

Kausazko koherentzia-erlazioen azterketa automatikoa euskarazko laburpen zientifikoetan*

MIKEL IRUSKIETA¹, MARIA JESUS ARANZABE², ARANTZA DIAZ DE ILARRAZA³
ETA MIKEL LERSUNDI²

¹ Hizkuntza eta Literaturaren Didaktika Saila, ² Euskal Hizkuntza eta Komunikazioa Saila,

³ Lengoia eta Sistema Informatikoak Saila

Ixa Taldea (UPV/EHU)

(Toward a computational approach of causal coherence relations in scientific abstract texts)

DOI: 10.1387/gogoa.15628

Abstract

Detecting automatically the cause relations of a text may be useful in question answering tasks and event information extraction. The aim of this paper is to study how to detect coherence relations of the cause subgroup (CAUSE, RESULT and PURPOSE). To achieve this aim we have used the Rhetorical Structure Theory (RST) and some automatic linguistic information from different tools developed by IXA Group. We have used a corpus of 60 scientific abstracts, the Basque RST Treebank (Iruskietta et al., 2013), of different domains: science, medicine and terminology. A linguist has annotated all the signals of that corpus and described the most important problems in such task. To report the reliability of this annotator, two linguists have annotated the signals of the cause subgroup and all the annotations were compared and evaluated. After that, a superannotator has harmonized all the signals of those cause relations. Finally, we show the most important signals for such relations.

Keywords: coherence, connexion, cause, signals, RST.

* Lan hau burutu dugu proiektu hauen laguntzei esker: i) NewsReader: ICT Call 8 FP7-ICT-2011-8-316404 (Europako Batzordea). ii) Ber2tek: IE12-333 (Eusko Jaurlaritza).

1. Sarrera

Azken urteotan hizkuntza aztertze eta kontsultatzeko erak asko aldatu dira Hizkuntzaren Prozesamenduak izan duen garapenari esker; hau da, konputazioan oinarriturik corpusak eta hiztegiak erabiltzeko eta informazioa biltzeko era asko aldatu da.

Euskararen kasuan ere lan handia egin da, besteak beste, informazio morfosintaktikoa automatikoki aztertze tresnetan.¹ Esate baterako, Maltixa analizatzaile sintaktikoa (Bengoetxea eta Gojenola, 2007) eta Eustagger analizatzaile morfologikoa (Artola et al., 2005) baliatuta, (1) adibidearen analisi automatikoa (ikusi 1 Taula) lortuko dugu:

- (1) Gabonetan karroza-desfile ikusgarria egiten dute.

1. Taula

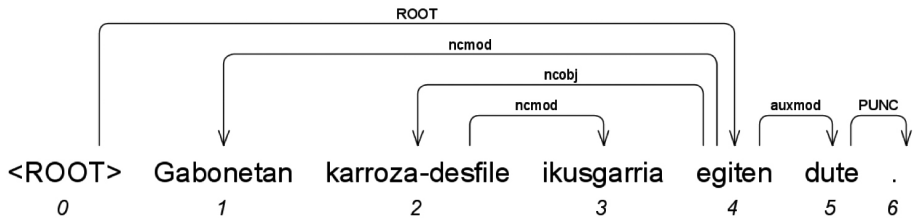
Maltixaren irteera

Syntax analysis of: Gabonetan karroza-desfile ikusgarria egiten dute. is:							
Index	Word	Head	Relation	Lemma	Features	Category	Subcategory
1	Gabonetan	4	ncmod	Gabonak	KAS:INE NUM:P	IZE	IZE_IZB
2	karroza-desfile	4	ncobj	karroza-desfile	—	IZE	IZE_ARR
3	ikusgarria	2	ncmod	ikusgarri	KAS:ABS NUM:S	ADJ	ADJ_ARR
4	egiten	0	ROOT	egin	ASP:EZBU	ADI	ADI_SIN
5	dute	4	auxmod	*edun	—	ADL	ADL
6	.	5	PUKC	.	—	PUNT	PUNT_PUNT

Analisi automatiko hori ulertze aldera, 1 Taulan informazioa honela adierazi da: *Index* zutabean, (1) adibideko esaldiaren hitz kopurua modu ordenatuan; *Word* zutabean, hitz bakoitza zein den; *Head* zutabean, zenbakiz adierazten da hitzaren burua zein den; *Word* zutabeko hitza esaldiko zein hitzaren mende dagoen: adibidez, esaldiko 2. hitza (*karroza-desfile*) esaldiko 4. hitzaren (*egiten*) mende dago; *Relation* zutabean, beti bi hitzen artean gauzatzen den dependentzia-erlazioa zein den adierazi da dependentzia-etiketak baliatuta: adibidez, *karroza-desfile* hitza *egiten* aditzaren objektua dela adierazi da *ncobj* (*non-clausal object*) dependentzia-etiketaren bidez; *Lemma* zutabean, esaldiko hitz forma bakoitzaren lema; *Features* zutabean, hitz bakoitzaren

¹ <http://ixa.si.ehu.es/Ixa>.

ezaugarri morfologikoak: izenetan, kasu-marka eta numero singularra edo plurala (KAS:INE/NUM:P) eta aditzetan, aspektua (ASP:EZBUrutua); *Category* zutabea, hitz bakoitzaren kategoria (IZE_{na}, ADJektiboa, ADItza, AditzLaguntzailea, PUNTuzazioa) eta *Subcategory* zutabea, hitz bakoitzaren azpikategoria (IZE_IZB: izen berezia, IZE_ARR: izen arrunta...). Analisi hau bera irudikatu da dependentsia-zuhaitzaren bitartez 1 Irudian.



1. Irudia

Dependentsien irudikapen grafikoa, 1 Taularena

Behin mota horretako analisi automatikoak lortuta, testua ulertzea helburu duten sistemak gai dira beste urrats bat egiteko. SENEKO programak (Lopez-Gazpio eta Marichalar Anglada, 2013; Aldabe, 2011), esaterako, esaldi batean hainbat galdera eta erantzun sor ditzake automatikoki:

- (2) Noiz egiten dute karroza-desfile ikusgarria?
- (3) Zer egingo dute Gabonetan?

Galdera horiek egitea sinplea dirudien arren, hainbat informazio hartu behar da kontuan (Aldabe et al., 2013): *i*) inesiboarekin non edo noiz galdetu behar den, *ii*) sintagma zer den, *iii*) galderetan sintagmen ordena zein den, *iv*) izena +/- bizidun den, *v*) izen berezia edo arrunta den. Informazio hori bera izan behar da galdera horiek erantzuteko.

Galdera-erantzunak automatikoki sortzeko sistema horiek oso baliagarriak dira irakaskuntzaren zenbait ataza lantzeko: kasu-markak, osagaien hurrenkera, eta abar. Baina esaldia luzea denean, esaldi barruko informazio linguistiko horrekin galderak automatikoki sortzea eta erantzutea zailago gertatzen da eta are zailagoa testu osoa interpretatu nahi dugunean esaldien arteko erlazioak gauzatu. Esaldiko osagaiei buruzko galderak egin eta erantzun daitezke, baina nola ziurta daiteke zuzenak direla egin diren galderak eta eman diren erantzunak? Nola egiazta daiteke irakurleak testua ulertu duela?

Horri guztiari irtenbidea emateko, beharrezkoa da testu bateko erlaziozko diskurtso-egiturari dagokion informazioa izatea; hau da, testuaren egitura hierarkikoa zein den jakinda gai izango ginatke testu-zatien arteko erlazioak gauzatzeko. Horretarako, beharrezkoa da honako ataza hauek automatikoki

egitea: *i*) testuak oinarritzko diskurtso-unitateetan (aurrerantzean, EDU) zatitzea, *ii*) testuko ideia nagusia edo gaia (unitate zentrala) detektatzea eta *iii*) ideia nagusiari lotutako koherentzia-erlazio esanguratsuak detektatzea.

Lehen bi ataza horiek aztertuak izan direnez Iruskieta eta Zapirain (2015) eta Iruskieta et al. (2015a) lanetan, artikulua honetan hirugarren ataza gauzatzeko oinarriak finkatuko ditugu; zehatzago esanda, kausa taldeko koherentzia-erlazioak automatikoki detektatzeko patroiak deskribatuko ditugu euskarazko laburpen zientifikoez osatutako corpora (Iruskieta et al., 2013) oinarri hartuta.

Koherentzia-erlazioen artean kausazkoen azterketari lehenetsia eman diogu, Girjuk (2003) erakusten duelako galde-erantzun sistemak hobetu daitezkeela aditzen argumentuen arteko kausazko erlazioak eraziz; horretarako, *WordNet*eko informazioan oinarritu da bera. Halaber, erlazio-egitura erabiltzeko edota esaldia gainditzeko ikuspegi horretatik, kausazko erlazio erretorikoak beste erlazioak baino hobeto ulertzen direla diote Knott eta Dalek (1994). Horrez gain, Osinagaren (2001) arabera, diskurtso mota bakoitzak baditu erlazio erretoriko bereziak eta, beraz, arrazoizkoa dirudi aztertzea ikuspegi teorikotik antzekotasuna duten erlazioak, batetik, eta diskurtso motari lotutakoak, bestetik. Horiek horrela, zientziako testuetan kausa taldeko erlazioak eta bere seinaleak aztertzea erabaki dugu, horiek direlako gure corpora ezaugarritzen duten erlazio erretorikoak. Azterketa horretan kontuan izango dugu batetik, KAUSA, ONDORIO eta HELBURU erlazioak direla kausazkotzat hartu dituztenak Benamara eta Taboada (2015) *Rhetorical Structure Theory* (RST) eta *Segmented Discourse Representation Theory* (SDRT) (Asher eta Lascarides, 2003) hurbilpenetako koherentzia-erlazioak konparatu ondoren eta, bestetik, ikuspegi analitikotik, ez dela erraza sailkatzea Euskara Institutuaren (2015) arabera hiru erlazio horiek (kausa, ondorioa eta helburua) adierazteko perpausak, semantikoki oso hurbil daudelako.

Egin dugun beste lan guztia hurrengo atal hauetan azalduko dugu, lehendabizi, zein den arloko egoera eta zein den kausazko erlazioak detektatzeko erabiliko dugun teoria (2. atala); ondoren, marko metodologikoa zein den (3. atala), eta horren ostean, etiketatzaileen adostasuna eta kausa taldeko seinaleak zeintzuk diren aipatuko ditugu (4. atala). Bukatzeko, ondorio nagusiak eta etorkizuneko lana zein izango den aipatuko dugu (5. atala).

2. Egitura Erretorikoaren Teoria (RST)

Hizkuntzaren Prozesamenduan testuak deskribatzeko gidalerroak definitu dira teoria desberdin hauetan oinarrituta: *i*) *Segmented Discourse Representation Theory* (SDRT)² (Asher eta Lascarides, 2003), *ii*) *Discourse-Lexical-*

² Aipagarria da SDRT euskarara ekartzeko saiakera (Gómez, 1996).

ized *Tree Adjoining Grammar* (D-LTAG) (Webber et al., 2003) eta *iii) Rhetorical Structure Theory* (RST) (Mann eta Thompson, 1987).

Lan hau aipaturiko azken teoria horretan oinarritu da, RSTn,³ hain zuzen ere. Testuaren egitura edo koherentzia deskribatzen duen teoria da Egitura Erretorikoaren Teoria (RST). Marcuren (2000) arabera, RST da hizkuntzaren sorkuntza automatikoan dabiltzanen artean teoriarik erabiliena. RSTren bidez koherenteak diren testu mota gehientsuenak deskriba daitezke, hizkuntza batean edo bestean izan (Mann eta Thompson, 1987). Mann and Thompsonen (1988) arabera testu-unitatea edo diskurtso-unitatea osatzen dute perpausek (*clause*), subjektu edo objektu funtzioa ez duten perpau adjuntuek. Har dezagun RSTko adibide ezagun hau, (4a) adibidea, diskurtso-unitatea zeri esaten dioten ikusteko.

- (4) a. [The next music day is scheduled for July 21 (Saturday), noon-midnight.]₁ [I'll post more details later,]₂ [but this is a good time to reserve the place on your calendar.]₃⁴ (Mann and Thompson, 1988, 252)

Esate baterako, (4a) adibidea Mann eta Thompsonen (1988) hiru segmentutan zatitzen dute eta segmentu bakoitzak proposizio bat hartzen du. Segmentu bakoitza giltzekin zedarritu eta besteetatik bereizteko segmentu-amaieran azpiindizeak jarri ditugu. Segmentazioa 2 Irudian errepresentaturiko zuhaitz-egituran ere ikus dezakegu, testu-segmentuak fisikikoki bereizita baitaude.

