Universidad del País Vasco Euskal Herriko Unibertsitatea

Koherentziazko diskurtso erlazioen detekzio automatikoa patroien bidez, XMLko erlazio-egiturak oinarri hartuta

Egilea: Asier Kortajarena
Tutorea: Mikel Iruskieta

hap

Hizkuntzaren Azterketa eta Prozesamendua Masterreko titulua lortzeko bukaerako proiektua

2016ko ekaina

Sailak: Lengoaia eta Sistema Informatikoak, Konputagailuen Arkitektura eta Teknologia, Konputazio Zientziak eta Adimen Artifiziala, Euskal Filologia, Elektronika eta Telekomunikazioak.

Laburpena

Hizkuntzaren prozesamenduan testu koherenteetan kausa taldeko erlazioak (KAUSA, ONDORIOA eta HELBURUA) automatikoki hautematea eta bereiztea erabilgarria da galdera-erantzun automatikoko sistemak eraikitzerako orduan. Horretarako Egitura Erretorikoaren Teoria (Rhetorical Structure Theory, aurrerantzean RST) eta bere erlazioak erabiliko ditugu, corpus bezala RST *Treebank*-a (Iruskieta et al., 2013) hartuta, zientziako laburpen-testuz osatutako corpusa, hain zuzen ere. Corpus hori *XML* formatuan deskargatu eta hortik *XPATH* tresnaren bidez informazio garrantzitsuena eskuratzen dugu. Lan honek 3 helburu nagusi ditu: lehendabizi, kausa taldeko erlazioak elkarren artean bereiztea, bigarrenez, kausa taldeko erlazio hauek beste erlazio guztiekin bereiztea, eta azkenik, EBALUAZIOA eta INTERPRETAZIOA erlazioak bereiztea sentimendu analisian aplikatu ahal izateko. Ataza horiek egiteko, *RhetDB* tresnarekin eskuratu diren patroi ensaguratsuenak erabili eta bi aplikazio garatu ditugu. Alde batetik, bilatu nahi ditugun patroiak adierazi eta erlazio-egitura duen edonolako testuetan bilaketak egiten dituen bilatzailea, eta bestetik, patroi esanguratsuenak emanda erlazioak etiketatzen dituen etiketatzailea. Bi aplikazio hauek gainera, ahalik eta modu parametrizagarrienean erabiltzeko garatu ditugu, kodea aldatu gabe edonork erabili ahal izateko antzeko atazak egiteko. Etiketatzaileak ebaluatu ondoren, identifikatzeko erlaziorik errazena HELBURUA erlazioa dela ikusi dugu eta KAUSA eta ONDORIOA bereizteko arazo gehiago dauzkagula ere ondorioztatu dugu. Modu berean, EBALUAZIOA eta INTERPRETAZIOA ere elkarren artean bereiz dezakegula ikusi dugu.

Abstract

At language processing an automatic detection of causal relations (CAUSE, RESULT and PURPOUSE) would be useful in coherent texts, specially building automatic Question Answering(QA) systems. Achieving this task, we use RST (Rethorical Structure Theory) relations and RST *Treebank* (Iruskieta et al., 2013) basque corpus which have many scientific abstract texts. We have download this corpus in *XML* format and get the most important data using *XPATH* for information extraction. This work has 3 goals: firstly, we want to distinguish the causal relation set among themselves, secondly, we want to distinguish the cause subgroup relations from other relations, and finally, distinguish EVALUATION and INTERPRETATION relation to apply on sentiment analysis. To do so, we use some meaningful patterns extracted from *RhetDB* tool and we build two programs. On the one hand, we will develop a search tool which match patterns on the structured relation texts, and on the other hand, we will develop a program which tags relations of a *XML* structured text. Both programs are also easily configurable for anyone. After evaluating the taggers, we conclude that the easiest relation to identify is PURPOUSE and a harder task is to distinguish CAUSE and RESULT relations. More over, we have seen that we can distinguish EVALUATION and INTERPRETATION among themselves.

