

EUDIMA CORPUSETIK ADIBIDEAK ERAUZTEKO LAN-TRESNA: UNITATE FRASEOLOGIKO-DISKURTSIBOAK AZTERTZEKO LANABESA¹

Xabier Alberdi

(Universidad del País Vasco/Euskal Herriko Unibertsitatea, UPV/EHU)

Josu Landa

(Ametzagaiña I+G+B)

[**AURKIBIDEA**](#)

Laburpena

Lan honetan diskurtso-markatzaileei buruzko ikerketa-proiektu baterako² garatu dugun lan-tresna bat aurkeztu dugu. “[EUDIMA corpusetik adibideak erauzteko lan-tresna](#)” da aplikazioaren behin-behineko izena eta, nahiz eta berariaz diskurtso-markatzaileak aztertzeke diseinatuta egon, guztiz baliagarria izan liteke etorkizunean —behin hobekuntza eta egokitzapen batzuk eginda— egungo euskararen erabiltzen diren beste unitate fraseologiko-diskurtsibo aldagaitz batzuk ikertzeke, hala nola lokuzioak, kolokazioak, hizketa-fórmulak eta paremiak.

Gako-hitzak: diskurtso-markatzaileak, birformulatzailak, unitate fraseologiko-diskurtsiboak, corpusak, adibideen erauzketa, lan-tresna.

1. Sarrera

Artikulu honetan UPV/EHUK finantzaturako “Euskarazko diskurtso-markatzaileen hiztegia (I)” ikerketa-proiekturako garatu dugun lanabes bat aurkeztuko dugu. “[EUDIMA³ corpusetik adibideak erauzteko lan-tresna](#)” da lanabesaren behin-behineko izena eta, nahiz eta berariaz diskurtso-markatzaileak aztertzeke diseinatuta egon, guztiz baliagarria izan liteke diskurtsoaren mailako beste unitate fraseologiko-diskurtsibo batzuk ikertzeke.

Esan bezala, gure ikerketa-proiektuaren helburua zen diskurtso-markatzaile mota bat aztertzea: birformulatzailak. Zehatzago esanda, euskarazko birformulatzailen hiztegi bat egitea zen helburua: Briz, Pons eta Portolés-ek (2008) koordinatu duten *Diccionario de partículas discursivas del español* eta Fuentes-en (2009) *Diccionario de conectores y*

¹ Artikulu hau Universidad del País Vasco/Euskal Herriko Unibertsitateak (UPV/EHU) ‘Unibertsitatea-Gizartea’ modalitatean finantzatu duen US10/10 “Euskarazko diskurtso-markatzaileen hiztegia (I) / Diccionario de marcadores discursivos del euskera (I)” ikerketa-proiektuaren emaitzariko bat da. Euskal Irrati Telebistarekin (EITB) eta Elhuyar Fundazioarekin sinatu ditugun lankidetzeta-hitzarmenei esker garatzen ari gara proiektua. Artikulua, halaber, ni partaide naizen IT665-13 talde egonkortuaren ikerketaren emaitzen artean sartzen da.

² UPV/EHUK finantzaturako US10/10 ikerketa-proiektua (‘EUDIMA’).

³ ‘EUDIMA’ da UPV/EHUK finantzaturako US10/10 ikerketa-proiektuaren izen osoaren akronimoa.

operadores del español hiztegien tankerako hiztegi prototipo bat egitea zen gure helburua. Eta horretarako corpus zabal bat osatzea erabaki genuen.

Gure ikerlanerako osatu dugun corpusak —EUDIMA corpusak— egungo euskara zaindua islatu nahi du eta honela osatuta dago⁴:

| AZPICORPUSA | TESTU-HITZ KOPURUA | EZAUGARRIAK |
|--|-----------------------|---|
| ZIO corpusa ⁵ | 1,0 milioi | - hizkuntza idatzia |
| UPV/EHUko eskuliburu itzuliak | 5,5 milioi | - azalpenezko testuak |
| Zuzenbidea / lege-testuak | 4,1 milioi | - jatorrizkoak/itzuliak |
| Zuzenbidea / testu akademikoak | 2,5 milioi | - erregistro zaindua |
| Pentsamenduaren Klasikoak ⁶ | 10,7 milioi | |
| ZT Corpusa ⁷ | 6,6 milioi | |
| EPD liburuak ⁸ | 10,2 milioi | - eleberria / saiakera - jatorrizkoak/itzuliak |
| EPD prentsa ⁹ | 10,5 milioi | - prentsa |
| ETBko dokumentalak | 0,5 milioi | - entzunak izateko idatziak - zabalkunderako - itzuliak - erregistro zaindua |
| Goenkale corpusa ¹⁰ | 11,0 milioi | - mintzatuaren “imitazioa” - jatorrizkoak - lagunartekoaren “imitazioa” |
| GUZTIRA | 62,6 milioi | |

1.a. taula. EUDIMA corpusaren osaera

⁴ Corpusari dagokionez, eskerrak eman behar dizkiogu lehenik UPV/EHUko Euskara Institutuaren zuzendari Pello Salabururi (<http://www.ei.ehu.es>), Institutuaren corpusak gure zerbitzura jarri dituelako (ZIO, Pentsamenduaren Klasikoak, Ereduzko Prosa Dinamikoak eta Goenkale). UPV/EHUko Euskara Zerbitzuaren zuzendari Jesus Mari Makazagari eskertu behar diogu, halaber, EHUko ikasliburu itzuliak gure eskura jarri izana. EITBrekiko lankidetzari esker eta, zehazki, Asier Larrinagaren bitartekotzari esker eskuratu dugu ETBko dokumentalen corpusa. Elhuyar Fundazioarekiko lankidetzaren baten bidez eskuratu ditugu Zientzia eta Teknologiaren corpuseko testuak (ZTC). Eta, azkenik —baina ez azken—, Joseba Ezeiza sailkide eta adiskideari zor dizkiogu zuzenbideko testuen corpusak (Garalex corpusa: <http://www.hizlan.org/db-zuzenbidea/>).