Segmentatu ondoren, unitateek elkarren artean dituzten koherentziazko erlazioak (*coherence relation*) ezartzen dira, 2 Irudian erakusten den moduan. Testu-zatien arteko koherentziazko erlazioei erlazio erretoriko (*rhetorical relation*) esaten zaie. Erlazio erretorikoekin soilik deskribatzen da koherentzia RSTn. RSTn ez da bestelako kontzepturik behar, ezta diskurtso-markatzailena ere⁵ Oatesen (1999) arabera, irakurleak erlazio implizituen⁶ kasuan diskurtso-markatzailerik gabe ere interpretatu behar duelako testu zatien arteko erlazio erretorikoa.

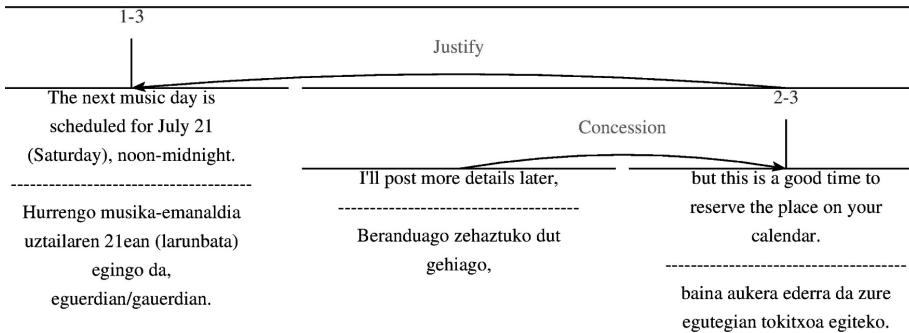
³ RSTri buruzko informazio gehiago (Iruskieta et al., 2015b) lanean aurki daiteke. Euskaraz RSTko web-orrian: <http://www.sfu.cwa/rst/07basque/index.html>.

⁴ Hau da adibide ezagun horren itzulpena: «Hurrengo musika-emanaldia uztailaren 21ean (larunbata) egingo da, eguerdian/gauerdian. Beranduago zehaztuko dut gehiago, baina aukera ederra da zure egutegian tokitxo egiteko.»

⁵ Diskurtso-markatzailerekin erlazio erretorikoak nabarmendu eta seinalatu egiten ditu idazleak eta irakurleak erlazio-egitura erakutsi.

⁶ Erlazio implizituei bibliografian modu askotan esaten zaie: *implicit, unsignal, unmarked* edo *underspecific*.

Unitate bat beste unitate bati erlazio erretorikoaz lotzeko unitateak koherentetzat hartu behar ditu irakurleak; izan ere, unitateen artean koherentziarik ez badago, ez da unitateen arteko koherentzia-erlaziorik egongo edo, zehazkiago esanda, irakurleak ez du ulertuko zein den unitateen arteko erlazio erretorikoa. Horregatik esaten zaie RSTn koherentzia-erlazio erlazio erretorikoei (Pardo, 2005).



2. Irudia

RSTko (4a) adibidearen errepresentazioa

Erlazio erretorikoz osaturiko egitura hierarkikoa edo zuhaitz erretorikoa behetik gora azaltzen badugu, ikus dezakegu 2 Irudian 2 unitatea lotzen zaiola 3 unitateari, CONCESSION erlazioaz. Izan ere, idazleak badaki 2 unitatean *musika-emanaldi* hori irakurlearen asmoetan edo planetan sartzeko beharrezko informazioa (lekua eta ordua) falta dela eta informazio hori eman gabe plana egiteko eskaera kontraesankorra dela; bada, irakurleak ezin du jakin *musika-emanaldira* joateko bere egutegian tarterik duen. Beraz, bi egoeren arteko kontrakotasun hori idazleak onartuta, irakurleari adierazten dio 3 unitatean *musika-emanaldia* aipatutako egunean izango dela eta agendan ekitaldi horri tokia egin diezaiola. RSTn erlazio erretorikoak errekurtsiboak direnez, testu koherenteetako unitate guztiak zuhaitz-egituran txerta daitezke; beraz, erlazio bat beste erlazio baten unitatea izango da, edo, bestela esanda, erlazioak lotzen dituen unitate-multzoa (*span*) beste erlazio-egitura baten unitate izan daiteke. 2-3 *spana* JUSTIFY erlazioaz lotzen zaio 1 unitateari. Hau da, 2-3 unitate-multzoan (*spanean*) irakurleak uste du idazleak 1 unitatea esateko arrazoirik baduela, nahiz eta *musika-emanaldiaren* lekua ez zehaztu.

Testu koherenteetan alboko segmentuak (*adjacent segments*) loturik egon behar dira eta ez da *non-sequitur* edo hutsunerik egongo; beraz, RSTn testuko unitate guztiak lotu beharko ditu etiketatzaileak. Lotura horiek zein diren adierazteko, erlazio erretorikoez lotzen dira alboko segmentuak. Horrela, RSTrekin irudikatutako testu koherentearen erlazio-egitura aztertzean, unitate guztiak lo-

tuta egongo dira eta zuhaitz-egitura osatuko dute. Testua edo testu-zatiren bat inkoherentea denean, ordea, unitate batean esaten denak ez du bestean esaten denarekin erlaziorik izango (*non-sequitur*) eta, hortaz, etiketatzaileak ezin izango die erlazio erretorikorik ezarri koherentziarik gabeko unitate horiei.

Unitate bat beste bati lotzen zaiola esan dugunez, erlazio-egituran unitate bakoitzak duen garrantzia zein den azaldu beharrean gaude. Idazleak testuaren mezua edota intentzioa komunikatzeko, unitate batzuk gako izango dira ideia nagusia azaltzeko, eta beste unitate batzuk, bigarren mailakoak, ideia nagusi horren xehetasunak azaltzeko erabiliko dira. Hortaz, diskurtso-unitate batzuk beste batzuk baino garrantzitsuagoak dira. Horiei nukleo-unitate (*nuclear unit*, N) esaten zaie eta besteei satellite-unitate (*satellite unit*, S). Unitate bat nukleo ala satellite den esateari nukleartasuna (*nuclearity*) ebaztea deritzo. Nukleartasun ezberdineko erlazio errekurtsiboekin sortzen dira RSTko zuhaitz-egitura hierarkikoak.

Nukleo bakarreko erlazioetan satellitea da nukleoari erlazio erretorikoz lotzen zaiona. Horiei, nukleo-satelitez (N-S)⁷ osaturikoei *erlazio hipotaktiko* deritze eta nukleo-nukleoz (N-N)⁸ osaturiko erlazioei *erlazio parataktiko* edo nukleoaniztunak (*multinuclear*).

- i) Erlazio nukleoaniztunak (N-N) nukleoz bakarrik osatuta daudenez, erlazio nukleobakarretan ez bezala, nukleo-unitate guztiek garrantzi bera dute eta, beraz, hierarkiarik ez dago lotzen dituzten unitateen artean. Hauek dira sailkapen hedatuko erlazio nukleoaniztunak: LISTA, DISJUNTZIOA, BATERATZEA, BIRFORMULAZIO MULTINUKLEARRA, SEKUENTZIA, KONTRASTEA eta KONJUNTZIOA.
- ii) Erlazio nukleobakarrek (N-S) maila ezberdinetako proposizioak lotzen dituzte. Nukleo-unitateak predikazioaren edo komunikatu nahi denaren zatirik garrantzitsuenak dira; satellite-unitateek, berriz, gaia (zeina nukleo-unitatean adierazten den) hobeto edo zehatzago ulertzen laguntzen dute. Erlazio hauek lotzen dituzten segmentuei dagokienez, bai esaldi barnean, bai esaldien edo paragrafoen barnean egon daitezke eta bi motatakoak dira (Mann eta Thompson, 1988):
 - a) Edukizko (E) erlazioak (*subject matter*) dira; unitate bat besteari lotzen zaion efektuaren izaera semantikoa bada, hau da, idazlearen intentzioa da erlazio-unitateen artean erlazio bera dagoela jakinaraztea irakurlari: ELABORAZIOA, METODOA, ZIRKUNSTANTZIA, ARAZO-SO-

⁷ Lan honetan N-S (Nukleo-Satelite) erabiliko dugu erlazio hipotaktiko edo nukleobakarra adierazteko, bai Nukleo-Satelite, bai Satellite-Nukleo segida adierazteko.

⁸ Erlazio parataktikoei edo nukleoaniztunei erreferentzia egiteko N-N (Nukleo-Nukleo) erabiliko dugu, bai bi nukleoko erlazioak izendatzeko, bai nukleo gehiago dituzten erlazioak izendatzeko.

LUZIOA, BALDINTZA, AUKERA, ALDERANTZIZKO BALDINTZA, EZ-BALDINTZATZAILEA, INTERPRETAZIOA, EBALUAZIOA, KAUSA, ONDORIOA eta HELBURUA.

- b) Aurkezpenekoek (A) (*presentational*), ordea, izaera erretorikoa dute; beraz, unitateen arteko loturak irakurlearengan efektu jakina eragiteko asmoa du: PRESTATZEA, TESTUINGURUA, AHALBIDERATZEA, MOTIBAZIOA, EBIDENTZIA, JUSTIFIKAZIOA, ANTITESIA, KONTZESIOA, BIRFORMULAZIOA eta LABURPENA.

RSTn erlazioak nola definitzen diren erakusteko (4a) adibideko erlazio biak —KONTZESIOA eta JUSTIFIKAZIOA— nola definitzen dituzten erakutsiko dugu 2 Taulan.

2. Taula

KONTZESIOA eta JUSTIFIKAZIOA erlazioen definizioak

	Arauk Sn eta Nn	Arauk S-Nn	Efektua
KONTZESIOA	Nn: idazleak Nrekiko aldeko iritzia du: Sn: idazleak ez du erakusten S onartzen ez duenik	Idazlearentzat N eta S onargarriak izan badaitezke ere, bien artean aurkakotasunak egon daitezkeela onartzen du: irakurleak aurkakotasunezko egoera hori onartzean, Nrekiko aldeko iritzia handitzen du	N eta S onargarriak izan arren, aurkakotasunak daude bi egoerotan. Aurkakotasun horrek irakurlearengan Nrekiko aldeko iritzia handitzea dakar
JUSTIFIKAZIOA	Ez dago baldintzarik	Irakurleak S ulertzean, idazleak N aurkezteko egokitasuna areagotzen da	Irakurleak idazleari N aurkezteko egokitasuna onartzen dio

KONTZESIOA erlazioaren definizioari dagokionez, batetik, unitate bat KONTZESIOA erlazioarekin lotzeko, nukleo-unitatean bete behar den baldintza da idazleak aldeko jarrera izatea unitatean bertan aipatzen den egoerarekiko. Satelite-unitateko baldintza da idazleak ez duela errefusatzen satelitean agertzen den egoera. Bestetik, bi unitateak erlazonatzerakoan bete behar den baldintza da idazlearentzat bi unitateetako egoerak onargarriak badira ere, badagoela bi egoeren arteko aurkakotasuna. Beste alde batetik, idazleak uste du bi egoera horien aurkakotasun ez erabatekoa erakutsiz indartuko duela irakurleak nukleo-unitatean agertzen den egoerarekiko duen aldeko iritzia.

JUSTIFIKAZIOA erlazioaren definizioari dagokionez, batetik, nukleo-unitatean eta satelite-unitatean ez dago baldintzarik. Bestetik, nukleoaren eta satelitea-

ren arteko erlazioan baldintza da nukleo-unitatean aurkezturiko egoeraren azalpena ematen duela idazleak satelite-unitatean eta azalpen horretan nukleo-unitateko egoeraren motiboak edo zergatiak aipatzen direla. Beste batek, horrek irakurlearengan duen efektua da irakurleak idazleari nukleo-unitateko egoera aurkezteko zilegitasuna onartzen diola.

Beraz, RSTrekin (4a) adibidea eta bere erlaziozko diskurtso-egitura (2 Iru-dia) kontzeptu horiekin guztiekin deskribatzen da.

Erlazio erretorikoen detekzioari dagokionez, erlazio-funtzioa (konexioa) duten baliabide linguistikoak (diskurtso-markatzaileak, zentzu zabalean ulertuta) izan dira Hizkuntzaren Prozesamenduan gehien aztertu direnak. Izan ere, diskurtso-markatzaileek orokorrean koherentziatzeko erlazio bakar bat seinalatzen ez badute ere (van Dijk, 1998), zenbait erlazio zedarrizten dituzte. Zenbait autorek (Knott eta Sanders, 1998; Risselada eta Spooren, 1998) diskurtso-markatzaileak erlazio erretorikoak seinalatzen dituzten hizkuntzazko ebidentziatzat hartzen dituzte, irakurlea gidatzeko edota erlazio-egituraren berri emateko. Ikuspegi hori oso erabilgarria izan da (Georg et al., 2009; Taboada eta Mann, 2006; Pardo eta Seno, 2005) erlazioak aztertzeko edo automatikoki detektatzeko.