Gaien aurkibidea

Irudien zerrenda

Taulen zerrenda

1 Proiektuaren definizioa

Hizkuntzalaritza konputazionalean diskurtsoaren egiturari buruz hitz egiten denean bertako lanik esanguratsuenak, erlaziozko egituren identifikazioa (Iruskieta, 2014; Mann eta Thompson, 1988) eta diskurtsoan zehar dauden korreferentziak aukeratzean datza (Goenaga et al., 2012). Master amaierako lan honetan lehenengo aukera landuko dugu eta horretarako diskurtso-egiturako RST (*Rethorical Structure Theory*) deritzon teorian oinarritu gara. Honek gainera, badu euskarazko bertsio bat deskribatuta Euskal RST *Treebank* (Iruskieta et al., 2013) deritzona.

Lan honetan euskarazko RST *Treebank* hori erabiliz diskurtso-egitura aztertzen da, RST-ren barnean aurkitzen diren erlazioak begiratuz eta erlazio horien diskurtso markatzaileak identifikatuz. Erlaziozko diskurtso-egitura esaten dugunean, testu batean koherentzia duten erlazio multzoari buruz ari gara. Erlazio-egitura hori ezagutzeko hizkuntzalaritza konputazionalean koherentziari dagozkion gertaera guztiak hartzen dira kontuan eta zehaztasunez deskribatu behar ditugu diskurtso-egituraren tratamendu automatikoa ondo egiteko. Esaterako har ditzagun 3 adibide hauek:

(1) Mikel etxera joan da. Gripea dauka.

(2) Bilbora edo Donostiara noa edota Gasteizera noa.

(3) Mikel lanera berandu iritsi da. Kotxeko gurpila zulatu zaio.

(1) Adibidea hartzen badugu, ikus dezakegu bigarren esaldia (gripea izateak) lehenengoaren (etxera joatearen) zehaztapena dela eta bien arteko erlazioa adieraztea nahikoa da. Batak bestea zehazten duenez, ELABORAZIOA erlazioa — aurrerago 2 Taulan zerrendatzen dira RST-ren erlazio guztiak— dela esan daiteke, hau da, erlazio semantiko inplizitu bat gertatzen dela esaten da, ez dagoelako diskurtso-markatzailerik. (2) Adibidean, aldiz, argi ikus daiteke SEKUENTZIA bat dela, *edo* eta *edota* juntagailuez markatuta baitago. (3) Adibidean, berriz, interpretazio bikoitza atera dezakegu, koherentziazko erlazio bat baino gehiago atera ditzakegulako. Mikel lanera berandu iritsi izanaren arrazoia gurpila zulatzea dela pentsatzen badugu, *erlazio semantiko inplizitua* dago; izan ere, gurpila zulatzea KAUSA erlazio bat da. Era berean, lanera berandu iristea gainera KAUSA horren ONDORIOA erlaziotzat hartu dezakegu. Hala ere, ez da interpretazio posible bakarra, kotxeko gurpila zulatu izana lanera berandu iristeko JUSTIFIKAZIO bat dela pentsa baitezakegu, baina kasu horretan ez dago esplizituki erlazio semantikorik eta interpretazio honi *erlazio pragmatiko* edo *erretorikoa* esaten zaio.

Lan hau egiterako orduan, gertaera eta interpretazio guzti horiek kontuan hartzea beharrezkoa da. Beraz, gaur egungo testu corpus guztietan egoera hau gerta daitekeenez, dagoeneko existitzen den erlazio semantiko, pragmatiko eta inplizituak barne hartzen dituen marko teoriko bat, Egitura Erretorikoaren Teoria (Mann eta Thompson, 1987) hartuko dugu, hain zuzen ere, lan honetan hemendik aurrera RST bezala ezagutuko duguna. RST-ren bidez koherentzia daukaten testu edo corpus gehienak deskribatu daitezke zein hizkuntza

