

# ANÁLISIS DE LOS MARCADORES DEL DISCURSO PARA EL EUSKERA: DENOMINACIÓN, CLASES, RELACIONES SEMÁNTICAS Y TIPOS DE AMBIGÜEDAD<sup>1</sup>

MIKEL IRUSKIETA - ARANTZA DIAZ DE ILARRAZA - MIKEL LERSUNDI  
*Universidad del País Vasco*

## RESUMEN

*El propósito de este artículo es demostrar el interés de una clasificación unificada de los marcadores del discurso para el euskera. Para hacer dicha clasificación es necesario trasladar la centralidad que le ofrece la gramática a la conjunción coordinante hacia una arquitectura discursiva basada en las relaciones semánticas: las relaciones estructurales del discurso. El efecto inmediato de este movimiento es la desclasificación de algunas conjunciones coordinantes, conectores y conjunciones subordinantes y la inclusión de aquellas que poseen una función retórica. Presentamos las paradojas de relación y de función y los diferentes tipos de ambigüedad que surgen al analizar los marcadores del discurso desde el punto de vista exclusivamente gramatical.*

*Palabras clave: Procesamiento del lenguaje natural, estructura del discurso, relaciones semánticas, marcadores del discurso, conjunción coordinante, conector, conjunción subordinante.*

## ABSTRACT

*The aim of this work is to attempt a unified classification of discourse markers in the Basque language. To get that classification it is necessary to emphasize the importance of speech over grammar. Thus the semantic relations will replace the coordinative conjunction role. The objective of this study concerns therefore the structural relations of speech. Owing to this, some coordinative conjunctions, connectives and subordinative conjunctions were taken out and all the ones with rethoric functions were put in. We point out relation and function paradoxes and different kinds of ambiguity that arise while analyzing the discourse markers from just a grammatical point of view.*

*Keywords: Natural language processing, discourse structure, discourse relations, discourse markers, coordinating conjunction, connector, subordinating conjunction.*

## 1. INTRODUCCIÓN

Un discurso es una secuencia de oraciones producida por una o más personas cuyo objetivo es producir cierta información. Para comprender cada oración es necesario contextualizarla y conocer su relación con las demás. Una de las ramas de la lingüística computacional tiene como objetivo llegar a la comprensión del discurso; para ello, es importante saber interpretar los fragmentos del discurso y, sobre todo, qué tipos de relaciones existen entre ellos. Los *marcadores del discurso* constituyen un aspecto importante a la hora de interpretar y estudiar las relaciones estructurales del discurso. Por todo esto, en este artículo analizaremos algunos ámbitos relacionados con los *marcadores del discurso*.

Comenzando con la denominación Jucker y Ziv (1998: 1) mencionan que para designar este objeto se usan diferentes términos:

A variety of terms are used to refer to these elements. Among them are discourse marker (e.g. Schiffrin 1987), pragmatic marker (e.g. Fraser 1996, Brinton 1996), discourse particle (e.g. Schourup 1985; Abraham 1991, Kroon 1995), pragmatic particle (e.g. Östman 1981), pragmatic expression (e.g. Erman 1987) or connective (e.g. Blakemore 1987, 1988).

En euskera tenemos la misma dispersión denominativa tal y como se ve en la Tabla 1.

Euskaltzaindia (1990)	Azkue (1891)	Altube (1929)	Lafitte (1929)	Arejita (1978)	Osa (1990)	Agirre (1991)	Larringan (1995)
<b>Menderagailu</b> (Conjunción subordinante)	Alkargai	Morfema modal Conjuntivo	Suffixes Prefixes	Lokarri Morfema	Marka Klausula adberbiala	Elkartitza	Testu antolatzaile
<b>Juntagailu</b> (Conjunción coordinante)	Alkargai	Conjuntivo	Coordination conjunctions	Lokarri Esaldi- lokarri	Juntagailu	Elkartitza Marka Lokarri	Testu antolatzaile
<b>Lokailu/ Esapide</b> (Conector/ locución)	Alkargai	Conjuntivo	Coordination conjunctions	Lokarri	Juntagailu	Elkartitza Marka Lokarri	Testu antolatzaile