Hala izanik ere, Taboadak (2006) dio diskurtso-markatzaileez seinalaturiko erlazioak gutxiago direla seinalatu gabeko erlazioak baino. Lan horretan erlazioen % 43,48 soilik daude seinalaturik. Beraz, diskurtso-markatzaileekin ez dira erlazio guztiak seinalatzen (Iruskieta et al., 2009).⁹

Diskurtso-markatzaileen anbiguotasunari nahiz erlazio implizituen arazoei aurre egiteko eta erlazio erretoriko gehiagoren ebidentzia linguistikoak izateko, zenbait autorek (Das et al., 2015; da Cunha et al., 2012; Pardo and Nunes, 2004) bestelako elementuak (erlazio funtziorik ez duten esapideak, aditzak, izenak, besteak beste) ere aztertu dituzte erlazio erretorikoen seinale¹⁰ gisa; nahiz eta seinale horiek diskurtso-markatzaileak baino anbiguoagoak izan.

Adibidez, Das et al.ek (2015) diotenez *RST Discourse Treebank*eko test-corpuseko (38 fitxategi) erlazio-seinaleen (edozein multzotako seinaleak) maiztasuna % 81,81 da (1127 erlazio seinalaturik, 1304 erlazioetan). Beraz, garrantzitsua da Hizkuntzaren Prozesamenduan erlazioak detektatzeko metodo bat ala bestea aukeratzea, batek (diskurtso-markatzaileenak) erlazioen % 43,48 baino ez baititu deskribatzen eta besteak (erlazio-seinaleenak) erlazioen % 81,81.

⁹ Erlazio erretoriko guztiak diskurtso-markatzaileekin ez seinalatzeko zuzenbide-egituraren goiko mailetan (errekurtsibitate gehiago dagoen mailetan) sarriago gertatzen da beheko mailan baino (Longacre, 1985).

¹⁰ van Dijkek (1998) 'seinale' terminoa erabili beharrean 'missing link' edo 'konektatzaile eliditua' terminoa erabili du bi proposizioen arteko lotura eragiten duen edozein elementu denotatzeko.

Erlazio erretorikoak detektatzeko helburuarekin, RSTn, erlazio erretoriko-koen seinaleak corpusetan aztertu dituzten lanak hauexek dira:

- Portugeserako Pardo eta Nunesek (2004) 100 zientzia-artikuluren sar-
rerrak (4100 erlazio erretoriko) aztertu dituzte eta seinale-zerrendak
eman dituzte.
- Gaztelaniarako da Cunhak (2013) 84 zientziako testu (1275 erlazio
erretoriko) aztertu ditu eta horietako batzuk desanbigutzeko errege-
lak egin ditu. Gaztelaniako corpus hori bera erabili du Duquek (2014)
KAUSA eta ONDORIOA erlazio-seinaleak aztertzeko.
- Ingeleserako Das et al.ek (2015) egindako lan sakonean, egunkarie-
tako 385 artikuluko (21404 erlazio erretoriko) corpora aztertu dute.
Lan horretan, kategoria eta maila guztietako seinaleak proposatu di-
tuzte.

Das et al.en (2015) sailkapenean oinarrituta etiketatzaile batek (E_1 ek) cor-
puseko erlazio guztien seinaleak etiketatu ditu Rhetorical DataBase (RhetDB)
tresna (Pardo, 2005) erabiliz. Etiketatzaile horrek etiketaturiko adibideak era-
biliko ditugu, 3 Taulako sailkapena azaltzeko.

3. Taula

Erlazio erretoriko-koen seinale motak euskararako

Seinale motak	Seinaleak
Juntagailuak eta lokailuak	eta, baina, hala ere, horrenbestez, izan ere
Morfologikoak	-t(z)eko, -lako, -gatik, bait-
Lexikoak	ez, lortu, helburu, emaitza, asmoa
Entitateak	[izen bereziak]
Semantikoak	<i>Antonimia</i> : handitu (...) txikitu, ugari (...) gutxi, ir- tenbide (...) arazo, <i>hiponimia</i> : patologia (...) gaixotasun, legamia (...) onddo
Sintaktikoak	[galde-perpaua] (...) [erantzun-perpaua]
Grafiko-nermerikoak	1. (...) 2., a (...) b)

Kausa taldeko erlazioak detektatzeari dagokionez, Khoo et al.ek (1998)
kausa-efektu erlazioa detektatzeko patroï linguistikoak definitu dituzte eta
aztertu dute bi etiketatzailek erlazioen % 68an bat egiten dutela ingelesez
idatzitako kazetaritza testuetan. Patroïak egiteko ezaugarri morfologikoak,
kausazko aditz-zerrendak, diskurtso-markatzaileak, izenordainak eta beste-

lako informazioa erabili dute. Esaterako, honela definitu dute satellite-nukleo (SN edo 2-1) ordena duen KAUSA erlazioa detektatzeko patroia (5).

- (5) [2] (AND THIS) & AUX &ADV because of [1].
The car failed to brake in time and this was (&ADV) because of it
was raining heavily.

Orobat, da Cunha et al.ek (2012) esaldi barruko koherentzia-erlazioak detektatzeko, hiru motatako informazioan oinarrituriko erregelak egin dituzte: *i)* diskurtso-markatzaileetan (*ya que* modukoak) oinarrituriko 155 erregela, *ii)* izenetan eta aditzetan (*metodologia, usar* modukoak) oinarrituriko 77 erregela, eta *iii)* aditz-egituretan (*para + INFINITIVO* modukoak) oinarrituriko 41 erregela. Adibidez, KAUSA erlazioa detektatzeko diskurtso-markatzaileetan oinarrituriko erregela bat da honako hau (6):

- (6) SATELLITE-NUCLEUS (S-N) RULES: -*ya que* («since») => Satellite of Cause of the next EDU

Euskaraz kausazko erlazioak detektatzeko ez da egin Hizkuntzaren Prozesamenduan erabil litekeen deskribapenik. Edonola ere, badaude zenbait gramatika-lan kausa taldeko erlazioak seinatzeko baliabideak deskribatzen dituztenak (Euskaltzaindia, 1999, 1990). Lan horietatik eta IXA ikerketataldean osatutako Euskararen Datu Base Lexikaetik (EDBL) (Aldezabal et al., 2001) erauzi ditugu 4 Taulako diskurtso-markatzaileak.¹¹

4. Taula

Kausa taldeko diskurtso-markatzaileak
(morfologia mailakoak, juntagailuak eta lokailuak)

Kausazkoak	Ondoriozkoak	Helburuzkoak
-(e)lako	alabaina	(*erakuslea*)... ezen/non... bai-(ADI)
(ADI)-eta	bada	(*erakuslea*)... ezen/non... -n
bait-(ADI)	ezen (bada)	orduan
-t(z)eagatik	haatik	bada
-t(z)earren	hargatik	beraz
-nez (gero)	horratik	hainbestez
-la eta	horregatik	hala
-la bide	izan ere	halatan
-la kausa	zergatik	honenbestez
-la medio	zeren (eta)	horrenbestez
		hortaz
		ondorioz
		-t(z)eko
		-t(z)era
		-t(z)eko asmotan
		-t(z)ekotan
		-t(z)eagatik
		-t(z)earren
		(subjuntiboa)

¹¹ Taulan *erakuslea* jartzen duen tokian: *hain, non, honen, horren* eta *horrenbeste* jar daitezke.

Horrez gain, Oyharçabalek (2002) eta Ormazabalek (2007) euskarazko aditz kausatiboen berri ematen dute. Lan horietan *i*) egoera fisikoaren al-daketak adierazten dituztenak dira: *hustu, zikindu, berotu, hautsi, apurtu, urratu...*, *ii*) psiko-kausatiboak: *kezkatu, alaitu, beldurtu...*, eta *iii*) toki-aldaketak adierazten dituztenak: *hurbildu, bideratu, aurreratu, atzeratu...* Bestalde, Nedjalkov eta Silnickij (1973) lanean, guri aditz kausatibo ezberdinen artean ezberdintzen lagunduko ligukeen sailkapena egin dute: *i*) sinpleak (*simple causatives*), *cause, lead to, generate, allow...*, *ii*) ondoriozkoak (*resultative causatives*), *kill, melt, dry...*, *iii*) instrumentalak (*instrumental causatives*): *hang, punch, clean*, besteak beste.

Euskaraz egin diren lan gehienetan, diskurtso-markatzaileen balioak eta esaldian duten kokalekua aztertzen da; gure lanean seinaleak esaldian duen kokagunea zein den aipatzeaz gain, perpaus-elkarketan zein diskurtso-unitatetan dagoen (lehenengoan edo bigarreanean) eta unitate horrek diskurtsoan duen nukleartasun-funtzioa zein den (satelitea edo nukleoa) desberdinduko dugu.

Aipatutako euskal gramatiketan eta RSTko lanetan oinarrituko gara zientziako laburpen-testuez osatutako Euskal RST *Treebankean* dauden erlazio-seinaleen patro linguistikoak deskribatzeko.

3. Euskal RST *Treebankeko* seinaleak

3.1. Corpusaren osaera eta etiketatze-lana

Euskal RST *Treebanka* hiru domeinuko testuekin osatu dugun corpora da:

- Medikuntzako testuak *Gaceta Médica de Bilbao* aldizkariko artikuluetako laburpenak dira.
- Terminologiako testuak HAEE-IVAPek eta UZEIk antolatuta, Donostian 1997. urtean egin zen Nazioarteko Terminologia Biltzarreko aktetako hitzaldi eta komunikazioen laburpenak dira.
- Zientziako testuak Euskal Herriko Unibertsitateko (UPV/EHU) Zientzia eta Teknologia Fakultateak 2008ko maiatzean antolaturiko Zientzia eta Teknologia Fakultateko 1. Ikerkuntza Jardunaldiak izeneko komunikazioetako laburpen batzuekin osatu da.

Corpusaren ezaugarriak 5 Taulan aurkezten ditugu:

Corpus horretan erlazio-egitura etiketatuta dago; beraz, testuak segmentatuta eta beraien arteko koherentzia-erlazioak etiketatuta daude. Iruskieta-*ren* (2014) arabera corpus horretan koherentzia-erlazioetan dagoen fidagarritasun maila hau da: *i*) nukleartasunean % 81,73 (f-neurria) eta *ii*) erlazio

erretorikoetan % 61,47 (f-neurria). Corpusaren fidagarritasuna neurtu ondoren, etiketatzaileetako batek ezadostasunak ebazteko lanak egin ditu, fidagarritasun gehiago eman nahian.

Erlazio-seinaleen etiketatzaileei dagokienez, hiru etiketatzaileak (E_1 , E_2 eta E_3) hizkuntzalariak dira eta hizkuntzako beste maila batzuk (morfosintaxia, sintaxia eta semantika) zein erlazio-egitura etiketatzen esperientzia izan dute. Lan honetan erlazio-seinaleak etiketatu dituzte.

5. Taula

Aztergai dugun euskarazko corpusaren deskribapena

Azipicorpusa		Testuak	Esaldiak	Hitzak	EDUak	Erlazioak
MEDIKUNTZA	GMB	20	198	3.010	283	253
TERMINOLOGIA	TERM	20	253	5.664	584	504
ZIENTZIA	ZTF	20	352	6.892	603	535
Guztira		60	803	15.566	1.470	1.292

Zuhaitz erretorikoetako seinaleak etiketatzeko, *Rhetorical Data-Base* (RhetDB) tresna erabili dugu. Tresna horri esker, etiketatzea (baita kontsultak ere) erlazioz erlazio egin daiteke. Hiru etiketatzaileek erlazio baten (KAUSARENAK esaterako) seinale guztiak etiketatu dituzte lehenbizi eta gerorako utzi dituzte falta zaizkien bi erlazioak (ONDORIOA eta HELBURUA).

3.2. Erlazio-seinaleak etiketatzeko gidalerroak

Das et al.en (2015) lanari jarraituz, etiketatzaile batek (E_1) diskurtso-markatzaileak¹² eta esanguratsuak iruditu zaizkion beste elementu batzuk hartu ditu seinaleztat, Euskal RST *Treebankean*. Horren ostean, etiketatzaileak izan dituen arazoen berri eman eta seinaleen adibide batzuk prestatu ditu beste bi etiketatzaileek (E_2 eta E_3) kausa taldeko erlazio-seinaleak etiketatzeko.