⁵ <http://www.ehu.es/ehg/zio/>

⁶ <http://ehu.es/ehg/pkc/>

⁷ Zientzia eta Teknologiaren Corpusa: <http://www.ztcopusa.net/cgi-bin/kontsulta.py>

⁸ <http://ehu.es/ehg/epd/>

⁹ <http://ehu.es/ehg/epd/>

¹⁰ <http://ehu.es/ehg/goenkale/>

| | |
|--------------|------------------------|
| Jatorrizkoak | 35,7 milioi testu-hitz |
| Itzulpenak | 26,9 milioi testu-hitz |
| GUZTIRA | 62,6 milioi testu-hitz |

1.b. taula. EUDIMA corpusaren osaera

Guztira 62,6 milioi testu-hitz ditu corpusak eta halako oreka bat lortu da testu itzulien eta jatorrizkoen artean. Testu mota batzuk biltzen ditu (akademikoak, lege-testuak, prentsa, saiakera) eta, bereziki, ahozko hizkuntzara hurbiltzen diren bi azpicorpus ditu: alde batetik, Goenkale, zeinak hein batean lagunarteko erregistroa imitatzen baitu; eta, bestetik, ETBko dokumentalen corpusa, zeinak, entzunak izateko testuen bidez, erregistro zaindua islatzen baitu.

Behin EUDIMA corpusa osatu ondoren, corpusetik adibideak —diskurtso-markatzaileen adibideak— erauzteko lanabes bat behar genuen, eta testuinguru hartan sortu zuen Josu Landak orain hizpide dugun lan-tresna, ikertaldearen iradokizun eta beharrak kontuan harturik.

2. Lan-tresnaren funtzionalitateak (<http://www.ehu.es/ehg/eudima/corpusa/>)

Gorago azaldu dugunez, diskurtso-markatzaileen adibideak erauzteko pentsatuta dago lan-tresna hau eta, markatzaileok unitate aldagaitzak —hots, hitz edo esapide finkatuak (erdi)gramatikalizatuak edo (erdi)lexikalizatuak— direnez, lematizatu gabe dago EUDIMA corpusa. Corpusetik markatzaile birformulatuak erauztea zen gure premia: hau da, 2. taulan agertzen diren unitateen modukoak erauztea (ikus Alberdi 2011a, 2011b, Alberdi 2013a, Alberdi & Ezeiza 2013, Ezeiza & Alberdi 2013). Beraz, lan-tresna honen bidez lexikalizatuta edo lexikalizatzeko bidean dauden hitz eta esapideak erauz ditzakegu corpusetik (1. irudian ikus daiteke aplikazioaren interfazea).

| BIRFORMULATZAILEAK | |
|--|--|
| a) Esplikatiboak | <i>alegia, hau da, hots, erran nahi baita (...), bestela esanda, beste era batera esanda (...), beste modu batera esanda (...), beste hitz batzuetan esanda (...), beste hitz batzuetan...</i> |
| b) Zehaztaileak (“de concreción”) | <i>zehatzago esanda (...), zehatz-mehatz, zehazkiago, hurrenez hurren...</i> |
| c) Laburbiltzaileak (gazt. “recapitulativos”) | c1) Laburbiltzaile hutsak <i>laburbilduz, laburbilduta (...), labur esanda, labur esateko (...), hitz batean, hitz gutxitan, hitz batean esateko...</i> |
| | c2) Berrazterketa edo berrausnarketa adieraztekoak («de reconsideración») <i>azken batean, azken finean, azken buruan...</i> |
| d) Urruntzaileak (gazt. “de distanciamiento”) | <i>hala eta guztiz ere (...), dena dela, dena den (...), edozein modutan (ere)(...), nolana (ere) (...), edozelan (ere) (...), edozein kasutan (ere), esanak esan...</i> |
| e) Zuzentzaileak (gazt. “de corrección”) | <i>hobeto esanda (...), (edo) hobeto, are hobeto, are hobeki (...), barkatu, zer diot? ...</i> |

2. taula: Euskarazko diskurtso-markatzaile birformulatzaileen sailkapena

www.ehu.es/ehg/eudima/corpusa/ mac imprimir pantalla

ES • EU • Euskalterm • Elhuyar • Labayru • Zehazki • ZT Hiztegia • Energia • Telekomunikazioak • Danobat • Batua • OEH • Sin.UZEI • ItzuL • Harluxet • Wikipedia • Mokoroa • Intza • Eurovoc • Bergara • EPG • Klasikoak • ZT Corpusa • Lexikoaren Behatokia

banakoak multzoak

—berria txertatu—

1. **agerian utzi** 613 [0.97]
1.60 | 0.83 | 0.36 | 0.48 | 0.40
0.80 | 3.11 | 0.62 | 1.00 | 0.23

2. **agerian utzi+** 0 [0.00]
0.00 | 0.00 | 0.00 | 0.00 | 0.00
0.00 | 0.00 | 0.00 | 0.00 | 0.00

3. **aitzitik** 5349 [8.54]
17.30 | 6.41 | 4.14 | 8.24 | 29.11
5.33 | 2.86 | 7.00 | 2.20 | 0.12

4. **alde hartatik** 18 [0.02]
0.20 | 0.00 | 0.00 | 0.00 | 0.04
0.07 | 0.00 | 0.01 | 0.20 | 0.00