den kontuan izan gabe. Horren adibide da gaur egun teoria hau, gaztelaniaz[1], frantsesez[2] , portugesez [3] edota euskaraz[4] probatu dela arlo eta esparru ezberdinetako testu corpusak erabiliz azterketa horietan. Lan hau egiterako orduan erabili dugun datu-iturria, Euskal RST *Treebank*-eko (Iruskieta et al., 2013) zientziaren inguruko eta domeinu ezberdinetako 60 euskarazko laburpen-testu zientifikoa izan da, medikuntza (GMB), zientzia-teknika (ZTF) eta terminologia (TERM) alorreko testuak, hain zuzen. Testu hauen xehetasunak 1 Taulan ikus ditzakegu:

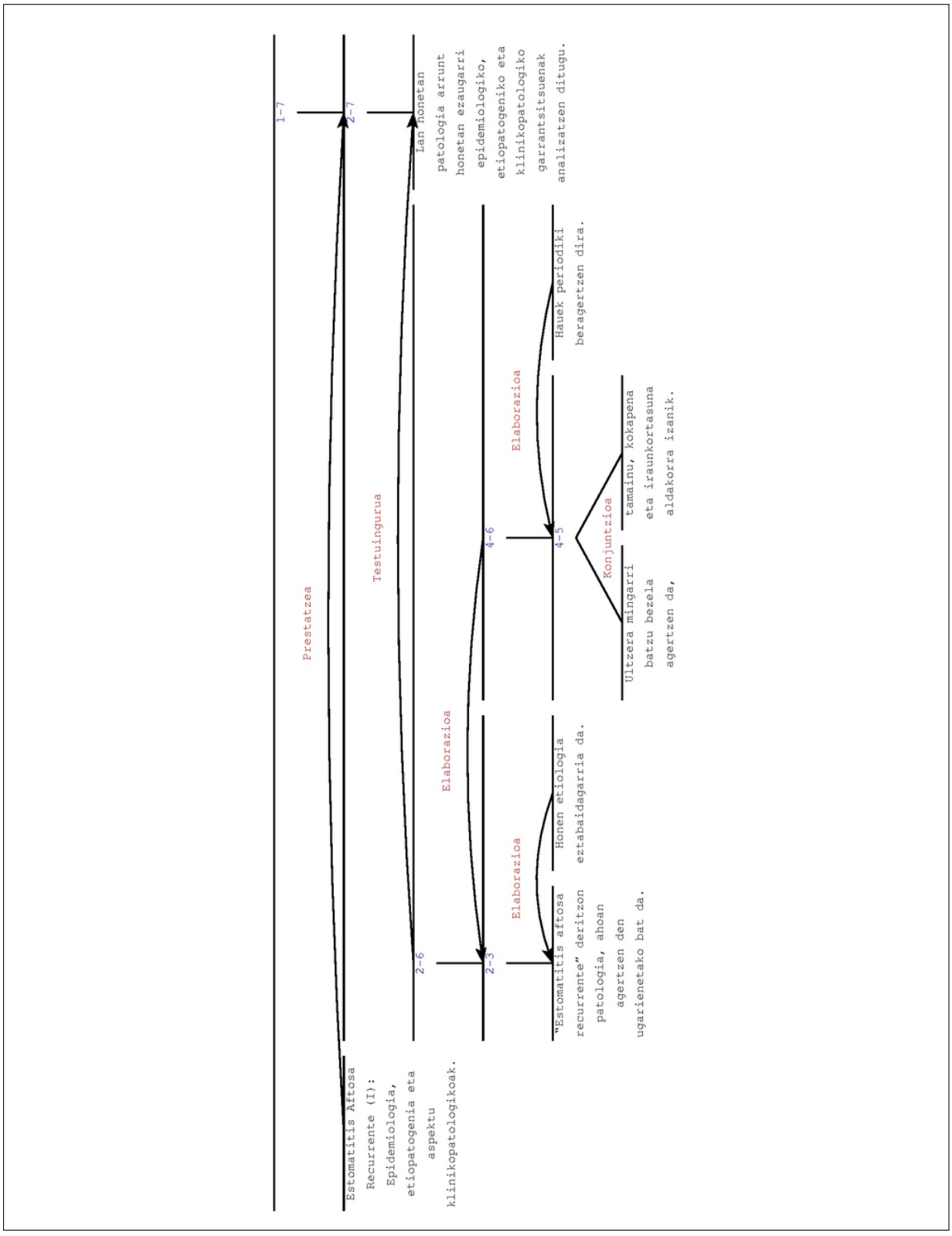
ID	ARLOA	TESTUAK	ESALDIAK	HITZAK	EDU-ak
GMB	Medikuntza	20	198	3010	283
ZTF	Zientzia	20	352	6892	603
TERM	Terminologia	20	253	5664	584
GUZTIRA		**60**	**803**	**15566**	**1470**