Adibidez, KONTRASTEAREN erlazioa etiketatzean nukleoan edo satelitean ez partikula etiketatzen dugu, erlazio-seinale horrek (7) adibidean KONTRASTEAREN ez-

¹² Etiketatzailearen batentzat diskurtso-markatzaileak ez dira beti esanguratsuak izan aztertu dituen erlazioetan. Edonola ere, diskurtso-markatzaileak etiketatzea erabaki dugu, beraien funtzioa erlazio-egitura esplizitu egitea delako; horren ondorioetako bat da seinale horiek erlazio erretoriko bat baino gehiago seinale dezaketela.

berdintzen duelako KONJUNTZIOA erlaziotik;¹³ baina beste erlazioetan agertzen diren ez partikula batzuk ez ditugu etiketatu, (8) adibidean, esaterako, ez delako seinale hori (ez hitza) KONJUNTZIOA erlaziorako adierazgarria; bai, ordea, *eta* juntagailua.¹⁴

- (7) [Gure parametroen arabera definitutako fenotipo «basala» korrelazioan jarri zen nabarmenki p53 mutatuaren agerpenarekin (p = 0.0001) 3. gradu nuklearra (p < 0.0001), eta % 60 edo hortik gorako Ki67 agerpen ehunekoarekin; bestalde, korrelazioa oso txikia izan zen 3. gradu histologikoarekin (p = 0.045).]_{1N} KONTRASTE[A] [Ganglioen inbasioarekin (p = 0.51) ez zen korrelazio gradurik hauteman.]_{2N} KONTRASTE[A] GMB0702
- (8) [Horien artetik, 369 NOS kartzinoma duktal iragazkorrak izan ziren,]_{1N} KONJUNTZIOA [*eta* horiexek baino ez dira aztertu.]_{2N} KONJUNTZIOA GMB0702

Euskal RST *Treebankean* ez partikula 70 aldiz agertzen da eta horietako 4tan etiketatu da KONTRASTEAREN seinaleztat eta 1ean KONTZESIOAREN seinaleztat. Beraz, ez partikularen agerpen gehienak ez ditu etiketatzailerak seinaleztat hartu.

Ondorengo azpiataletan hizkuntza-maila ezberdinetan aurkitu ditugun erlazio-seinaleen berri emango dugu. Aurretik erakutsi dugunez, zenbait kasutan seinale konposatuekin seinalatzen dira koherentziazko erlazioak. Erlazio-seinaleak etiketatzeko ataza zeinen konplexua den erakusteko, beste mailako seinaleen berri oin-ohar bidez emango dugu, baita bestelako interpretazioen berri ere.

3.2.1. Lokailuak eta juntagailuak erlazioen seinale

Diskurtso-funtzioa (EDUen arteko erlazio-funtzioa) duten juntagailuak eta lokailuak soilik etiketatu ditugu. Beraz, EDUen edota *spanen* arteko juntagailuak edota lokailuak baino ez ditugu etiketatu.¹⁵ Gure corpusean *hala ere* diskurtso-markatzaileak KONTZESIOA erlazioa bakarrik seinalatzen du, adibidez, (9) adibidean.¹⁶ Erlazio bakarra seinalatzen duen beste diskurtso-mar-

¹³ Korrelazio eta gradu hitzak ere seinaleztat har daitezke, nahiz eta etiketatzailerak batek ere ez dituen etiketatu. Orobat, puntuazioa ere ez da seinaleztat hartu.

¹⁴ Nabarmendu behar da aurreko (7) adibidean ez dugula *eta* juntagailua etiketatu, bi segmentuen arteko diskurtso-funtziorik ez duelako. Juntagailuak edota lokailuak seinaleztat hartzeko, diskurtso-funtzioa izan behar dute aztertzen ari garen koherentzia-erlazioan.

¹⁵ (9) adibideko laugarren hitza (*eta* juntagailua), besteak beste, ez dugu seinaleztat hartuko, EDU₁ean hitzak koordinatzeko funtzio sintaktikoa duelako.

¹⁶ Adibide honetan ikus daiteke antonimo erlazioa egon litekeela *sistema* eta *gizaki aditu* hitzen artean, nahiz eta etiketatzailerak ez duten halakorik adierazi.

katzaile bat da *aldiz* lokailua. Lokailu horrek KONTRASTEA erlazioa seinalatzen du¹⁷ (10) adibidean ikus daitekeenez.¹⁸

- (9) [Termino espezializatuak artxibatzearekin eta aurkeztearekin lotuta dauden biltegitratze- eta berreskuratze-lan batzuk arindu dituzte gaur egun terminologia kudeatzeko **sistemek**.]₁₅ KONTZESIOA [**Hala ere, gizaki adituek** egiten dituzte bilketak, azterketak eta baliozkotzeak.]_{2N} TERM30
- (10) [Gaixo **bat** arnas azidosian aurkitzen zen;]_{1N} KONTRASTEA [**bostek, aldiz**, arnas alkalosi kronika zeukaten.]_{2N} KONTRASTEA GMB0001

Beste zenbait diskurtso-markatzailek erlazio erretoriko bat baino gehiago seinalatzen dute. Esaterako, *izan ere* diskurtso-markatzaileak (11) adibidean KAUSA eta (12) adibidean JUSTIFIKAZIOA seinalatzen du.¹⁹ Gauza bera gertatzen da gaztelaniaz, *ya que* diskurtso-markatzailearekin CAUSA eta JUSTIFICACIÓN erlazioak seinala baititzake (da Cunha, 2013).

- (11) [Aurreko hamarkadetan, serbierako zientzia-arloko ikertzaile askok joera bat nabaritu dute eta horren berri eman dute: ingeleseko unitate lexikalen maileguak eta unitate-egitura luzeagoen maileguak hartzen dira zientzia-erregistro zehatz baterako, itzulpenak edo kalkoak egin ordez.]_{1N} [**Izan ere**, iritzi ezberdinetako zientzialari serbiarrek adostasuna lortu dute eta aurreko hamarkadetan ingelesari eman diote zientzia-komunikaziorako hizkuntza bakarraren estatusa.]₂₅ KAUSA TERM18
- (12) [Helburuak ere bi jomuga zituen: alde batetik, lanbidearen errealitate-rako hurbilpen delako hori —kasurik onenean ere gelan egindako simulazio-saio batzuk baino urrunago ez doana— muturreraino eramatea, nonahi eta noiznahi aipatzen baitzaigu bai itzulpengintza eta interpretaritzari buruzko nazioarteko foroetan, bai halako diziplinen inguruko ikasgaietako ikasketa-planetan eta curriculumetan ere; eta beste alde batetik, gaur egun batetik bestera hain usu darabilgun diziplinartekotasuna gauzatzea behingoz zalantzarik tokirik utzi gabe.]_{1N} [**Izan ere**, Itzulpengintza eta Interpretaritzako ikasketetan ez dugu diziplinartekotasun hori baztertu baina, barneratu ere ez dugu erabat egin (behin betiko indartzeko arrazoi garbia, beharbada).]₂₅ JUSTIFIKAZIOA TERM48

¹⁷ Beste adibide honetan, *bat* eta *bost* zenbaki ezberdinen arteko erlazioa seinaleztat har daiteke, nahiz eta kasu honetan ere etiketatzaileek ez duten halakorik adierazi. Horrez gain, bi esaldiek erakusten dute egitura sintaktiko bera: ‘zenbakia + gaixotasuna + aditza’; beste seinaleak ez baleude, egitura sintaktiko hori egokia litzateke KONJUNTZIOA seinalatzeko.

¹⁸ Adibideetan ez dira EDU guztiak zenbakitu, irakurterraztasunaren mesedetan.

¹⁹ Diskurtso-markatzaileetan egon daitekeen anbiguotasuna desanbigua daiteke beste kategoria batzuetako seinaleak kontuan hartuz. Esaterako, (12) adibideko nukleoko eta sateliteko aditz nagusietan pertsona ezberdinak izatea baliagarri izan daiteke kasu honetan; izan ere, nukleoko aditz nagusia 3. pertsonan dago eta satelitean dauden aditzak 1. pertsona pluralean (-gu morfema) daude. Beraz, aurkezpeneko koherentzia-erlazioaren seinale izan daiteke hori. RSTn aurkezpeneko erlazioetan kausalitate kutsua izan dezakeen erlazioa da JUSTIFIKAZIOA.

3.2.2. Morfemak erlazioen seinale

Diskurtso-markatzaileez gain, ezaguna da zenbait menderagailu erabil daitezkeela esaldi barruko proposizio adjuntuen arteko loturak egiteko. Esate baterako, (13) adibidean *-t(z)eko* menderagailua erabili da satelite-unitatea HELBURUA erlazioz nukleo-unitateari lotzeko.²⁰ (14) adibidean nukleo-unitatean *arrazoi* izena, batetik, eta *-gatik* eta *-lako* menderagailuak, bestetik, KAUSA erlazioa seinalatzeko.

- (13) [TERMCAT Terminologia Zentroak Consell Supervisor izeneko organo iraunkor bat dauka]_{1N} [beste hizkuntza batzuetatik, batez ere ingelesetik datozen neologismo terminologikoetatik katalanera **babesteko**]_{2S} HELBURUA TERM21
- (14) [Eragozpen gehiago topatuko dugu ondoko hiru arrazoiengatik]_{1N} [bateratze-prozesua bukatzeke izateagatik, egindako ikerketak murriztak direlako eta hizkuntza eranskaria izateagatik]_{2S} KAUSA TERM31

Baina morfemekin ere gertatzen da juntagailuekin eta lokailuekin gertatzen den gauza bera, ez dela beti morfema bera erabiltzen erlazio bakarra seinalatzeko. Bada, *bait-* menderagailuak erlazio bat baino gehiago seinalatzen du: (15) adibidean JUSTIFIKAZIOA²¹ eta (16) adibidean KAUSA,²² adibidez.

- (15) [Helburuak ere bi jomuga zituen: alde batetik, lanbidearen errealitate-rako hurbilpen delako hori —kasurik onenean ere gelan egindako simulazio-saio batzuk baino urrunago ez doana— muturreraino eramatea]_{1N} [nonahi eta noiznahi **aipatzen** baitzaigu bai itzulpengintza eta interpretaritzari buruzko nazioarteko foroetan, bai halako diziplinen inguruko ikasgaietako ikasketa-planetan eta curriculumetan ere]_{2S} JUSTIFIKAZIOA TERM48
- (16) [Euskararen kasuan **beharrizan hori areagotu** egin da]_{1N} [esparru horretan gure hizkuntzaren **egoera ez** baita primerakoa, ezta hurrik eman ere]_{2S} KAUSA TERM25

3.2.3. Lexikoa erlazioen seinale

Aipatu dugun bezala, Pardo eta Nunesek (2004) erakutsi dute diskurtso-markatzaileez gain, zenbait aditz ere erlazio erretorikoen seinaleztat har dai-

²⁰ *Babestu* aditza ere seinaltzat har daiteke; izan ere, aditz horrek '*purpose+*' eta '*cause+*' ezaugarri ontologikoak ditu WordNet-eko zenbait adieratan.

²¹ Adibide horretan sateliteko aditzean dagoen *-gu* morfema ere seinaleztat har daiteke, JUSTIFIKAZIOA erlazioaren seinaleztat. Bestalde, *aipatu* aditzak ere '*purpose+*' eta '*cause+*' ezaugarriak ditu WordNet-eko zenbait adieratan.

²² Beste adibide honetan, *egoera* eskas batek eragin dezake *beharrizan*etako bat.

tezkeela, baina era berean adierazi zuten seinale horiek ez dutela balio testu-genero guztietarako. Adibidez, (17) adibideko *lor*tu aditza testu zientifikoetan ONDORIOA erlazioaren seinaleztat hartu dugu.²³

Aditzez gain, izenek ere seinalatzen dute erlaziorik; (18) adibidean esaterako, *helburu* izenak HELBURUA erlazioa seinalatzen du.²⁴

- (17) [Bestetik, trantsizio-metalen fosfato, fosfito eta artseniatoak sintetizatu dira,]_{1N} [horrela, baldintza hidrotermal leunetan eta diaminen presentzian, material mikroporodun hibrido inorganiko-organiko lortu dira.]_{2s} ONDORIOA ZTF17
- (18) [Bukatzeke, komunikazio sistemetan zarataren eragina kontuan hartu beharreko gai garrantzitsua da.]_{1N} [Alor honetan, gure ikerkuntza taldearen helburu nagusiak bi dira. Alde batetik, anplifikadore eta nahasgailuen zarata-figuraren neurketen zehaztasuna hobetzea, eta beste alde batetik, zirkuitu eta sistemetan (anplifikadore, osziladore eta nahasgailuetan bereiziki) seinalearen presentzian anplitudezarata eta fase-zarata neurtu ahal izateko tekniken garapena.]_{2s} HELBURUA ZTF13

Bestalde, *emaitza* hitzak ONDORIOA seinalatzen du,²⁵ (19) adibidean bezala; nahiz eta gure corpusean INTERPRETAZIOA seinalatzen duen gehiagotan, beste seinale batzuek ere eragiten dutelako. Esaterako, (20) adibidean *emaitza* badugu ere, aditz nagusia (*zailtzen*) gailentzen da predikazioan eta hori hartu dugu INTERPRETAZIOAREN seinale nagusitzat.