Tabla 1: dispersión terminológica de los marcadores del discurso.<sup>2</sup>

Para realizar esta tabla hemos tomado como referencia la terminología que utiliza la Real Academia de la Lengua Vasca Euskaltzaindia (1990: 1-31; 1994: 16-17) (1ª columna): *conjunción subordinante*, *conjunción coordinante* y *conector/locución*. En las demás columnas hemos enumerado otros autores con su correspondiente denominación. Además de la falta de consenso, también existen problemas para delimitar o clasificar los *marcadores del discurso*: para Osa (1990: 250) los adverbios oracionales como “*zinez*”<sup>3</sup> y “*zorionez*”<sup>4</sup> entrarían dentro del objeto de estudio, y para Azkue (1891: 347) los elementos de comparación como “*beste*”<sup>5</sup>.

Este problema de denominación en euskera surge debido a que la aproximación metodológica no ha sido la adecuada: se ha tratado de explicar un fenómeno que trasciende a la gramática, como es la función de los *marcadores del discurso*, desde un punto de vista exclusivamente gramatical.

Aunque en la tabla no aparezca distinción entre los *conectores* y las *locuciones*, mismo grupo en la Tabla 1, Euskaltzaindia los define de un modo terminológico, aunque no funcional y, por eso, señala que no hay delimitación definitiva entre los *conectores* y las *locuciones*.

En el apartado 2. de este artículo explicaremos que, en nuestra opinión, no hay fundamento discursivo para la distinción *conjunción coordinante* vs. *conector*. Es más, esa distinción crea ciertas paradojas y puntos débiles en la explicación que comentaremos en los apartados 2.1. y 2.2. Para ejemplificar estas paradojas hemos analizado la *conjunción copulativa “eta”* en el corpus *EPEC*<sup>6</sup> construido por Aduriz *et al.* (2006: 1-15).

Además la distinción *conjunción coordinante* vs. *conector* también crea ciertas ambigüedades como son la ambigüedad de clase y la de relación que veremos en el apartado 3.

Para argumentar nuestras hipótesis (además de la *conjunción copulativa “eta”*) hemos analizado los *marcadores del discurso* que marcan “causa” y “adversidad”, ya que estas relaciones son expresadas tanto por *conjunciones coordinantes* como por *conectores*. Para ejemplificar estas ambigüedades hemos vuelto a extraer datos del corpus *EPEC*.

## 2. DIFUSIÓN DE LA DISTINCION CONJUNCIÓN COORDINANTE VS. CONECTOR

En este apartado analizaremos desde un punto de vista discursivo la necesidad o no de las clases de los *marcadores del discurso* y, si lo son, bajo qué fundamento. Vamos a realizar un acercamiento crítico sobre la descripción teórica de Euskaltzaindia (1990). Esta descripción se limita, y así lo reconoce, al punto de vista gramatical, aunque acepta que puede haber otros modos de considerar este fenómeno.

Basándonos en la división propuesta por la Academia definimos la *conjunción coordinante*, los *conectores* y la *conjunción subordinante*:

1. La *conjunción coordinante* es la única clase que junta o relaciona dos proposiciones o unidades discursivas.
2. Los *conectores* necesitan de la conjunción para juntar o relacionar dos unidades discursivas. Cuando esta conjunción no aparece explícitamente se habla de una conjunción abstracta.
3. La *conjunción subordinante* pone las unidades discursivas o proposiciones en relación de una forma especial: subordinándolas.

Euskaltzaindia analiza cada clase por separado. Mientras la *conjunción subordinante* tiene unas reglas gramaticales más diferenciadas, para nosotros, no existe diferencia tan clara entre las *conjunciones coordinantes* y los *conectores*. Esta delimitación tan clara genera dos paradojas.

### 2.1. Primera paradoja: relaciones semánticas

Siguiendo con la descripción la única relación semántica que se puede expresar con la *conjunción copulativa* “*eta*”<sup>7</sup> es la “copulativa”, pero no se pueden expresar otras relaciones como “causa”, “consecuencia”, “tiempo” y “adversidad” que tan fácilmente se expresan con los *conectores*. Para corroborar esto, hemos tomado al azar 10 ejemplos del corpus *EPEC*, y hemos visto que en el corpus sí existe tal diferencia de relaciones semánticas.