Taula 1: Euskal RST *Treebank*eko corpusaren deskribapena

Hauetaz gain, gaur egun *Treebank* hau etengabe handitzen doa eta lanaren hasieran batez ere sentimendu analisirako erabiltzen diren 28 testu gehiago lortu ditugu, guztira erabili ditugun testu kopurua 88koa izanik. Testu zientifiko hauek guztiak, gainera, datubase batean bilduta daude eta bertan testuak segmentatuta ageri dira, honek gure lana konputazionalki egiteko modua errazten digularik. Segmentu hauek analizatzerako orduan faktore asko izan behar ditugu kontuan. Adibidez, argi izan behar dugu testu bateko unitate bakoitzak ez duela garrantzi berdina izango eta gero eta leku unitate gehiagotan gertatu, orduan eta probabilitate handiagoa izango du horrek testuaren barnean esanahi bat edo behintzat garrantzia izateko[5].

Hala eta guztiz ere, hemen azaltzen den lanean ez dugu maila horretan aztertuko;izan ere, dagoeneko Euskal RST *Treebank*ean dauden testuak segmentatuta daude eta hori da lan honek duen abiapuntua. Segmentu horiek gainera, elkarrekin koherentzia erlazioak dituzte eta hori da RSTren berezitasunetako bat; testu koherente bat bestearekin aztertzean beti egongo da nolabaiteko erlazio semantiko, logiko edo pragmatikoren bat. Beraz, segmentu batek beste batekin beti izango du loturaren bat, eta modu honetan, erlazioak identifikatuz eta testu-zati bakoitzak duen garrantzia zein den jakinda, zuhaitz egiturako eskemak lortuko ditugu, 1 Irudian ikus dezakegun bezala.

[1]http://www.sfu.ca/rst/08spanish/index.html
[2]http://www.sfu.ca/rst/07french/index.html
[3]http://www.sfu.ca/rst/07portuguese/index.html
[4]http://www.sfu.ca/rst/07basque/index.html
[5]Baina horrek ere ez du zertan esan nahi garrantzia izango dutenik, esaterako *eta* edo *hala ere* karaktere sekuentziak oso ohikoak dira testu zatietan, baina hizkuntzalaritza konputazionalean horiek duten esanahi semantikoa ezerezean gelditzen da. Hori dela eta, analisia egiterakoan oso ohikoa izaten da askotan gertatzen diren hitz errepikatu — *stop words* (Silva eta Riveiro, 2003)— hauei garrantzirik ez ematea.



Irudia 1: Erlazio bidez lotutako zuhaitz-egitura

Beraz, zuhaitz hauek egiterako orduan, zati guztiek ez daukatela garrantzi berbera ikusten da, eta horretarako, diskurtso unitateei —aurrerantzean EDU— garrantzia esleitu behar zaie. Hau egiteko RSTn oso garrantzitsua den nukleartasuna (*nuclearity*) aztertu behar da. Testu-egitura koherente hauetan bi motatako erlazio-egiturak aurki daitezke:

a) Segmentuek garrantzi berbera dutenak, hauei nukleoaniztunak deitzen zaie.

b) Segmentu batek beste bat baino garrantzi handiagoa dutenean (Mann eta Thompson, 1987). Kasu honetan hierarkia bat osatuko dute segmentuek beraien artean.

