag, Holland Academic Graphics.
- Umbach, C. (to appear). ontrast and information structure: A focus-based analysis of 'but'. Linguistics.