5. **alde horretatik** 1121 [1.79]
4.60 | 0.56 | 0.70 | 2.08 | 1.91
1.50 | 3.40 | 1.16 | 1.40 | 1.48

6. **alderantziz** 4278 [6.83]
9.80 | 6.65 | 1.36 | 5.28 | 15.49
5.98 | 2.60 | 9.98 | 1.60 | 3.79

7. **alegia** 19544 [31.22]
94.90 | 21.34 | 9.60 | 73.68 | 61.91
33.05 | 16.29 | 37.31 | 11.20 | 8.70

8. **are gehiago** 4104 [6.55]
11.40 | 11.36 | 2.00 | 7.60 |

EUDIMA - Corpusetik erazutako adibideak

—2.1 bertsioa (12-02-13)—

alegia

| | agerpenak | maiztasuna | | agerpenak | maiztasuna |
|--------------------------------|--------------|--------------|--------------|-----------|------------|
| A1. ZIO corpusa | 949 | 94.90 | | | |
| A2. EHUko eskuliburu itzuliak | 1174 | 21.34 | | | |
| A3a. Zuzenbidea / lege-testuak | 394 | 9.60 | | | |
| A3b. Zuzenbidea / akademikoak | 1842 | 73.68 | jatorrizkoak | 8486 | 23.77 |
| A4. Pentsamenduaren Klasikoak | 6625 | 61.91 | itzulpenak | 11058 | 41.10 |
| B1a. EPD liburuak | 3372 | 33.05 | | | |
| B1b. EPD prentsa | 1711 | 16.29 | | | |
| B2. ZT Corpusa | 2463 | 37.31 | | | |
| C1. ETBko dokumentalak | 56 | 11.20 | | | |
| C2. Goenkale corpusa | 958 | 8.70 | | | |
| denera | 19544 | 31.22 | | | |

testuinguru murrizketa: | eten mota: | bidali

ZIO EHU Zuzen. LT Zuzen. TA PKC EPD lib. EPD pren. ZTC ETB doku. Goenkale

agerpenak: 949 | liburuak: 14 | maiztasuna: 94.90

Loroaren teorema, Denis Guedj / Jon Muñoz (EHU, 2005)

- Jonathan-eta-Léa antz-antzekoak ziren, nola diren elkarren antzeko bi ur tanta; **alegia**, tantek bezala, ez zuten elkarren antzik batera. Orr.: 20 (0-0)
- Talesek, berriz, bestelakorik egin zuen: bere buruari galderak egin, **alegia**. Orr.: 33 (0-0)
- Harrera egin zion liburuzain andreak galdetu zion ea irakaslea edo ikertzailea zen; ea ikerlan bat egiten ari zen, **alegia**, eta zein ikerlan zen, eta zein irakasleren gidaritzapean; ea ikaslextartelak zuenetz; ea... Orr.: 34 (0-0)
- Kareleri helduta, Joniako lur gero eta urrunagoari begira zegoen Tales; ordu arte bizileku izan zuen tokiari, **alegia**. Orr.: 43 (0-0)
- "Itzalak zirkunferentzia doi-doi ukitzen - itzalaren luzera bere neurriaren berdina zenean, **alegia** -, aurrez adostutako garrasia egin zuen. Orr.: 45 (0-0)
- Heriva du, bain zuzen, berezitasun geometriko: oinarri bat izateko, non barmatua **alegia**. Orr.: 46 (0-0)

1. irudia: lan-tresnaren interfazea

2.1. Sarrerak eraztea

Lan-tresna honen bidez, lehenik, aztergai ditugun hitz eta esapideak erazuzi behar ditugu: gure ikerketan, birformulatzaileak¹¹. Bigarren irudian ezker aldeko erlaitzean ageri dira EUDIMA proiekturako erazuzi ditugun sarrerak¹².

The screenshot shows the EUDIMA web interface. On the left, there is a list of terms with their respective counts and various statistics. A red box labeled 'SARRERAK' points to the first item, 'agerian utzi'. The main content area displays the detailed view for the term 'alegia', including a table of counts for different categories and a list of example sentences.

| | agerpenak | maiztasuna | | agerpenak | maiztasuna |
|--------------------------------|--------------|--------------|--------------|-----------|------------|
| A1. ZIO corpusa | 949 | 94.90 | | | |
| A2. EHUko eskuliburu itzuliak | 1174 | 21.34 | | | |
| A3a. Zuzenbidea / lege-testuak | 394 | 9.60 | | | |
| A3b. Zuzenbidea / akademikoak | 1842 | 73.68 | jatorrizkoak | 8486 | 23.77 |
| A4. Pentsamenduaren Klasikoak | 6625 | 61.91 | itzulpenak | 11058 | 41.10 |
| B1a. EPD liburuak | 3372 | 33.05 | | | |
| B1b. EPD prentsa | 1711 | 16.29 | | | |
| B2. ZT Corpusa | 2463 | 37.31 | | | |
| C1. ETBko dokumentalak | 56 | 11.20 | | | |
| C2. Goenkale corpusa | 958 | 8.70 | | | |
| denera | 19544 | 31.22 | | | |

2. irudia: sarrerak erazteko aukera






¹¹ Birformulatzaileen multzoan sartzekoak dira lan-tresnan erazuzitako sarrera gehienak, baina ez guzti-guztiak.

¹² EUDIMA proiektuko ikertaldearen helburuetarako diseinatuta dago lan-tresna eta, beraz, oraingoz EUDIMAKO ikertzaileek baino ezin dituzte erazuzi edo ezabatu sarrerak. Etorkizunean funtzionalitate horiek ere ikertzaileen eskura nola jarri aztertuko dugu (ikus 3. atala).