Hain zuzen ere, nukleo-unitatea (N) da diskurtso egituran garrantzia ematen zaion kontzeptua. Nukleoa erlazioko zatirik garrantzitsuentzat hartzen da eta aldiz satelite-unitatea (S) garrantzi gutxiagoa duena. Beste modu batera esanda, nukleoak alderdi orokorra hartzen du bere baitan eta aldiz sateliteak nukleoaren zehaztapen bat hartzen du normalean. Bi hauek ongi identifikatzen badira ohikoena da nukleoak sateliterik gabe ere zentzua aurkitzea (Mann eta Thompson, 1988) eta aldiz satelitea bere horretan irakurtzen bada informazio garrantzitsua falta zaiola ikusten da, argi eta garbi testu ulertezin eta koherentziarik gabeko bat irakurtzen delako. Horregatik, oso garrantzitsua da testu egitura hauetan nukleartasuna ondo ebaztea.

Nukleartasun hau irizpidetzat hartuta, RSTren barnean guztira 30 erlazio ezberdineko sailkapena egiten da, *RST-ko sailkapen klasiko hedatua* (Mann eta Taboada, 2010) izenez ezagutzen dena.

<table>
<tr><th colspan="2">Erlazio nukleobakarrak (S-N)</th><th>Erlazio nukleoaniztunak (N-N)</th></tr>
<tr><td>Edukizkoak:</td><td>ELABORAZIOA, METODOA, ZIRKUNSTANTZIA, ARAZO- SOLUZIOA, BALDINTZA, AUKERA, ALDERANTZIZKO BALDINTZA, EZ-BALDINTZATZAILEA, INTERPRETAZIOA, EBALUAZIOA, ONDORIOA, KAUSA eta HELBURUA</td><td rowspan="2">LISTA, DISJUNTZIOA, BATERATZEA, BIRFORMULAZIOA-NN, SEKUENTZIA, KONTRASTEA eta KONJUNTZIOA</td></tr>
<tr><td>Aurkezpenezkoak:</td><td>PRESTATZEA, TESTUINGURUA, AHALBIDERATZEA, MOTIBAZIOA, EBIDENTZIA, JUSTIFIKAZIOA, ANTITESIA, KONTZESIOA, BIRFORMULAZIOA eta LABURPENA</td></tr>
</table>

Taula 2: RSTko sailkapen klasiko hedatua

2 Taulan ikusten den bezala, erlazio nukleo bakarretan azpi-sailkapen bat egiten da. Alde batetik, bertako EDU-en loturak esanahi semantikoa badute, orduan edukizko (*subject*

matter) erlazioa dela esaten da, hau da, idazlearen helburua irakurleak horko unitateen artean erlazio bat jakinarazi nahi dionean gertatzen da. Bestetik, irakurlearengan berak interpretatzeko efektu bat egin nahi denean, orduan erlazioa aurkezpenezkoa (*presentational*) dela esaten da, aurrez aipatutako 3. adibidean erlazio erretorikoak gertatzen diren kasuetan, hain zuzen.

Erlazio guzti hauei, ordea, lan hau egiterako orduan ez diogu garrantzi berbera esleituko. Izan ere, gure lanean berebiziko garrantzia dauka galdera-erantzun sistemetarako (*Question Answering, QA*), laburpen automatiko sistemetarako (*Automatic summarization*) edota sentimendu analisirako baliagarri izan dakigukeen atazak garatzea. Laburpen automatikoak egiteko testuan garrantzi gehien duten segmentuak bilatu eta sailkatu behar direnez, hor ezinbestekoa da unitate zentrala (UZ) detektatzea, azkenean testuko ideia nagusia izan behar delako testu baten laburpenaren zatirik nagusiena. UZ-ren detekzio hori, dagoeneko ikasketa automatikoko (*Machine Learning, ML*) metodo edota erregelen bidez (Iruskieta et al., 2015) egiten da. Galdera-erantzun sistemetan aldiz, galderak automatikoki erantzun behar direnez, ohikoena da galderaren gakoan galdetzailea identifikatu eta horren emaitza lortzea. Har dezagun honako adibidea:

(4) Non bizi da Aitor?