	Copulativa	Causa	Tiempo
Eta (y)	6 %60	2 %20	2 %20

Tabla 2: datos sobre las relaciones semánticas de la *conjunción* “eta”.

En esos 10 ejemplos hemos observado que existen diferencias, pues aparecen 3 relaciones semánticas: “copulativa”, “causa” y “tiempo”, tal y como se aprecia en los ejemplos [1], [2] y [3].

[1] Jon etxera joan da **eta** jaten hasi da.<sup>8</sup> “Tiempo”.

[2] Kanpora Julen etorri da **eta** Martin hotelean geratu da.<sup>9</sup> “Copulativa”.

[3] Gortinak sua hartu zuen **eta** etxe guztia erre zen.<sup>10</sup> “Causa”.

Según el acercamiento de la Academia la única relación posible es la “copulativa”, pero desde un punto de vista discursivo, es necesario distinguir dichas relaciones semánticas para analizar las relaciones entre las unidades del discurso. Por esto concluimos que la expresividad semántica de la *conjunción* “eta”, dado que puede marcar las mismas relaciones que los conectores, difumina el fundamento de la distinción *conjunción coordinante vs. conector*.

## 2.2. Segunda paradoja: delimitación de funciones

Según la Academia la *conjunción coordinante* sólo junta sintagmas u oraciones simples. Y cuando relaciona enunciados, entonces pierde su función de juntar y es necesaria la presuposición de una *conjunción abstracta* para realizar la *yuxtaposición*. Esta presuposición va en contra del *principio de economía* propuesto por Ockham.<sup>11</sup>

[4] Jon **eta** Martin etorri dira.<sup>12</sup> JUNTA.

[5] Jon etxera etorri da **eta** Martin hotelean geratu da.<sup>13</sup> JUNTA.

[6] Jon etxera etorri da. **Eta** Martin hotelean geratu da.<sup>14</sup> NO JUNTA.

Para corroborar las funciones de los *marcadores del discurso*, hemos extraído del corpus *EPEC* 20 ejemplos tomados al azar por cada *marcador* analizado. Hemos elegido los *marcadores* que expresan “contrariedad” y “causa” porque estas relaciones son expresadas por

elementos que han sido clasificados entre las *conjunciones coordinantes* y entre los *conectores*.

Hemos observado que los *marcadores del discurso* relacionan oraciones simples, enunciados y párrafos, mientras que sus homólogos relacionan sintagmas. Las unidades que relacionan sintagmas poseen función sintáctica dentro de la oración; en cambio, las unidades que unen oraciones y expresan una relación semántica poseen una función retórica. En la Tabla 3 hemos puesto los datos de los *marcadores del discurso*: “*baina*”<sup>15</sup>, “*baizik*”<sup>16</sup>, “*hala ere*”<sup>17</sup>, “*ordea*”<sup>18</sup> e “*izan ere*”<sup>19</sup>.

Euskaltzaindia		Función sintáctica	Función retórica		
		Sintagma	Oración s.	Enunciado	Párrafo
Coordinativa	<i>baizik</i>	15 80%	5 20%	--	--
<b>Coor. / subord.</b>	<b><i>baina</i></b>	1 5%	15 75%	4 20%	--
<b>Conectores</b>	<b><i>hala ere</i></b>	--	2 10%	11 55%	7 35%
	<b><i>ordea</i></b>	--	4 20%	11 55%	5 25%
	<b><i>izan ere</i></b>	--	3 15%	17 85%	--

Tabla 3: información sobre las funciones de *conjunción coordinante*, *conjunción subordinante* y *conexión*

La *conjunción coordinante* “*baizik*” tiene en un 80% de los casos veces función sintáctica, característica propia de la *conjunción coordinante*. Aunque también posee en un 20% función retórica que le permite ser considerada como un *marcador del discurso*. Sin embargo “*baina*” funciona el 95% de las veces como *marcador del discurso* y, a pesar de estar clasificada como *conjunción coordinante*, relaciona sintagmas sólo el 5% de los casos.

Con las *conjunciones subordinantes* sucede algo distinto: algunas de ellas no poseen función retórica en ninguna ocasión, y aún así son clasificadas junto con otras que sí poseen función retórica. Éste es el caso de las *oraciones sustantivas* que funcionan como sujeto y complemento.