2.2. Sarrerak erauztea eta eguneratzea

Hirugarren irudian ikus daitekeenez (infra 3. irudia), aplikazioak aukera ematen digu sarrera berriak gaineratzeko (ezkerreko erlaintzeko “berria txertatu” aukeratuta) eta sarrerak ezabatzeko (zakarrontziaren irudian klik eginda).

EUDIMA - Corpusetik erauzitako adibideak

| banakoak | multzoak |
|--|---|
| —berria txertatu— | |
| 1.  agerian utzi+ | 0 [0.00] 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 |
| 2.  alegia | 19544 [31.22] 94.90 21.34 9.60 73.68 61.91 33.05 16.29 37.31 11.20 8.70 |
| 3.  are gehiago | 4104 [6.55] 11.40 11.36 2.00 7.60 9.51 7.29 7.20 5.39 2.40 1.88 |
| 4.  are hobeki | 37 [0.05] 0.30 0.00 0.00 0.00 0.20 0.10 0.00 0.00 0.00 0.00 |
| 5.  are hobeto | 61 [0.09] 0.00 0.20 0.00 0.00 0.24 0.13 0.01 0.09 0.00 0.01 |

3. irudia: sarrerak eguneratzeko aukera

2.3. Sarrera bakoitzaren agerpen kopurua eta maiztasuna

Beheragoko irudian ikus daitekeenez (4. irudia), behin ezkerreko erlaintzeko sarrera baten gainean klik eginda (*alegia* sarreran, adibidez), azpicorpusez azpicorpus eta orotara ematen dizkigu aplikazioak sarrera horren agerpen kopurua eta maiztasuna. Maiztasunari dagozkion datuak 100.000kotan emanak daude.

EUDIMA - Corpusetik erauzitako adibideak

—2.1 bertsioa (12-02-13)—

| | alegia | | | agerpenak | maiztasuna |
|--------------------------------|--------------|--------------|--------------|-----------|------------|
| | agerpenak | maiztasuna | | | |
| A1. ZIO corpora | 949 | 94.90 | | | |
| A2. EHUko eskuliburu itzuliak | 1174 | 21.34 | | | |
| A3a. Zuzenbidea / lege-testuak | 394 | 9.60 | | | |
| A3b. Zuzenbidea / akademikoak | 1842 | 73.68 | jatorrizkoak | 8486 | 23.77 |
| A4. Pentsamenduaren Klasikoak | 6625 | 61.91 | itzulpenak | 11058 | 41.10 |
| B1a. EPD liburuak | 3372 | 33.05 | | | |
| B1b. EPD prentsa | 1711 | 16.29 | | | |
| B2. ZT Corpora | 2463 | 37.31 | | | |
| C1. ETBko dokumentalak | 56 | 11.20 | | | |
| C2. Goenkale corpora | 958 | 8.70 | | | |
| denera | 19544 | 31.22 | | | |

4. irudia: sarrera bakoitzaren agerpen kopurua eta maiztasuna

2.4. Sarrera bakoitzaren adibideak azpicorpusez azpicorpus

Ondorengo irudian ikus daitekeenez (5. irudia), behin sarrera bat hautatu ondoren, aplikazioak azpicorpusez azpicorpus erakusten dizkigu adibideak.

testuinguru murrizketa: | eten mota:

| | | | | | | | | | |
|-----|-----|-----------|-----------|-----|----------|-----------|-----|-----------|----------|
| ZIO | EHU | Zuzen. LT | Zuzen. TA | PKC | EPD lib. | EPD pren. | ZTC | ETB doku. | Goenkale |
|-----|-----|-----------|-----------|-----|----------|-----------|-----|-----------|----------|

agerpenak: 949 | liburuak: 14 | maiztasuna: 94.90

Loroaren teorema, Denis Guedj / Jon Muñoz (EHU, 2005)

- Jonathan-eta-Léa antz-antzekoak ziren, nola diren elkarren antzeko bi ur tanta; **alegia**, tantek bezala, ez zuten elkarren antzik batere. Orr.: 20 (0-0)
- Talesek, berriz, bestelakorik egin zuen: bere buruari galderak egin, **alegia**. Orr.: 33 (0-0)
- Harrera egin zion liburuzain andreak galdetu zion ea irakaslea edo ikertzailea zen; ea ikerlan bat egiten ari zen, **alegia**, eta zein ikerlan zen, eta zein irakasleren gidaritzapean; ea ikasletxartelik zuenetz; ea... Orr.: 34 (0-0)
- Karelari helduta, Joniako lur gero eta urrunagoari begira zegoen Tales; ordu arte bizileku izan zuen tokiari, **alegia**. Orr.: 43 (0-0)
- "Itzalak zirkunferentzia doi-doi ukitzean - itzalaren luzera bere neurriaren berdina zenean, **alegia** -, aurrez adostutako garrasia egin zuen. Orr.: 45 (0-0)