- (19) [Zentzu honetan, gure ikerkuntza taldeak mikrouhin-zirkuituen dinamika ez-linealaren arloan eten-gabeko lana egin du azken urte haue-tan zehar.]_{1N} [Lan hauen emaitza izan da polo-zero identifikazioan oinarritutako egonkortasun analisirako tekniken garapena.]_{2s} ONDORIOA ZTF13
- (20) [Gainerakoan, prozesu adierazle egokiak daude, baina altan dagoen gaitxoaren ahalmen funtzionalaren erregistro urria antzematen da,]_{1N} [eta horrek osamen arretaren emaitzen alderaketa zailtzen du.]_{2s} INTERPRETAZIOA GMB0504

²³ Kasu honetan *lor*tu aditzak ez du WordNet-eko euskarazko bertsioan 'cause+', badute, ordea, *to get* aditzaren zenbait adierak. Horrez gain, *horrela* diskurtso-markatzaile funtzioa duen adberbioa ere hartu da seinaleztat.

²⁴ Bestalde, *helburu* izenaren ondoren jarritako *nagusi* adjektiboak idazleak dituen helburuez ari dela adieraz dezake, kasu askotan.

²⁵ Nabarmendu nahi dugu KAUSA eta ONDORIOA erlazioak nukleartasunak erabakita aukeratzen direnez, *emaitza* hitza anbigua dela eta KAUSA nahiz ONDORIOA seinala dezakeela. Gehienetan ONDORIOA erlazio seinalatzen du, horri buruzko predikazioaz ari baita (19) adibidean bezala.

3.2.4. Entitateak erlazioen seinale

(21) adibideko *Estibaliz Apiñaniz Fdez de Larrinoa* eta *Damian Martin Rodriguez* izenak LISTA erlazioaren seinaleztat har daitezke.²⁶ Baita (22) adibideko izen bereziak ere: '*Ln1-xAxM1-yM'yO3*' izenak eta '*A2BB'O6*' izenak.²⁷

- (21) [*Estibaliz Apiñaniz Fdez de Larrinoa* «Theoretical and experimental study of the relationship between structural and magnetic properties in Fe-Al alloys» Julio de 2003.]_{1N} LISTA [*Damian Martin Rodriguez* «Estudio sobre la influencia del orden intermetálico en las propiedades magnéticas de las aleaciones Fe-Al ricas en Fe» Febrero de 2007.]_{2N} LISTA ZTF10
- (22) [*Ln1-xAxM1-yM'yO3* motako oxido mistoek (Ln = lantanidoa; A = katioi dibalentea; M = Mn, Fe; M' = trantsizio-metalak) perovskita motako egitura erakusten dute, baita magnetorresistentziaren balio altuak eta propietate magnetiko bereziak (exchange bias) ere.]_{1N} LISTA [*A2BB'O6* motako perovskita bikoitzak ere ikertu dira (A = lantanidoak edo lurralkalinoak; B eta B' = trantsizio-metalak).]_{2N} LISTA ZTF17

3.2.5. Erlazio lexiko-semantikoak erlazioen seinale

(23) adibidean EDU₁₃an *eraso* (*eraso terminologiko gogorak jasan*) eta EDU_{2N}an *irtenbide* (*irtenbide terminologikoekin leundu*) antonimoak dira; izan ere, satelitean *erasoa* arazotzat hartzen du. Antonimia horrek ARAZO-SOLUZIOA seinalatzen du.²⁸ Bestelako ikuspuntu batetik, erlazio semantiko horiek ko-referentziak sortzen dituzte.²⁹

(24) adibidean argi eta garbi ikusten da *handitu* eta *txikiagotu* antonimoak direla,³⁰ kasu honetan antonimo horiek ANTITESIA erlazioaren seinale dira, idazleak ezin baitu bi proposizioen artean egia-balio bera komunikatu; bataren ala bestearen alde egin behar du.³¹

²⁶ Egia esateko, erlazio horretan entitateez gain, badaude beste seinale batzuk. Horietako bat egitura sintaktikoaren errepikapena da: IZENA-IZENBURUA-DATA egitura bi EDUetan errepikatzen denez, hori ere LISTA erlazioaren seinaleztat har daiteke. Horrelakorik, ordea, ez du etiketatzailerik batek ere etiketatu.

²⁷ Adibide horretan *ere* lokailua ere kontuan hartzekoa da.

²⁸ Adibide horretan *gogor* eta *leun* hitzen artean ere antonimia erlazioa dago, baina horrelako seinaleak aurkaritzako erlazioan etiketatu izan dira eta ez dute ARAZO-SOLUZIOA erlazioa seinalatzen.

²⁹ Horrela, erreferentziatzko diskurtso-egituraren eta erlaziozko diskurtso-egituraren arteko harreman argia ikus daiteke.

³⁰ Euskal WordNet-en esaterako *handitu*₁ adieradun aditza *txikiagotu*₈ adieradun aditzaren antonimoa da. EusWordNet hemen kontsulta daiteke: <http://ixa2.si.ehu.es/cgi-bin/mcrt/public/wei.consult.perl>.

³¹ Aipatu behar da adibide horretako menderagailuak ere *-tu beharrean* aurkakotasun-zentzua indartzen duela.

(25) adibidean bi antonimia erlazio daude, batetik, *aurrizki* eta *atzizki* hitzen artekoa eta, bestetik, *ugari* eta *gutxi* terminoen artekoa.³²

- (23) [Teknologia berrien mundura sartzen ari den neurrian, eraso terminologiko **gogor**ak jasaten ari da galiziera, eta arlo hori ere gaztelerarekiko menpekotasunaren beste adar bat izan daiteke.]_{1S} ARAZO-SOLUZIOA [Galizierari emandako irtenbide terminologikoekin **leundu** egin nahi da arazo larri hori, batetik, gaztelerakadak besterik gabe onartzea saihestuz, eta, bestetik, gaztelerak moldatutako terminoak bere corpus terminologikora bilduz.]_{2N} TERM 17
- (24) [asetasun magnetizazioa [(**handitu beharrean**, ferromagnetiko arruntetan gertatzen den bezala)]_{1S} ANTITESIA txikiagotu egiten da.]_{2N} ZTF10
- (25) [a] euskarak atzizki ugari baditu ere,]_{1N} KONTRASTEA [oso aurrizki gutxi dauka,]_{2N} KONTRASTEA TERM50

(26) adibidean *minbizi* eta *tumore* hitzen arteko erlazio semantikoa sinonimia da eta LISTA erlazioa seinalatzen dute bi segmentuek nukleartasun bera dutelako eta zerrenda bat osatzen dutelako.³³

- (26) [Gure biztanleen artean, T1 bularreko minbizietako gongoil zentinelak aztertzean, oro har, lau gaixotik gongoil positibo bat hauteman ohi da. Kasuen erdian edo gehiagoetan, gongoil hori basapean egongo den positibo bakarra izango da.]_{1N} LISTA [T1a tumoreak eta edozein tamainatako kartzinoma tubularrak dira gongoil inbasioa izateko probabilitate txikiena dutenak.]_{2N} LISTA GMB0703

(27) adibidean, esaterako, hiponimia erlazio semantikoa dago eta ELABORAZIOA seinalatzen du *legamia* aurretik aipatu den *onddo* mota bat delako.³⁴

- (27) [Candida albicans animalia homeotermoetarako ohizko oportunistak den onddo patogenoa da.]_{1N} [Legamia honek odolez sakabanatzeko gaitasuna dauka arazo larriak sortuz immunokonprometituak dauden pazienteengan.]_{2S} ELABORAZIOA ZTF20

(28) adibidean *onddo* izenaren errepikapena ELABORAZIOA erlazioaren seinaleztat har daiteke.³⁵

³² Horrez gain, *ba... ere* menderagailuak aurkaritza adierazten duen beste erlazio bat seinalatzen du: kasu honetan KONTRASTEa da, bi EDUen arteko garrantzi maila berdina delako.

³³ *Kartzinoma* minbizi mota bat denez, hiponimia erlazioa dago nukleoan dagoen *minbizi* seinalearen eta satelitean dagoen *kartzinoma* seinalearen artean. *TI* terminoaren errepikapena ere izan daiteke seinalea, elementu baten ezaugarrien zerrendaketa batean.

³⁴ Horrez gain, hiponimia hori progresio linealarekin (*legamia* da nukleoko galdegaia eta *onddo* sateliteko mintzagaia) garatzen da, izaki horren gaitasuna zein den esateko.

³⁵ Pentsa liteke adibide horretan dagoen erlazioa bestelakoa dela, testu osoa kontuan hartu gabe; izan ere, bigarren perpausoko modalizazioa irakurriz (*aho batez onartzea*) pentsa liteke nu-

- (28) [Onddoak bizidun heterotroak dira eta ekosistemetako sare konplexuetan parte-hartzen dute, bereziki lurarretan.]_{1N} [Bertan, onddoek betetzen dituzten funtzioen garrantzia aho batez onartzen da gaur egun.]_{2S}
ELABORAZIOA ZTF15

3.2.6. *Sintaxia erlazioen seinale*

(29) adibidean honakoek seinalatzen dute ARAZO-SOLUZIOA: galdera-perpaseko (sateliteak) 'zeintzuk galdetzailea + aditza + subjektua + galdera marka' egitura sintaktikoak eta ondorengo erantzunean (nukleoan) dagoen egiturak 'galdegaia + aditza + mintzagai hanpatua'. Orobat, (30) adibidean DU_{1N}ak³⁶ baiezeko perpausaren ordena sintaktikoa du 'aditza + laguntzailea' eta DU_{2N}ak ezezeko perpausaren ordena sintaktikoa 'ez + laguntzailea + aditza'; egitura horrek KONTRASTEAREN erlazioa seinalatzen du.³⁷

- (29) [(...) Zeintzuk dira lexikoaren sorkuntzan gailentzen ari diren sistematik?] _{1S} ARAZO-SOLUZIOA [Ezaugarri berdin eta komun bat daukate hizkuntza guztiek: terminoak ingelesez sortzen dira, eta mailegu moduan sartzen dira beste hizkuntzetan.] _{2N} TERM38
- (30) [% 95 inguruko estaldura lortzeko doitasuna % 50-era jaitsi ohi da.] _{1N} KONTRASTEAREN [eta doitasuna % 85 ingurukoa izan dadin estaldura % 35era ere ez da iristen.] _{2N} KONTRASTEAREN TERM31

3.2.7. *Baliabide grafiko-nererikoak erlazioen seinale*

(31) adibideko 1. eta 2. zenbakiak edota (32) adibideko a) eta b) hizkiak LISTA³⁸ erlazioaren seinaleztat har daitezke, baita eta juntagailua ere.

- (31) [1. Gizentasun fenotipoen zehaztapena Euskadiko populazioan eta ijito etniako familietan.] _{1N} LISTA [2. Gizentasunarekin asoziazioa edo ligamendua erakusten duten geneak identifikatzea ijito etniako familietan

kleoa dela eta, beraz, hor beharko lukeen erlazioa TESTUINGURUA. Testu osoa kontuan hartzen bada, ordea, konturatu gara nukleoa lehenengo perpaus koordinatua dela; etiketatzaileek, behintzat, hala etiketatuta dute.

³⁶ RSTn diskurtso-unitatea (DU) konposatua izan daiteke; beraz, EDU bat baino gehigo duen kasuetan erabiltzen da.

³⁷ Besteak beste, *estaldura* eta *doitasuna* termino kontrajarriak ere KONTRASTEAREN seinaleztat har daitezke.