[7] Egingo **zuela** esan zuen.<sup>20</sup>

Como hemos dicho estas paradojas surgen porque se quiere analizar un fenómeno extragramatical con herramientas gramaticales. Lo que para la gramática es insuperable, la oración compuesta, es para nosotros normalmente la unidad mínima. Por este cambio de punto de vista se da un giro por el cual la *conjunción coordinante* pierde su status privilegiado. En nuestra opinión algunas *conjunciones coordinantes* y *subordinantes* y los *conectores* marcan de la misma forma relaciones estructurales en el discurso mediante relaciones semánticas. Esta afirmación es el fundamento para el análisis unificado de los *marcadores del discurso*.

### 3. AMBIGÜEDAD

La ambigüedad se nos presenta, a primera vista, de un modo más intenso. Hay *marcadores del discurso* que participan en las tres clases: *conjunciones coordinantes*, *conectores* y *conjunciones subordinantes*. Es la información gramatical o sintáctica la que determina si el *marcador del discurso* es de una u otra clase; cuando pertenece a más de una clase hablamos de ambigüedad de clase. En cuanto a las relaciones un mismo *marcador del discurso* puede expresar más de una dependiendo del contexto, lo que produce la *ambigüedad de relación* (Iruskieta *et al.* Por publicar).

#### 3.1. Ambigüedad de clase

Un elemento puede formar parte de más de una clase. Es suficiente la información gramatical para distinguir entre las diferentes clases del *marcador del discurso* “*baina*”.

[8] Pandoraren kutxa itxi egin zuen, **baina**, han geratu zenak ez zion lasai egoteko arrazoirik ematen.<sup>21</sup>

[9] Denetatik zegoen Pandoraren kutxan, zorigaitza ugari, oso ugari **baina**.<sup>22</sup>

En los 2 ejemplos se expresa una relación de adversatividad pero en “*baina*” del ejemplo [8] en medio de las dos oraciones (tal como lo delata en este caso el uso especial de la puntuación), es una *conjunción*

*coordinante*. Mientras que el del ejemplo [9], dentro de la segunda oración, es un *conector*.<sup>23</sup>

[10] Internetez erositako ordenagailua apurtuta etorri zitzaion **eta** beste bategatik aldatu egin zuen.<sup>24</sup>

[11] Internetez erositako ordenagailua aldatu zuen, apurtuta etorri zitzaion **eta**.<sup>25</sup>

En estos 2 ejemplos se expresa una relación de “causa”. En el del ejemplo [10], el elemento de unión entre oraciones es una *conjunción coordinante* y el del [11], detrás del verbo, es una *conjunción subordinante*.

Este no será el objetivo principal de nuestro estudio, a pesar de que en algunos *marcadores del discurso* la desambiguación de la clase signifique la desambiguación de la relación semántica.

### 3.2. Ambigüedad de relación

Un elemento puede expresar más de una relación. Cuando nos encontramos frente a la ambigüedad de relación la sintaxis no siempre es útil para su desambiguación. Sólo tenemos la información del contexto o la dirección que sugieren las proposiciones. Presentaremos aquí algunos ejemplos para ilustrar este aspecto:

[12] Pandoraren kutxa ez duzula ekarri? **Bada**, itxaron egin beharko duzu.<sup>26</sup>

[13] Pandoraren kutxa ez zuen ukitu; **bada**, beldurra ematen zion handik atera zitekeenak.<sup>27</sup>

[14] Pandoraren kutxa ez zela ireki behar? **Bada**, nik ireki egin dut.<sup>28</sup>

El *marcador del discurso* “*bada*” es un *conector*, pero mientras en [12] marca la relación semántica de “consecuencia”, en [13] marca “causa” y en [14] “adversidad”.

Desambiguar el tipo de relación existente entre las unidades es uno de los objetivos principales de la lingüística computacional que trabaja en el ámbito del discurso.

La ambigüedad no afecta sólo a la clase o a la relación semántica por separado, también puede afectar a ambas:

[15] Amerikatik ekarri zuen ordenagailua **eta** bidaian hautsi egin



zitziaion.<sup>29</sup>

[16] Ordenagailua aldatu behar izan zuen, apurtuta etorri zitziaion **eta**.<sup>30</sup>

Mientras que en el [15] la *conjunción coordinante* “eta” marca la relación semántica de “consecuencia” en el [16] la *conjunción subordinante* marca la relación de “causa”. Esta *conjunción subordinante* siempre expresa “causa”, por lo que desambiguar la clase nos ayudaría a desambiguar la relación.