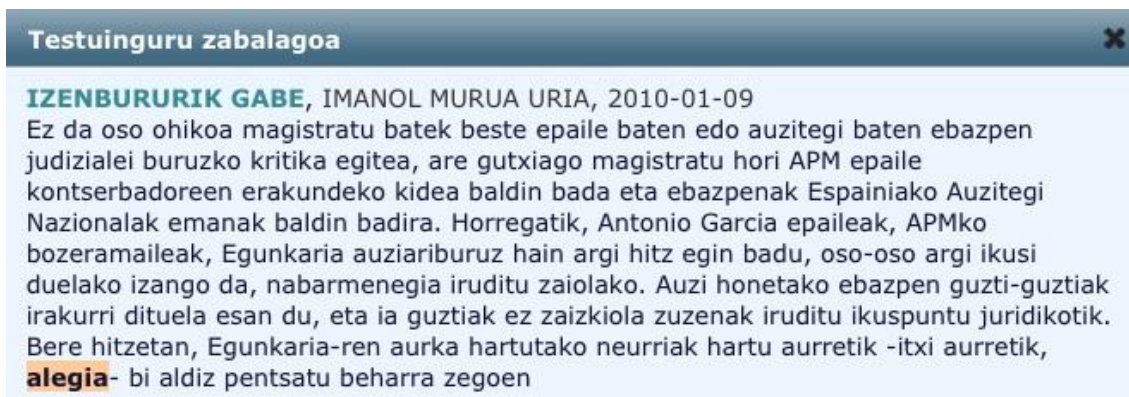
5. irudia: adibideak azpicorpusez azpicorpus

2.5. Sarrera bakoitzaren adibideak: testuinguru murriztua eta testuinguru zabalagoa

Ondorengo irudietan ikus daitekeenez, adibide bakoitzaren ondoren ageri den laukitxoaren gainean klik eginda (6.a. irudia), adibidearen testuinguru zabalagoa (6.b. irudia) erakusten digu lan-tresnak. Gure ikerketan ezinbestekoa zen testuinguru zabalagoa ikusteko aukera izatea. Izan ere, hartzaileari diskurtsoetik inferentziak ateratzen laguntzeko gida edo jarraibide moduko batzuk eskaintzea da diskurtso-markatzaileen¹³ eginkizun nagusia eta, beraz, haien eginkizun diskurtsiboaz jabetzeko ezinbestekoa da diskurtso-markatzaile bakoitzaren aurreko eta ondorengo enuntziatuak ikustea.



6.a. irudia: adibideen testuinguru murriztua



6.b. irudia: adibideen testuinguru zabalagoa

¹³ Martín Zorraquinok eta Portolésesek (1999: 4057) honela definitzen dituzte diskurtso-markatzaileak: “Los 'marcadores del discurso' son unidades lingüísticas invariables, no ejercen una función sintáctica en el marco de la predicación oracional —son, pues, elementos marginales— y poseen un cometido coincidente en el discurso: el de guiar, de acuerdo con sus distintas propiedades morfosintácticas, semánticas y pragmáticas, las inferencias que se realizan en la comunicación.”

2.6. Agerpen kopurua eta maiztasuna: jatorrizko testuak / itzuliak

Ondorengo irudian (7. irudia) ikus daitekeenez, sarrera batek testu itzulietan eta jatorrizkoetan dituen agerpen kopurua eta maiztasuna ematen dizkigu lan-tresnak. Datu estatistiko hori kontuan hartzekoa eta are adierazgarria ere izan daiteke hizkuntzaren bilakaera aztertzeko: *beste hitz batzuetan esanda* birformulatzailerak eta beraren aldaerak (*beste hitz batzuetan esateko*, *beste hitz batzuekin esanda/esateko*), esate baterako, azkozaz ere maizago agertzen dira testu itzulietan (90 agerpen) jatorrizkoetan baino (28 agerpen) eta agerpen kopuru apala (118 agerpen) dute oro har; horrek guztiak pentsarazten digu beharbada itzulpenetaren eraginez —kalko bidez— “sortu” dugun diskurtso-markatzailea izan daitekeela (Alberdi 2013a). *Alegiari* buruzko datuek ere (7. irudia) arrastoren bat ematen digute azterketa kontrastiboei ekiteko: itzulpenetan maizago erabiltzen dela ikusita, egiaztatu beharreko hipotesi bat da pentsatzea erdaretako egitura sintaktiko bihurriak —erlatiboak, adibidez— itzultzeko ere erabili ote den.

| | agerpenak | maiztasuna |
|--------------|-----------|------------|
| jatorrizkoak | 8486 | 23.77 |
| itzulpenak | 11058 | 41.10 |

7. irudia: *alegia* birformulatzailerak (agerpenak eta maiztasuna)

2.7. Testuinguru-murrizketak: etena, hasiera/bukaera, koma

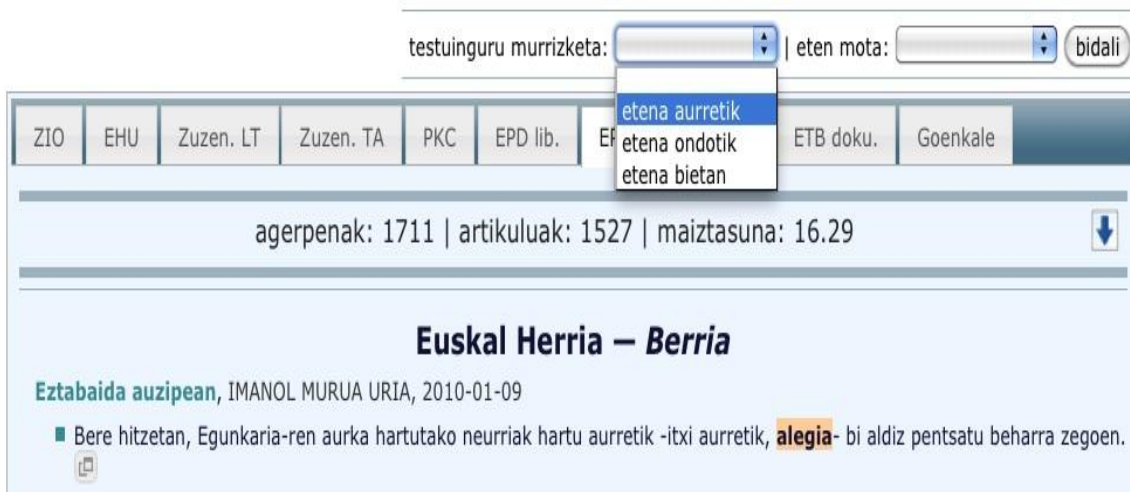
Bilaketak testuinguruaren arabera murrizteko eta diskurtso-markatzaileak markatzaile ez direnetatik bereizteko bide batzuk eskaintzen ditu lan-tresnak. Izan ere, diskurtso-markatzaileen ezaugarri bat da enuntziatuetan bakartuta agertzea, tartekien posizioan (cfr. Martín Zorraquino 2010: 104-113; Martí 2011: 18): “[Los marcadores discursivos] Distribucionalmente, aparecen en el enunciado aislados, ocupando la posición de inciso, la propia de modificadores, explicaciones y paréntesis.”