³⁸ Adibide honetan KONTRASTEAREN erlazioa dagoela pentsa dezake norbaitek, bi segmentu horiek soilik irakurriz; baina bi segmentu horien aurreko esaldiak determinatu egiten du bi esaldi horien arteko LISTA erlazioa. Horrela, testuko goiko mailako erlazioek beheagoko erlazioak (makroegiturak mikroegitura) nola determinatzen duen azter daiteke, besteak beste. Hauxe da adibide horren aurretik dagoen esaldia: «Edozein azalpen teorikok argitu behar dituen bi desberdintasunak, zehazkiago, hauexek dira:» (ikus hemen testu horren zuhaitz-egitura: http://ixa2.si.ehu.es/diskurtsoa/diskurtsoa_jpg/TERMS0-GS.jpg).

eta Euskadiko populazioan, asoziazio eta ligamendu analisiak erabiliz.]_{2N}
LISTA ZTF8

- (32) [a] euskarak atzizki ugari baditu ere, oso aurrizki gutxi dauka,]_{1N} LISTA
[eta b] desoreka hori ez dago inola ere hizkuntza erromanikoetan.]_{2N}
LISTA TERM50

3.3. Ebaluazioa

Corpuseko erlazio-seinale guztiak etiketatu dituenaren (E_1 ek) fidagarritasuna neurtzeko, beste bi etiketatzailerik (E_2 k eta E_3 k) etiketatu dituzten kausa taldeko (KAUSA, ONDORIOA eta HELBURUA) 233 erlazioen seinaleak (corpuseko erlazio guztien % 17,72) baliatu dira.

Erlazio-seinaleak etiketatzean, Duquek (2014) ez bezala, etiketatzailerik erlazio-seinale bat baino gehiago jar dezaketenez, bi etiketatzailerik bat datozela diogu soilik etiketatzailerik etiketatu dituzten seinale guztiak bat datozenean; beraz, nahikoa da adostasunik ez dagoela esateko, seinale ezberdin bakar bat. Etiketatzailerik arteko adostasunik 6 Taulan aurkezten ditugu.³⁹

6. Taula

Kausa taldeko erlazio-seinaleetan E_1 en fidagarritasuna, E_2 ren eta E_3 ren arteko adostasunarekin konparatua

Etiketatzailerik	KAUSARENAK		ONDORIOARENAK		HELBURUARENAK	
E_1 - E_2	40/56	71,43	40/67	59,70	99/110	90,00
E_1 - E_3	38/56	67,86	34/67	50,75	89/110	80,91
E_1 - E_2 - E_3	33/56	58,93	25/67	37,31	83/110	75,45
E_2 - E_3	41/56	73,21	25/67	37,31	86/110	78,18

Fidagarritasun handiena HELBURUAREN seinaleetan dago eta txikiena ONDORIOAREN seinaleetan. Duquek (2014) ere ondorio bera ateratzen du, gaztelaniako RST corpuseko KAUSA eta ONDORIOA erlazioen seinaleak etiketatzean, bertan ere KAUSA erlazioko seinaleetan adostasun handiagoa dago eta txikiagoa ONDORIOAREN seinaleetan. Gure etiketatzelanean, horren arrazoiak zein den jakiteko, hiru etiketatzailerik arteko adostasuna zein seinale-multzotan gertatu den erakusten dugu 7 Taulan.⁴⁰

³⁹ Helburua E_1 en fidagarritasunaren berri ematea denez, E_2 - E_3 etiketatzailerik arteko adostasunaren berri bigarren mailan ematen dugu.

⁴⁰ 'Seinalerik-ez' jarri dugu E_1 ek seinalerik jarri ez duenean eta adostasuntzat hartu dugu beste bi etiketatzailerik ere ez dutenean seinalerik jarri.

7. Taula

Kausa taldeko erlazio-seinaletan E₁en, E₂ren eta E₃ren arteko adostasuna multzoka

	KAUSA	ONDORIOA	HELBURUA
Seinalerik-ez	3tik 0	8tik 1	1etik 0
Juntagailuak eta Lokailuak	6tik 2	15etik 5	0tik 0
Morfemak	30etik 28	6tik 5	77tik 71
Lexikoa	8tik 3	19tik 7	6tik 7
Konposatuak	10etik 0	19tik 7	16tik 5
Guztira	57tik 33	67tik 25	110etik 83

7 Taulan nabarmen agertzen da ONDORIOA erlazio-seinaleei buruzko adostasuna lortzea zailagoa izan dela. Hori gertatzearen arrazoia izan daiteke ONDORIOAN maiztasun handia duen seinalerik ez topatu izana. Erlazio horretan maiztasun handiena *eta* juntagailuak du (4 aldiz etiketatua) eta ez da erlazio horren seinale bereizgarria. KAUSAN adostasuna handiagoa da; bada, KAUSAN 16tik 9tan etorri dira bat bi etiketazaille (% 56,25) eta ONDORIOAN 53tik 17tan (% 32,07). Nabarmenezkoa da juntagailuetan eta lokailuetan ez dela adostasunik egon; izan ere, etiketazailleetako batek ez ditu erlazio horien bereizgarritzat hartu seinale horiek. Lexikoan ere ezadostasunak egon dira etiketazaileren artean, batek jarrera kontserbakorra hartu du eta seinale gisa oso lexiko mugatua etiketatu du.

7 Taularen arabera etiketatu diren seinale-multzo guztietan ezadostasunak badaude ere, alde batetik, morfologia mailako seinaleetan dago adostasun gehien % 92,36 (113tik 104) eta, beste aldetik, adostasun gutxien, honako hiru multzo hauetan dago: *i*) seinale konposatuetan % 26,67 (45etik 12) *ii*) juntagailuetan eta lokailuetan % 33,33 (21etik 7)⁴¹ eta *iii*) lexikoan % 39,53 (43tik 17).

Emaitzei dagokienez, 6 Taularen arabera ONDORIOAREN seinaleetan egon da hiru etiketazaileren artean adostasun txikiena (% 37,31); HELBURUAREN seinaleetan, ostera, adostasun handiena (% 75,45). Erlazio seinaleetat diskurtso-marketzaileak baino fenomeno gehiago hartzen ditugunean, ataza konplikatu egiten da. Zailtasun hori gehitutakoan, subjektibotasuna neurtzea are garrantzitsuagoa da. Horren ebidentzia da, 7 Taularen arabera, gure la-

⁴¹ Juntagailu eta lokailuetan egon den ezadostasuna seinaleak etiketatzeko bi estrategia ezberdinei dagokie: *a*) Etiketazaille batek *eta* juntagailua beti etiketatu du erlazio-seinaletzat, baita kausa taldeko erlazioaren adierazgarri izan ez denean ere. *b*) Beste etiketazaille batek *eta* juntagailua ez du kausa taldeko erlazioen seinaleletzat hartu.

nean hiru etiketazaileen arteko subjektibotasuna txikiagoa dela diskurtso-markatzaileei (juntagailuei, lokailuei eta zenbait morfemari) dagokienez: % 82,84ko adostasuna (134tik 111). Bestelako seinaleetan (konposatuak barne) adostasuna askoz txikiagoa da: % 32,95ekoa (88tik 29).

Etiketazaileen arteko adostasunaren deskribapen kualitatiboari dagokionez, hiru etiketazaileen adostasuna seinale sinpleetan oinarritzen da. Seinale sinple esaten diogu multzo bakarreko seinaleari, esaterako, (33) adibidean, etiketazaile guztiak bat etorri dira seinale sinple bakarra aukeratzeko; hirurek *-lako* baino ez baitute etiketatu.

- (33) [Bere gorputzaren ohiko funtzionamendua narriatu *delako*]_{1s} KAUSA
[dago ospitalean.]_{2N} GMB0501

Ezadostasunaren arrazoi handienetakoa seinale konposatuekin gertatu da, esate baterako (34) adibidean hiru etiketazaileak seinaleetako batean ados egonda ere, ez daude ados zenbat elementuk osatzen duten seinalea. E₁ek *eragin-da* etiketatu du, E₂k *-da* eta E₃k *loturak eta... trukaketak eragin-da*. Hiru etiketazaileak bat etorri dira *-da* morfemak KAUSA erlazioa seinalatzen duela, E₁ek eta E₃k morfema horri lotutako *eragin* aditza gehitu diote (adiera askotan ‘*cause+*’ du WordNet-en) eta horrez gain, E₃k aditz horren argumentu koordinatua *loturak eta trukaketak* etiketatu ditu.⁴²

- (34) [Alde batetik, gero eta indartsuagoa da nazioarteko harmonizazioa lortu beharra,]_{1N} [ekonomian, politikan eta kultura eta gizarte gaietan etenik gabe sortzen ari diren loturak eta elkarren arteko trukaketak *eraginda*]_{2s}
KAUSA TERM 19

Kausa taldeko erlazio-seinaleak etiketatu ondoren, epaileak desadostasunak ebatzi ditu eta erlazio-seinaleen urre-patroizko corpusa eraiki du. Urre-patroizko corpus hori da ondoren deskribatuko duguna.

4. Kausa taldeko erlazioen patroiak

Atal honetan kausa taldeko seinaleekin osaturiko urre-patroizko corpuseko seinale esanguratsuenak zeintzuk diren zehaztuko ditugu.

Etiketazaileen arteko desadostasunak deskribatu ondoren, epaileak seinaleak harmonizatzekeo prozesua egin du, irizpide hauek jarraituz:

⁴² Intuizioan oinarritu beharrean, epaileak erlazio-egituraz eta informazio morfosintaxikoaz aberasturiko corpusa beharko luke, zalantza dagoenean maiztasunean oinarrituriko irizpideren bat jarraitzeko.

- i) Diskurtso-markatzaileak beti etiketatzea, nahiz eta erlazio jakin baten seinale ez izan.⁴³
- ii) Bi etiketatzaile bat etorririko seinaleak onartzea.⁴⁴

Harmonizazio lan horren ondorioz, sistema automatikoa sortzeko urrepatroitzat hartuko diren kausa taldeko seinaleak zehaztu dira. Etiketatzeko erabili dugun RhetDB tresnak berak estatistikak ateratzen ditu eta horri esker hurrengo taula hauetan aurkezten ditugu erlazioak eta bere seinaleak:⁴⁵ *i)* kausazkoak 8 Taulan, *ii)* ondoriozkoak 10 Taulan eta *iii)* helburuzkoak 12 Taulan.

4.1. Kausazko seinaleak eta anbiguotasuna

KAUSA erlazioaren seinalerik nabarmenenak hauek dira: SN ordenan⁴⁶ satelitean *-nez* seinalea etiketatu da 14tan, NS ordenan satelitean *bait-* 9tan, SN ordenan satelitean *-ela eta* 6tan, eta bi ordena posibleetan, bai nukleoan, bai satelitean *-gatik* 6tan. KAUSAzko seinale nabarmen gehiago erakusten ditugu 8 Taulan.

8. Taula

KAUSAREN PATROIEN LABURPENA

JNT eta LOK	Izenak	Aditzak	Morfologia
honegatik	arrazoi	ekarri	-nez (gero)
horregatik	emaitza	eragin	bait-
horretarako	efektu	etorri	-ela eta / bide / tarteko
horren ondorioz	eragile		-gatik
izan ere			-lako(an)
eta			-ta

Kausa taldeko seinaleen anbiguotasuna. Kausa taldeko erlazio-seinale garrantzitsuenen baliagarritasuna eta anbiguotasuna aztertu dugu, erlazio horiek detektatu nahi ditugulako.

⁴³ Zenbait diskurtso-markatzaile zenbait erlaziotan agertzen dira, baina ez beste batzuetan eta horrek balio dezake ikasketa automatikoarekin erlazioak bereizteko.

⁴⁴ Kasu gutxi batzuetan epaileak etiketatzaile bakarrak hautaturiko seinalea onartu izan du. Geroko lanetarako utziko da seinale horien egokitasuna neurtzea. Harmonizazio lan honen helburua ahalik eta seinale gehien etiketatzea izan da.

⁴⁵ Harmonizaturiko seinaleak hemen kontsulta daitezke: <http://ixa2.si.ehu.es/diskurtsoa/>.

⁴⁶ SN-rekin satelitearen eta nukleoaren ordena adierazten dugu, baita NS-rekin alderantzizko ordena.

KAUSA erlazioa maiztasun handiarekin seinalaturik dago: % 94,64 (56 erlaziotatik 53 daude seinalaturik). Lagin esanguratsua aztertu dugu 9 Taulan 47 seinale aztertzean: % 83,93 (56tik 47).

9. Taula

KAUSA seinalatzen duten seinaleen anbiguotasuna

Seinaleak	KAUSA	JUSTIFIKAZIOA	Besteak
ADI-nez	21	12 57,14	9
bait-ADI	19	9 47,37	4
ADI-ela eta	7	6 85,71	1
ADI-gatik	3	3 100,00	
izan ere	8	4 44,44	4 44,44
arrazoia	3	3 100,00	
Guztira	61	37	10 14

Lagin horretan, KAUSA gehienbat menderagailuekin seinalatzen da, adibidez, 'ADI-nez', 'bait-ADI', 'ADI-ela eta' eta 'ADI-gatik' baita *izan ere* lokailuarekin eta *arrazoi* izenarekin ere. Ikus dezakegu 9 Taulan zenbait seinalek, kasu honetan *izan ere* lokailuak eta *bait-* menderagailuak, KAUSA eta JUSTIFIKAZIOA erlazioak seinalatzen dituztela, (35) eta (36) adibideetan ikus daitekeen bezala.