#### 4. CONCLUSIONES

En este artículo hemos analizado los *marcadores del discurso* del euskera desde diferentes aspectos: denominación, clases, relaciones semánticas y tipos de ambigüedades.

En cuanto a la denominación, en euskera tenemos un problema de diversidad denominativa (tal y como ocurre en otras lenguas). Este problema se debe a que se ha tratado de explicar un fenómeno que trasciende a la gramática desde un punto de vista gramatical.

Por otro lado defendemos que no existe fundamento discursivo para la distinción entre *conjunciones coordinantes* y *conectores*. Una distinción muy marcada de estos dos elementos crea dos paradojas: la de las relaciones semánticas y la de las funciones.

En cuanto a las relaciones semánticas, una distinción estricta entre *conjunciones coordinantes* y *conectores* no nos permite marcar de manera adecuada las relaciones que existen entre unidades en el corpus. Hemos visto el ejemplo de la *conjunción copulativa* “eta”, que cuando funciona como *conjunción* sólo podría expresar relaciones “copulativas”, afirmación que no se corrobora en los ejemplos del corpus.

Siguiendo con la división estricta entre *conjunciones coordinantes* y *conectores*, nos encontramos con el problema de que una *conjunción* puede tener dos funciones: sintáctica (cuando une sintagmas) y retórica (cuando une oraciones). Además, hemos visto, en algunos ejemplos del corpus, que algunas *conjunciones* aparecen uniendo enunciados. Los elementos que unen sintagmas no son considerados como *marcadores del discurso*, pero sí los que unen oraciones. Por lo tanto, una *conjunción*

*coordinante* puede ser *marcador del discurso* o no dependiendo de las unidades que junte.

Las paradojas de relación y de función surgen por querer analizar un fenómeno extragramatical con herramientas gramaticales. Lo que para la gramática es insuperable, el enunciado, es para nosotros normalmente la unidad mínima. Por este cambio de punto de vista se da un giro que hace perder a la *conjunción* su status privilegiado. En nuestra opinión las *conjunciones coordinantes* y *subordinantes* y los *conectores* marcan de la misma forma relaciones estructurales en el discurso.

Hemos tratado también las ambigüedades de clase y de relación. El objetivo de la lingüística computacional que trata el discurso es el de desambiguar las relaciones entre unidades. Estas relaciones pueden ser ambiguas dependiendo del elemento que aparece entre las unidades. La distinción entre *conjunciones coordinantes* y *conectores* crea una ambigüedad de clase, que no será el objetivo principal de nuestro estudio, a pesar de que a veces desambiguar la clase de un elemento puede ayudar a desambiguar la relación semántica que esconde.

Con este movimiento, de desclasificar algunas *conjunciones coordinantes*, *conectores* y *conjunciones subordinantes*, y clasificar nuevas *locuciones* y *organizadores discursivos*, queda patente la necesidad de un análisis unificado de los *marcadores del discurso*.

Por lo tanto, a la lingüística computacional que trabaja con el euskera se le abre una nueva dimensión, la del discurso, que puede aportar a través de la descripción del uso de los *marcadores del discurso* generalizaciones sobre la teoría del lenguaje

## NOTAS

1. Este trabajo se ha realizado gracias a los proyectos de la Universidad del País Vasco (GIU05/52) y del Gobierno Vasco (IT-397-07).
2. Mantenemos los términos en lengua original.
3. De veras, de verdad.
4. Por suerte.
5. Tanto ... como, tan ... como.
6. EPEC es el corpus de referencia que usamos para el procesamiento del euskera. Este corpus tiene 300.00 palabras. De momento ha sido etiquetado manualmente a nivel morfológico y sintáctico.