Hala, bada, lan-tresnak eskaintzen dituen testuinguru-murrizketetariko bat etenetan oinarritzen da (ikus 8a irudia), etentzat harturik edozein puntuazio-zeinu¹⁴. Bilaketa murrizteko aukera hauek eskaintzen ditu aplikazioak: “etena aurretik”, “etena ondotik”, “etena bietan”. Esate baterako, izen diren *alegiak* birformulatzailerak direnetatik bereizteko estrategia bat izan daiteke puntuazioa kontuan hartzea: adibidez, etena aurretik eta ondotik duten *alegia* gehienak birformulatzailerak izango dira. Orobat, *hurrenez hurren* esapidearen adibideen artetik birformulatzailerak (gazt. “respectivamente”) erauzteko¹⁵ eta adberbio-erabilerak (“bata bestearen ondotik, ondoz ondo”) alde batera uzteko¹⁶ estrategia emankorra izan daiteke eten motaz baliatzea: etena aurretik eta ondotik duten *hurrenez hurren* gehienak izango dira birformulatzailerak.

¹⁴ Koma, puntua, gidoia, parentesia...; baita perpaus-hasiera eta perpaus-bukaera ere.

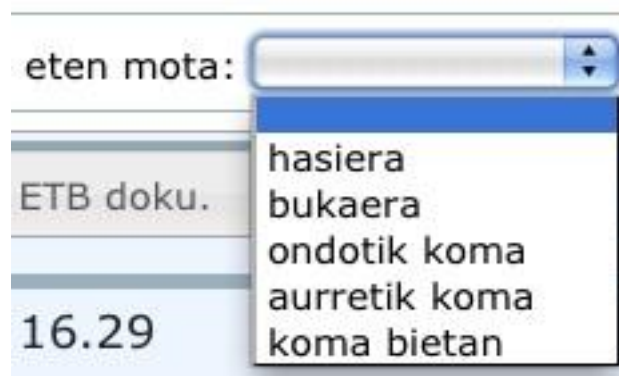
¹⁵ Birformulatzailerak erabileraren adibidea: “Gaztelaniatzko lanei dagokionez, 400 eta 280 aurkeztu dira, *hurrenez hurren*, narratiban eta poesian.”

¹⁶ Adberbio-erabileraren adibidea: “Larrutan bildu eta izoztutako hiru amuarrain artiko (Salvelinus alpinus) *hurrenez hurren* jarri, (...)”



8.a. irudia: testuinguru-murrizketa (etena)

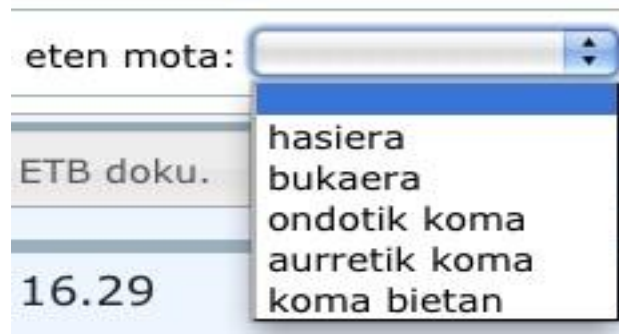
Posizioa da diskurtso-markatzaileen azterketan kontuan hartu beharreko beste alderdi bat, eta lan-tresna honek aukera ematen digu bilaketetan hasierako eta bukaerako posizioak¹⁷ bereizteko (ikus 8.b. irudia): horrelako murrizketa baten bidez egiazta dezakegu, esate baterako, *hurrenez hurren* markatzailea —*hurrenez hurren* adberbioa ez bezala— nekez agertu ohi dela perpaus-hasieran (maiz agertzen da, ordea, perpaus-bukaeran edota perpausaren erdian —enumerazio baten ondoren eta atal birformulatzailearen hasieran—); *bestela esanda* markatzaileak ere, perpausak lotzen dituenen, perpaus-hasieran agertu behar du ia ezinbestean.



8.b. irudia: testuinguru-murrizketa (hasiera/bukaera)

Lan-tresnak aukera ematen digu, orobat, puntuazioa —komaren erabilera— aztertzeko (ikus 8.c. irudia). Gorago azaldu denez, perpausaren sintaxi-egituratik aparte edo tarteki gisa agertu ohi dira diskurtso-markatzaileak eta hori hizkuntza idatzian koma puntuazio-zeinuaren bidez islatzen da: aurretik etena eta ondotik koma duten *bestela esan* gehienak, adibidez, markatzaile izango dira eta ez predikatu arruntak. Bestalde, oro har, funtzionalitate honi esker markatzaileekin dauden puntuazio-sistemak eta -usadioak azter daitezke.