- (35) [eta buru modura jokutzen dutenak ere, lehenengo gunean kokatzen dira,]_{1N} [osagarria eskuinaldean hartzen baitute (eslomar-se / desplomarse, desbular / deshojar).]_{2s} KAUSA TERM50
- (36) [Bestalde, Euskaltzaindiak hitz elkartuen bidea (1995eko urtarrilaren 27an onartutako araua) proposatzen du adjektibo erreferentzialak itzultzeko, baina arauan bertan esaten denez, «...ahal den guztian...»],_{1N} [*izan ere*, eguneroko praktikak frogatzen *baitigu* ez dela beti posible izaten formula horretara jotzea]._{2s} JUSTIFIKAZIOA TERM22

4.2. Ondoriozko seinaleak eta anbiguotasuna

Oso esanguratsua da ONDORIOAREN seinale ia denak NS ordenan daudela, seinale bakarra dago SN ordenan; baina seinale hori ez da ONDORIOAREN adierazgarria, *bestalde*, emendiozko lokailua delako. Bestalde, beste hauek dira seinalerik esanguratsuenak NS ordenako satelitean *i*) juntagailuak: *eta* 8tan, *ii*) izenak: *ondorio(z)* 7tan eta *emaitza* 7tan (nukleoan ere etiketatu da), *iii*) aditzak: *lortu* 7tan, eta *iv*) morfologian: *infititiboa gehi -z* 5etan. ONDORIOZKO seinale nabarmen gehiago erakusten ditugu 10 Taulan.

10. Taula

ONDORIOAREN PATROIEN LABURPENA

JNT eta LOK	Izenak	Aditzak		Morfologia
eta ondorioz beraz hau dela eta horrenbestez orduan horrela bada hartara	(-en) ondorio(z) emaitza(k) eragina korrelazio datu	lor্তু eragin erdietsi ekarri sortu erakutsi baieztatu aurkitu	esan nahi frogatu inplikatu adierazi jakin areagotu aurkitu	-tuz -ez gero bait- ba- -larik -lako

Ondoriozko seinaleen anbiguotasuna. Bestalde, ONDORIOAN seinale guxtiago daude: % 79,1 (67 erlaziotik 53 daude seinalatuta) eta aztertu dugun lagina ere txikiagoa da; bada, 11 Taulan 35 seinale aztertu ditugu: % 52,24 (67tik 35), ondorio seinaleek maiztasun txikia baitute, lehenago ere aipatu dugunez.

11. Taula

ONDORIOA SEINALATZEN DUTEN SEINALEEN ANBIGUOTASUNA

Seinaleak	ONDORIOA	KAUSA	INTERPRETAZIOA	Besteak
beraz 7	3 42,86	3 23,08	1 7,69 8 57,14 1 16,67	4
ondorioa 13	8 61,54			2
emaitza 14	6 42,86			
lor্তু 6	5 83,33			
Guztira 40	22	3	10	6

ONDORIOAN agertzen diren seinaleak KAUSA eta INTERPRETAZIOAN ere erabili izan dira 11 Taulan ikus dezakegunez. Horren erakusgarri da, (37) adibidean ez dagoela seinalerik ONDORIOA dela adierazteko (soilik *eta* juntagailu orokorra) eta (38) adibidean ONDORIOA eta (39) adibidean INTERPRETAZIOA seinalatzen duela *emaitza* seinale anbiguoak.

(37) [P. aeruginosa bidezko kolonizazio-infekzioak bronkio-gaixotasun kronikoak dituzten pazienteei eragiten die]_{1N} [eta haien eboluzioa korapilatzen du;]_{2S} ONDORIOA GMB0201

(38) [Zentzu honetan, gure ikerkuntza taldeak mikrouhin-zirkuituen dinamika ez-linealaren arloan etengabeko lana egin du azken urte hauetan

- zehar.]_{1N} [Lan hauen emaitza izan da polo-zero identifikazioan oinarritutako egonkortasun analisirako tekniken garapena.]_{2S} ONDORIOA ZTF13
- (39) [Gainerakoan, prokasu adierazle egokiak daude, baina altan dagoen gairaren ahalmen funtzionalaren erregistro urria antzematen da,] _{1N} [eta horrek osasun arretaren emaitzen alderaketa zailtzen du.]_{2S} INTERPRETAZIOA GMB0504

4.3. Helburuzko seinaleak eta anbiguotasuna

HELBURUAREN seinale nabarmenak hauek dira: *i*) menderagailuak, bi ordenatan satellite bukaeran *-t(z)eko* 85etan, NS ordenan satelitean *-t(z)ea* menderagailua 15etan; *ii*) bi ordenatan (SN eta NS) maila lexikoko elementu hauek: satelitean *helburu* izena 19tan eta satellite nahiz nukleoan *asmo* izena 12tan, eta *iii*) SN eta NS ordenan subjuntiboko aditzak 5etan. SN ordenan etiketatu diren seinale ia guztiak perpaus subordinatuko seinaleak dira eta NS ordenan, ordea, aditz nagusiak ere hartu dira seinaleztat. Horiek horrela, HELBURUA-SN ordena ez-kanonikoa ahalbideratzen du euskaraz subordinazioak. Nabarmentzekoa da ez dela juntagailurik edo lokailurik etiketatu seinale gisa. HELBURUZKO seinale nabarmen gehiago erakusten ditugu 12 Taulan.

12. Taula

HELBURUAREN patroien laburpena

JNT eta LOK	Izenak	Aditzak	Morfologia
—	helburua	(subjuntiboa)	-t(z)eko
—	asmo	lortu	-t(z)ea
—	xede	burutu	-t(z)era
—	nahi	ikertu	
—		iritzi	

Helburuzko seinaleen anbiguotasuna. Azkenik, HELBURUA erlazioa ia beti dago seinalatuta: % 99,09 (110etik 109); hortaz, 13 Taulan azertu ditugun 100 seinaleak aztertzean, erlazio horren lagin esanguratsua azertu dugu: % 90,9 (110tik 100).

HELBURUA 13 Taulan erakutsi ditugun seinaleek ia erabateko adostasuna izan dute. Beste erlazioetan ez bezala, ondo seinalatuta egoteaz gain, erlazio horretarako erabiltzen diren seinaleak ez dira beste erlazioetarako erabiltzen; ez dago anbiguotasunik, beraz. Kasu honetan kontuan hartzeko maiztasuna duten seinaleak *-t(z)eko* menderagailuaz ((40) adibidea) osaturik daude, baita *helburu* ((41) adibidea) eta *asmo* izenez ere eta baita subjuntibozko adizkiez ere.

13. Taula

HELBURUA seinalaten duten seinaleen laburpena

Corpusa		HELBURUA		Besteak
ADI-t(z)eko	65	64	98,46	1
ADI-t(z)eko asmo(z/arekin)	12	12	100,00	
ADI-t(z)eko helburuarekin	4	4	100,00	
subjuntiboko marka	5	5	100,00	
helburua	15	15	100,00	
Guztira	101	100	1	

- (40) [Biopsia egiteko,]₁₅ HELBURUA [pistola automatikoa 18Gko orratzarekin erabili dugu.]_{2N} GMB0202
- (41) [VOCALL proiektua (Vocational Language Learning for Less Widely Used and Taught Languages-Lanbide Arloetako Hizkuntzaren Ikaskuntza Gutxi Erabiltzen eta Irakasten Diren Hizkuntzetan) Europako Batzordeak finantzatzen du Leonardo da Vinci Programaren barruan. Proiektuaren koordinadore da Dublin Hiriko Unibertsitatea (Dublin, Irlanda).]_{1N} [Informatika, bulego-lana eta eraikuntzako ikasleentzako hizkuntza ikasteko tresnak egitea da helburua.]₂₅ HELBURUA TERM23

Atal honen laburpena eginez, KAUSAN SN ordena duenean, nukleoan zenbait seinale agertzen dira 11 kasutan, gehienak diskurtso-markatzaileak dira 8 kasutan.⁴⁷ Bestalde, NS ordena duenean, nukleoan oso gutxitan agertzen dira seinaleak eta agertzen direnak maila lexikokoak izaten dira. HELBURUAN, ordea, horrelakorik ez da gertatzen, nukleoan seinaleak ez dira arruntak SN edo NS egituretan (3 guztira) eta agertzen diren horiek ez dira sateliteetan agertzen direnen ezberdinak. KAUSA eta ONDORIOA erlazioekiko beste ezberdintasun bat da diskurtso-markatzaileak ez direla HELBURUAN agertzen.

- i) Seinale gehienak satelitean daude. Salbuespena da KAUSA-SN ordenako erlazioa: nukleoa bigarren diskurtso-unitatea denean, diskurtso-markatzailea erabiltzen da seinaleztat.
- ii) Nukleoan diskurtso-markatzaileak edota maila lexikoko seinaleak etiketatu dira.
- iii) NSn ordena gailentzen da ONDORIOAN eta HELBURUAN, baina ez KAUSAN.

⁴⁷ Gauza bera gertatu da ONDORIOA erlazioko kasu bakarrean, baina kasu bakarra izanik ezin da orokortu.

4.4. Kausa taldeko erlazioen ordena kanonikoa

Hitz hurrenkerari buruzko lanean, Zabalaren (2000) arabera kausazko perpausak perpaus nagusiaren atzean ere joan litezke, azalpenak emateko erabiltzen diren perpausak baitira.

Erlazio hipotaktikoetan nukleartasun joera bi daude:

- a) Nukleo-satelitea (NS): ezkerreko unitatea nukleoa da eta eskuinekoa satelitea.
- b) Satelite-nukleoa (SN): ezkerreko unitatea satelitea da eta eskuinekoa nukleoa.

Erlazio batek joera horietako bat maiztasun handiagoarekin jarraitzen duenean, ordena kanoniko jakina duen erlaziotzat hartzen da erlazio hori. Nabarmendu behar dugu, seinaleak aztertzen ari garen lan honetan, ordena kanonikoa soilik definitzen dela nukleo eta satelite ordenarekin eta ez formazko elementuekin (seinaleekin edo diskurtso-markatzaileekin). Nukleartasunaren ordena kanonikoa garrantzitsua da koherentziaren berri ematean; izan ere, ordena kanonikoa jarraitzen duen testua hobeto ulertzen da ordena kanonikoa jarraitzen ez duen testua baino (Mann eta Thompson, 1988).

Mann eta Thompsonen (1988) erlazio erretoriko hauen ordena kanonikoa deskribatu dute:

- Satelite-nukleoa (SN): lehendabiziko unitatea satelitea denean eta ondorengo nukleoa. Ordena hori jarraitzen dutenak hauexek dira: ANTITESIA, TESTUINGURUA, KONTZESIOA, BALDINTZA, JUSTIFIKAZIOA, ARAZO-SOLUZIOA.
- Nukleo-satelitea (NS): lehendabizi nukleoa eta ondoren satelitea. ELABORAZIOA, AHALBIDERATZEA, EBIDENTZIA, HELBURUA, BIRFORMULAZIOA.

Kausa taldeko erlazioetan ordena kanoniko argiena du ONDORIOA erlazioak, NS ordena (0,98), hain zuzen ere. Bestalde, KAUSA eta HELBURUA erlazioen nukleartasuna ez dago horren ondo definituta.

KAUSA erlazioaren ordena nagusia SN da (0,61)⁴⁸ eta HELBURUA erlazioan, ordea, NS ordena (0,64). Orobat, nabarmentzekoa da seinaleek ez badute forma kanonikoa determinatzen, nukleartasunaren ordena bera jarraitzen dutela eta ia neurri berean, 14 Taulan ikus dezakegunez.

⁴⁸ Nahiz eta Zabalak (2000) dioen bezala kausazko perpausak perpaus nagusiaren atzean ere joan daitezkeen. ONDORIOA erlazioarekin alderantzizko nukleartasuna (SN) pentsaezina den zerbait da laburpen zientifikoetan.

14. Taula

KAUSA erlazioen ordena kanonikoa

Erlazioa	NS		SN		NS		SN	
	Zk.	%	Zk.	%	Seinaleak	%	Seinaleak	%
Kausa	22	0,39	34	0,61	20	0,38	33	0,62
Ondorioa	66	0,98	1	0,02	56	0,98	1	0,02
Helburua	71	0,64	39	0,36	71	0,65	38	0,35

Urre-patroizko erlazio-seinaleen corpora eskuragarri jarri dugu erabili nahi duen ororentzat Euskal RST *Treebankean*.⁴⁹ Egindako lana euskararen Hizkuntzaren Prozesamenduan egin behar diren zenbait atazatan erabilgarria da, besteak beste, kausazko koherentzia-erlazioak automatikoki detektatzeko erregelak proposatzeko eta, ondoren, erregela horiek galderen sorkuntzako sistemetan aplikatzeko.