7. Conjunción copulativa “y”.
8. Jon se ha ido a casa **y** ha empezado a comer.
9. Julen ha venido fuera **y** Martín se ha quedado en el hotel.
10. La cortina cogió fuego **y** se quemó toda la casa.
11. *Pluralitas non est ponenda sine necessitate (cuanto menos se supone, mejor)*.
12. Han venido Jon **y** Martín.
13. Jon ha venido a casa **y** Martín se ha quedado en el hotel.
14. Jon ha venido a casa. **Y** Martín se ha quedado en el hotel.
15. Pero, sino, aunque.
16. Sino, tan sólo.
17. Sin embargo, a pesar de todo.
18. Pero, en cambio, sin embargo.
19. De hecho.
20. Dijo **que** lo haría.
21. Cerró la caja de Pandora, **aunque** lo que allí quedó no le daba razones para estar tranquila.
22. Había de todo en la caja de Pandora, **pero, sobre todo**, mucho infortunio.
23. Así lo dice la Academia, pero desde nuestro punto de vista, el [8] sería un *conector*, ya que está fuera de la oración, y el [9] una *conjunción*, ya no posee función retórica.
24. El ordenador que compró por Internet le vino estropeado, **y** lo cambio por otro.
25. El ordenador que compró por Internet lo cambió, **ya que** le vino roto.
26. ¿Que no has traído la caja de Pandora? **Entonces** tendrás que esperar.
27. No tocó la caja de Pandora, **ya que** le daba miedo lo que podía salir de ahí.
28. ¿Que no había que abrir la caja de Pandora? **Pues** yo la he abierto.
29. Trajo el ordenador de América **y** se le estropeó en el viaje.
30. Tubo que cambiar el ordenador, **ya que** le vino estropeado.

#### REFERENCIAS BIBLIOGRÁFICAS

- Aduriz, I., Aranzabe, M., Arriola, J., Atutxa, A., Díaz de Ilarraza, A., Ezeiza, N., Gojenola, K., Oronoz, M., Soroa, A., Urizar, R. 2006. “Methodology and steps towards the construction of EPEC, a corpus of written Basque tagged at morphological and syntactic levels for the automatic processing”. *Language and Computers* 56: 1-15. Holanda. Rodopi.
- Agirre, J. M. 1991. *Euskal gramatika deskriptiboa*. Bilbo. Labayru Ikastegia.
- Altube, S. 1929. *Erderismos*. Bermeo. Gaubeka.
- Arejita, A. 1978. *Euskal Joskera*. Durango. Labayru Ikastegia.
- Azkue, R. M. 1891. *Euskal-Izkindea*. Bilbao. Tipografía de Jose de

- Astuy.
- Bosque, I. y Demonte, V., eds. 1999. *Gramática descriptiva de la Lengua Española*. Madrid. Espasa Calpe.
- Esnaola, P. 2002. "Testu-antolatzaileak". *Hizpide* 50: 25-36.
- Euskaltzaindia. 1990. *Euskal gramatika lehen urratsak* – III. Bilbao. Euskaltzaindia.
- \_\_\_\_\_ 1994. *Euskal gramatika lehen urratsak* – IV. Bilbao. Euskaltzaindia.
- Fraser, B. 1999. "What are discourse markers". *Journal of pragmatics* 31: 931-952.
- Iruskieta, M., Diaz de Ilarraza, A. eta Lersundi, M. POR PUBLICAR. "Diskurtso-markatzaileak euskarazko testu elektronikoetan: azterketa teorikoa eta tratamendu konputazionalerako corpus-beresketa". UPV-EHU.
- Jucker, A.H. & Ziv, Y., eds. 1998. "Discourse Markers: Introduction", in Jucker, A.H. and Ziv Y., eds. 1-12.
- \_\_\_\_\_ 1998. *Discourse Markers: Descriptions and theory*. Amsterdam. John Benjamins.
- Lafitte, P. 1929. *Grammaire basque*. Bayonne. Courrier.
- Larringan, L.M. 1995. Testu-antolatzaileak bi testu-motatan: testu informatiboa eta argudiazpenezkoa. Tesis doctoral. Vitoria-Gasteiz. UPV-EHU.
- Martín Zorraquino, M.A. y Portolés Lázaro J. 1999. "Los marcadores del discurso", en Bosque I. y Demonte V., eds. Vol. 3, 4051-4213.
- Osa, E. 1990. Euskararen hitzordena: komunikazio zereginaren arauera. Leioa. UPV-EHU.
- Portolés, J. 1998. Marcadores del discurso. Barcelona. Ariel.