¹⁷ Lan-tresnak perpaustzat jotzen ditu parentesi eta gidoi artean datozenak ere.



8.c. irudia: testuinguru-murrizketa (puntuazioa: koma)

2.8. Sarrerak multzoka antolatzeoko aukera

Sarrera baten aldaerak multzoka antolatzeoko aukera ematen digu lan-tresnak (ikus 9.a. irudia) eta aldaera bakoitzaren agerpen kopurua ikusteko aukera (9.b. irudia). Funtzionalitate horri esker, esaterako, badakigu *beste era batera esanda* dela, alde handiz gainera, dabilzan sintaxi-egituren artetik erabiliena (*beste era batean esanda, beste era batera esateko, beste era batera esateko, beste era batera esanik...*) eta lexikalizatzeko-bidean dagoela. Beraz, sarrerak multzoka aztertzeoko funtzionalitateari esker bereiz ditzakegu sarrera nagusiak eta haien aldaerak (ikus 3. taula).

EUDIMA - Corpusetik erazitako adibideak

—2.1 bertsioa (12-02-13)—

| alegia | | | |
|--------------------------------|--------------|--------------|-------------------------|
| agerpenak | maiztasuna | | |
| A1. ZIO corpusa | 949 | 94.90 | |
| A2. EHUko eskuliburu itzuliak | 1174 | 21.34 | |
| A3a. Zuzenbidea / lege-testuak | 394 | 9.60 | agerpenak maiztasuna |
| A3b. Zuzenbidea / akademikoak | 1842 | 73.68 | jatorrizkoak 8486 23.77 |
| A4. Pentsamenduaren Klasiakoak | 6625 | 61.91 | itzulpenak 11058 41.10 |
| B1a. EPD liburuak | 3372 | 33.05 | |
| B1b. EPD prentsa | 1711 | 16.29 | |
| B2. ZT Corpusa | 2463 | 37.31 | |
| C1. ETBko dokumentalak | 56 | 11.20 | |
| C2. Goenkale corpusa | 958 | 8.70 | |
| denera | 19544 | 31.22 | |

testuinguru murrizketa: | eten mota: | bidali

ZIO EHU Zuzen. LT Zuzen. TA PKC EPD lib. EPD pren. ZTC ETB doku. Goenkale

agerpenak: 949 | liburuak: 14 | maiztasuna: 94.90

Loroaren teorema, Denis Guedj / Jon Muñoz (EHU, 2005)

- Jonathan-eta-Léa antz-antzekoak ziren, nola diren elkarren antzeko bi ur tanta; **alegia**, tantek bezala, ez zuten elkarren antzik batere. Orr.: 20 (0-0)
- Talesek, berriz, bestelakorik egin zuen: bere buruari galderak egin, **alegia**. Orr.: 33 (0-0)
- Harrera egin zion liburuzain andreak galdetu zion ea irakaslea edo ikertzailea zen; ea ikerlan bat egiten ari zen, **alegia**, eta zein ikerlan zen, eta zein irakasleren gidaritzapean; ea ikasletxartelik zuenetz; ea... Orr.: 34 (0-0)
- Kareleri helduta, Joniako lur gero eta urrunagoari begira zegoen Tales; ordu arte bizileku izan zuen tokiari, **alegia**. Orr.: 43 (0-0)
- *Itzalak zirkunferentzia doi-doi ukitzean - itzalaren luzera bere neurriaren berdina zenean, **alegia** -, aurrez adostutako garrasia egin zuen. Orr.: 45 (0-0)
- Horixe du, bain zuzen, bereizitacun geometrikoa: pinacri bat iratza: non bermatua **alegia**. Orr.: 46 (0-0)

9.a. irudia: sarrera baten aldaerak multzoka antolatzeoko aukera

| | | | | |
|----|---|--------------------------------|-----|--------|
| 3. |   | beste era batera esanda | 462 | [0.73] |
| | | beste era batean esanda | 40 | |
| | | beste era batera esanda | 388 | |
| | | beste era batera esanez | 6 | |
| | | beste era batera esanik | 6 | |
| | | beste era batera esateko | 22 | |

9.b. irudia: sarrera baten aldaerak multzoka antolatzeko aukera

| BIRFORMULATZAILE ESPLIKATIBOAK: SARRERA NAGUSIAK ETA ALDAERAK | |
|---|---|
| <p>1. ALEGIA</p> <p>2. HOTS</p> <p>3. HAU DA</p> <p>4. ERRAN NAHI BAITA 2222 - erran nahi baita 455 - esan nahi baita 411 - erran nahi da 172 - esan nahi da 47</p> <p>5. HURRENEZ HURREN [zehaztaileak]</p> <p>6. BESTELA ESANDA 1032 - bestela esanda 738 - bestela esan 213 - bestela esateko 43 - bestela esanik 20 - bestela esatearren 9 - bestela esanez 1</p> | <p>7. BESTE ERA BATERA ESANDA 462 - beste era batera esanda 388 - beste era batean esanda 40 - beste era batera esateko 22 - beste era batera esanez 6 - beste era batera esanik 6</p> <p>8. BESTE MODU BATERA ESANDA 231 - beste modu batera esanda 166 - beste modu batean esanda 40 - beste modu batera esateko 61 - beste modu batean esateko 14</p> <p>9. BESTE HITZETAN 111 - beste hitzetan 93 - beste hitzez 18 - beste hitzez errana 1 - beste hitzez errana 0</p> <p>10. BESTE HITZ BATZUETAN ESANDA 148 - beste hitz batzuetan esanda 63 - beste hitz batzuekin esanda 19 - beste hitz batzuetan esateko 23 - beste hitz batzuekin esateko 13</p> |

3. taula: sarrera baten aldaerak multzoka antolatzeko aukera

2.9. Bilaketaren emaitzak (adibideak) azpicorpusez azpicorpus esportatzeko aukera

Lan-tresnak aukera ematen digu, azpicorpusen taulako gezia sakatuta (10.a. irudia), bilaketa baten emaitzak (hots, adibideak) azpicorpusez azpicorpus esportatzeko fitxategi binario batera (ikus 10.b. irudia).