5. Ondorioak eta etorkizuneko lanak

Lan honetan RST eta Euskal RST *Treebankeko* kausazko koherentzia erlazioak seinalez etiketatzeke lana azaldu dugu. Etiketatze-lan horri esker kausazko erlazioak detektatzeko erlazio-seinaleen deskribapena egin dugu. Horrez gain, seinaleez etiketaturiko corpora eskuragarri jarri dugu erabili nahi duen ororentzat Euskal RST *Treebankean*.

Kausa taldeko erlazio-seinaleen fidagarritasunari dagokionez, HELBURUA izan da etiketatzailen artean adostasun handiena izan duen erlazioa eta automatikoki bilatzen errazena izan daitekeena. KAUSAZKO eta ONDORIOZKO erlazioen artean Duqueren (2014) laneko konklusio bera atera dugu: ONDORIOAKO seinaleek sortu dute etiketatzailen artean ezadostasun gehien.

Erlazioek, nukleartasunari dagokionez, honako forma kanoniko hau jarraitzen dute: *a)* KAUSAk satelite-nukleo (SN: 0,61), *b)* ONDORIOAK nukleo-satelite (NS: 0,98), eta *c)* HELBURUAK nukleo-satelite (NS: 0,64). Koherentzia erlazioen forma kanonikoa zein den jakitea garrantzitsua da erlazioak detektatzeko atazetan; izan ere, erlazio-seinaleak satelite-unitatekoak eta eskuinteko azkenengo diskurtso-unitatean izaten dira ia gehienak (batez ere juntagailuak, lokailuak eta loturazko esapideak). Arau moduko hori bete egiten da ONDORIOAN eta HELBURUAN, NS forma kanonikoa dutelako; baina ez KAUSA erlazioan, zeinek SN forma kanonikoa duen. Horren ondorioz, nukleo-unitatea

⁴⁹ Ikus <http://ixa2.si.ehu.es/diskurtsoa/>

eskuineko unitateaenez, seinaleak egon daitezke nukleoan (ohikoa ez izan arren): juntagailu eta lokailu mailakoak, eta satellite-unitatean leudekeen seinaleak maila morfologiko eta lexikokoak lirerateke. Ber gauza gertatzen da HELBURUA-SN ordenako satellite-unitateko erlazio-seinaleekin.

5.1. Etorkizuneko lanak

Etorkizunean hurrengo lan hauek egiteko asmoa dugu:

- i) Kausa taldeko seinaleak detektatzeko sistema automatikoak osatu eta ebaluatu: a) erregela linguistikoetan oinarritutakoa eta b) ikasketa automatikoetan oinarriturikoa.
- ii) Kausa taldeko seinaleak genero eta domeinu ezberdinetan erabili ahal izateko honako zerrendak osatu:
 - a) aditz psikologikoen euskarazko zerrenda,
 - b) kausa taldeko diskurtso-markatzaileen zerrenda eta
 - c) kausa taldeko seinaleen zerrenda osatuagoa.
- iii) EusEduSeg⁵⁰ segmentatzaile automatikoan (Iruskieta eta Zapirain, 2015) integratu kausa taldeko seinaleak detektatzeko sistema automatikoa, galdera sorkuntzan eta galde-erantzunak aplikatzeko sistemetan erabiltzeko.

Erreferentzia bibliografikoak

- ALDABE, I. (2011). Automatic exercise generation based on corpora and natural language processing techniques. Unpublished doctoral dissertation, UPV/EHU, Donostia, Basque Country.
- ALDABE, I., GONZALEZ-DIOS, I., LOPEZ-GAZPIO, I., MADRAZO, I. and MARITXALAR, M. (2013). Two approaches to generate questions in basque. *Procesamiento del lenguaje natural*, 51:101-108.
- ALDEZABAL, I., ANSA, O. and ARRIETA, B., ARTOLA, X., EZEIZA, A., HERNÁNDEZ, G., LERSUNDI, M. (2001). EDBL: a general lexical basis for the automatic processing of Basque. In *IRCS Workshop on linguistic databases*. Philadelphia (USA).
- ARTOLA, X., DÍAZ DE ILARRAZA, A., EZEIZA, N., GOJENOLA, K., LABAKA, G., SOLOGAISTOA, A. and SOROA, A. (2005). A framework for representing and managing linguistic annotations based on typed feature structures. In *RANLP 2005*.
- ASHER, N. and LASCARIDES, A. (2003). *Logics of conversation*. Cambridge Univ Pr, Cambridge.
- BENAMARA, F. and TABOADA, M. (2015). Mapping different rhetorical relation annotations: A proposal. In *Proceedings of the 4th Joint Conference on Lexical and Computational Semantics (*SEM)*. Denver (USA).

⁵⁰ Ikus: <http://ixa2.si.ehu.es/EusEduSeg/EusEduSeg.pl>

- BENGOETXEA, K. and GOJENOLA, K. (2007). Desarrollo de un analizador sintáctico estadístico basado en dependencias para el euskera. *Procesamiento del lenguaje natural*, 39:5-12.
- DA CUNHA, I. (2013). A symbolic corpus-based approach to detect and solve the ambiguity of discourse markers. In *14th International Conference on Intelligent Text Processing and Computational Linguistics*, Samos, Greece.
- DA CUNHA, I., SANJUAN, E., TORRES-MORENO, J.-M., CABRÉ, M. and SIERRA, G. (2012). A symbolic approach for automatic detection of nuclearity and rhetorical relations among intra-sentence discourse segments in Spanish. *Computational Linguistics and Intelligent Text Processing*, 7181:462-474.
- DAS, D., TABOADA, M. and MCFETRIDGE, P. (2015). RST Signalling Corpus. LDC2015T10. Distributed through the Linguistic Data Consortium.
- DUQUE, E. (2014). Signaling causal coherence relations. *Discourse Studies*, 16(1):25-46.
- EUSKALTZAINDIA (1990). *Euskal gramatika. Lehen urratsak III (Lokailuak)*. Euskaltzaindia, Bilbo.
- EUSKALTZAINDIA (1999). *Euskal gramatika. Lehen urratsak- V (mendeko perpausak-I)*. Euskaltzaindia, Bilbo.
- EUSKARA INSTITUTUA, E. (2015). *Helburu perpausen bereizgarriak*. Sareko Euskal Gramatika (SEG), www.ehu.eus/seg. UPV/EHU, ISBN-978-84-693-9891-3.
- GEORG, G., HERNALDT, H., CAVAZZA, M., PRENDINGER, H. and ISHIZUKA, M. (2009). From rhetorical structures to document structure: shallow pragmatic analysis for document engineering. In *9th ACM symposium on Document engineering*, pages 185-192, Munich, Germany. ACM.
- GIRJU, R. (2003). Automatic detection of causal relations for question answering. In *Proceedings of the ACL 2003 workshop on Multilingual summarization and question answering-Volume 12*, pages 76-83. Association for Computational Linguistics.
- GÓMEZ, I. (1996). Euskararen zatiketa informazionalaren eredu baterantz. *Anuario del Seminario de Filología Vasca «Julio de Urquijo»*, 30(1):195-218.
- IRUSKIETA, M. (2014). Pragmatikako erlaziozko diskurtso-egitura: deskribapena eta bere ebaluazioa hizkuntzalaritza konputazionalan (a description of pragmatics rhetorical structure and its evaluation in computational linguistic). Phd-thesis, Euskal Herriko Unibertsitatea, Donostia. http://ixa2.si.ehu.es/~jibquirm/tesia/tesi_txostena.pdf.
- IRUSKIETA, M., ARANZABE, M.J., DIAZ DE ILARRAZA, A., GONZALEZ, I., LERSUNDI, M. and LOPEZ DE LA CALLE, O. (2013). The RST Basque Treebank: an online search interface to check rhetorical relations. In *4th Workshop «RST and Discourse. Studies»*, Brasil.
- IRUSKIETA, M., DIAZ DE ILARRAZA, A., LABAKA, G. and LERSUNDI, M. (2015a). The detection of central units in Basque scientific abstracts. In *5th Workshop «RST and Discourse Studies»*. Alicante, Spain.
- IRUSKIETA, M., DIAZ DE ILARRAZA, A. and LERSUNDI, M. (2015b). Koherentziazko erlazioak: marko teorikoa eta corpusaren deskribapena, pages 345-361. *Ibon Sarasola, Gorazarre. Homenatge, Homenaje*. UPV/EHUren Argitalpen Zerbitzua. Bilbo.
- IRUSKIETA, M., DIAZ DE ILARRAZA, A. and LERSUNDI, M. (2009). Correlaciones en euskera entre las relaciones retóricas y los marcadores del discurso. In *Proceedings of 27th AESLA International Conference*, pages 963-971, Ciudad Real, Spain.
- IRUSKIETA, M. and ZAPIRAIN, B. (2015). EusEduSeg: a dependency-based EDU segmentation for Basque. *Procesamiento del Lenguaje Natural*. 55: 41-48.
- KHOO, C.S., KORNFILT, J., MYAENG, S.H. and ODDY, R. (1998). Automatic extraction of cause-effect information from newspaper text without knowledge-based inferencing. *Literary and Linguistic Computing*, 13(4):177.

- KNOTT, A. and DALE, R. (1994). Using linguistic phenomena to motivate a set of coherence relations. *Discourse Processes*, 18:35-35.
- KNOTT, A. and SANDERS, T.J. (1998). The classification of coherence relations and their linguistic markers: An exploration of two languages. *Journal of Pragmatics*, 30(2):135-175.
- LONGACRE, R.E. (1985). Sentences as combinations of clauses. *Language typology and syntactic description*, 2:235-286.
- LOPEZ-GAZPIO, I. and MARICHALAR ANGLADA, M. (2013). Web application for reading practice. In *IADAT-e2013: Proceedings of the 6th IADAT International Conference on Education*, pages pp-74. IADAT-e2013. ISBN: 978-84-935915-3-3.
- MANN, W.C. and THOMPSON, S.A. (1987). Rhetorical Structure Theory: A Theory of Text Organization. *Text*, 8(3):243-281.
- MANN, W.C. and THOMPSON, S.A. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. *Text-Interdisciplinary Journal for the Study of Discourse*, 8(3):243-281.
- MARCU, D. (2000). *The theory and practice of discourse parsing and summarization*. The MIT press, Cambridge.
- NEDJALKOV, V. and SILNICKIJ, G. (1973). The topology of causative constructions. *Folia Lingüística*, (6):273-290.
- OATES, S.L. (1999). State of the art report on discourse markers and relations. Technical report, University of Brighton, Information Technology Research Institute.
- ORMAZABAL, J. (2007). Kausatibo aldizkatzeak euskaraz eta inguruko hizkuntzetan, pages 643-654. *Gramatika Jaietan. Patxi Goenagaren omenez*. ASJUren gehigarriak. UPV/EHU. Bilbo.
- OSINAGA, M.S. (2001). *Azalpenezko testu entziklopedikoaren azterketa eta didaktika*. Erein.
- OYHARÇABAL, B. (2002). Kausazio aldizkatzea euskal aditzetan. *Lapurdum*, 7(VII):271-294.
- PARDO, T.A.S. (2005). Métodos para análise discursiva automática. Master's thesis.
- PARDO, T.A.S. and NUNES, M.G.V. (2004). Relações Retóricas e seus Marcadores Superficiais: Análise de um Corpus de Textos Científicos em Português do Brasil. Technical Report NILC-TR-04-03.
- PARDO, T.A.S. and SENO, E.R.M. (2005). Rhetalho: um corpus de referência anotado retoricamente. In *Anais do V Encontro de Corpora*, São Carlos, Brazil.
- RISSELADA, R. and SPOOREN, W. (1998). Introduction: Discourse markers and coherence relations. *Journal of Pragmatics*, 30:131-134.
- TABOADA, M. (2006). Discourse markers as signals (or not) of rhetorical relations. *Journal of Pragmatics*, 38(4):567-592.
- TABOADA, M. and MANN, W.C. (2006). Rhetorical Structure Theory: looking back and moving ahead. *Discourse Studies*, 8(3):423-459.
- VAN DIJK, T.A. (1998). *Texto y contexto: semántica y pragmática del discurso*. Cátedra.
- WEBBER, B.L., STONE, M., JOSHI, A. and KNOTT, A. (2003). Anaphora and discourse structure. *Computational Linguistics*, 29(4): 545-587.
- ZABALA, I. (2000). Hitz-hurrenkera euskara tekniko-zientifikoan. *Ekaia: Euskal Herriko Unibertsitateko zientzi eta teknologi aldizkaria*, (12):143-166.