10.a. irudia: bilaketa baten emaitzak esportatzeko aukera



10.b. irudia: bilaketa baten emaitzak esportatzeko aukera

3. Etorkizunari begira: lan-tresna nola hobetu eta gizarteratu

Gorago azaldu dugunez, berariaz diskurtso-markatzaileak —elementu lexiko aldagaitz mota bat— aztertzeke helburuarekin sortu genuen lan-tresna hau, baina gerora konturatu gara erabilgarria izan litekeela beste hizkuntza-unitate lexiko aldagaitz batzuk ikertzeko: a) lokuzioak (*agi denez, aho batez, argi eta garbi, argi kontra...*); b) kolokazioak (*bidea galdu, bizia galdu, hostoa galdu...*); c) hizketa-formula eta paremiak (*egun on dizuela; Jainkoak daki; Jainkoak gorde gaitzala...*).

Orain edozein erabiltzailek ikus ditzake interneten EUDIMA ikerketa-proiektuko lan-taldeak erauzi dituen sarreren adibideak eta datuak; baita datuak esportatu ere. Baina kanpoko erabiltzaileak ezin ditu sarrera berriak erauzi edota zaharrak ezabatu. Dena dela, etorkizunean lan-tresna gizarteratzea eta bere funtzio guztiekin ikertzaileen zerbitzura jartzea da gure asmoa.

Beraz, etorkizunari begira gure asmoa da lan-tresna hobetzea eta gainerako ikertzaileei eskaintzea UPV/EHUko [Euskara Institutua](http://www.ehu.es/etc/)ren web-atarian. Alde batetik, corpus askozaz ere zabalago bati —200 milioitik gora testu-hitz dituen *Egungo Testuen Corpora (ETC)*¹⁸ corpusari— aplikatuko diogu lan-tresna; beste alde batetik, *ETC* lematizazioa eskainiko du eta horrek bidea erraztuko du kolokazioak hobeto aztertu ahal izateko.

¹⁸ <http://www.ehu.es/etc/>

Laburbilduz, egungo euskarari erabiltzen diren unitate fraseologiko-diskurtsiboak (diskurtso-markatzaileak, lokuzioak, kolokazioak, hizketa-formula eta paremiak) ikertzeko lan-tresna egoki bat jarri nahi genuke ikertzaileen zerbitzura etorkizunean.

Erreferentzia bibliografikoak

- ALBERDI, X. (2011a): "Diskurtso-markatzaile berri bat: *hurrenez hurren* birformulatzailea", *Anuario del Seminario de Filología Vasca Julio de Urquijo* (ASJU), XLV (45), 1, 301-325.
- ALBERDI, X. (2011b) [argitaratzeko]: "Erran nahi baita birformulatzaile esplikatiboa gaurko euskarari", *Anuario del Seminario de Filología Vasca Julio de Urquijo* (ASJU), XLV (45), 2.
- ALBERDI, X. (2013a) [argitaratzeko]: "Diskurtso-markatzaile berri batzuk: *beste hitz batzuetan* (*esanda/esateko*), *beste hitzetan* (*esanda/esateko*)", in Gomez, Ricardo & Ezeizabarrena, Maria Jose (arg.): *Eridenen du zeraz kontenta. Sailkideen omenaldia Henrike Knörr irakasleari (1947-2008)*, Bilbo: UPV/EHUren Argitalpen Zerbitzua.
- ALBERDI, X.; EZEIZA, J. (2013) [argitaratzeko]: "Birformulazioa eta birformulatzaile esplikatiboak hizkera juridikoan", *Euskera*.
- BRIZ, A.; PONS, S.; PORTOLÉS, J. (koord.) (2008): *Diccionario de partículas discursivas del español*. Internet: <http://www.dpde.es> [2013-02-11]
- EZEIZA, J.; ALBERDI, X. (2013) [argitaratzeko]: Los marcadores discursivos como rasgo distintivo de los textos del ámbito de las Ciencias Jurídicas en lengua vasca: aproximación estadística con proyección jurilingüística", *Signos*.
- FUENTES, C. (2009): *Diccionario de conectores y operadores del español*, Madril: Arco/Libros.
- LOUREDA, O. eta ACÍN, E. (koord.) (2010): *Los estudios sobre marcadores del discurso en español, hoy*, Madril: Arco/Libros.
- MARTÍ, M. (2011): "Los conectores discursivos (entre los otros marcadores discursivos y los otros conectores)", in CESTERO, A. M^a eta MARTÍ, M. (koord.), *Lingüística en la red, monográfico: I Jornadas de lengua y comunicación: marcadores discursivos*, A Coruña: Universidad de A Coruña.
- MARTÍN ZORRAQUINO, M^a A. (2010): "Los marcadores del discurso y su morfología", in LOUREDA, O. eta ACÍN, E. (koord.), *Los estudios sobre marcadores del discurso en español, hoy*, Madril: Arco/Libros, 93-181.
- MARTÍN ZORRAQUINO, M^a A.; PORTOLÉS, J. (1999): "Los marcadores del discurso", in BOSQUE, I. eta DEMONTE, V. (argit.), *Gramática descriptiva de la lengua española*, Madril: Espasa. Vol. III, 4051-4213.