(4) Adibideko galderari erantzuteko, testuan argi dago kokalekuaren entitateak identifikatu behar direla eta dagoeneko existitzen dira euskaraz hori egiten duten tresnak (Eihera[6]), leku, pertsona eta erakundeak bereizten dituena. Honek, ordea, entitatearen araberako bilaketa bat egiten du sintaxi mailan entitateak bilatuz, baina zer gertatzen da era berean testu handietan oso garrantzia duten 'zergatik?' edo 'zertarako?' bezalako galderak erantzuteko orduan? Maila horretan eta geroagokoetan lan egiteak ez du laguntzen eta, beraz, diskurtsoa aztertu beharra dago, erantzuna esaldi ezberdinetan egon daitekeelako. Esaterako, suposatu dezagun zientzia alorreko testu batean (5) Adibideko galdera egin nahi dela eta testuan, besteak beste, (6) Adibidean agertzen den testu-zatia daukagula:

(5) Zergatik dira interesgarriak oinarrizko ikuspegitik nanoegiturak?

(6) Bestetik, Nanoegiturak interesgarriak dira oinarrizko ikuspegitik, beraietan jokabide kuantikoa azaltzen baita.

(5) Adibideko galderari erantzuteko, ez da nahikoa entitateak detektatzea, KAUSA erlazio bat bilatu behar baitugu galderari erantzuteko eta horretarako (6) Adibideko testu segmentua aztertzen denean ikus dezakegu, testuaren amaieran agertzen den baldintzazko *bait-* menderagailuak adierazten digula *beraietan jokabide kuantikoa azaltzen* delako gertatzen dela interesgarria. Beraz, galdera hau erantzuteko ez da nahikoa entitateak detektatzea eta diskurtso-egitura aztertu beharra daukagu, kausa taldeko erlazioak esaterako (Girju, 2003).

[6] http://ixa2.si.ehu.es/demo/entitateak.jsp

Hau honela, diskurtso-egitura eta erlazioak aztertuz, galdera-erantzuneko sistema automatikoetarako honako atazak egingo ditugu lan honetan:

1 : KAUSA, ONDORIO eta HELBURUA erlazioak elkarren artean bereizi.

2 : KAUSA, ONDORIO eta HELBURUA erlazioak beste erlazio guztiekin bereizi.

3 : EBALUAZIOA eta INTERPRETAZIOA erlazioak bereizi, galdera erantzunetako sistemez gain, sentimendu analisirako ere baliagarria izateko.

Ataza hauek egiteko, dagoeneko eskuratuta ditugu 2 Atalean ikusiko dugun *RhetDB* tresnak eskaintzen dituen erlazio hauen seinaleak. Gure lana seinale horien fidagarritasuna ikustea izango da patroi horiek beste erlazio testuetan agertzen diren begiratuz eta horrela patroi horiek erlazio konkretu horretarako daukaten probabilitatea eskuratu eta esanguratsua den erabakitzeko. Teknika honekin emaitza onak espero behar ditugu 1 Atazarako, izan ere, 3 erlazio hauek batuta ez dugu testu kopuru handia eta erlazio jakin batean —esaterako, *-tzeko* HELBURU erlazioetan— behin baino gehiagotan agertzen diren diskurtso-unitateak (seinaleak) aurkitu eta esanguratsuak izateko aukera dago, modu honetan, KAUSA, ONDORIOA eta HELBURUA erlazioak elkarren artean bereiziz.

Errealitatean, ordea, ez dira testu asko izango 3 erlazio horiek bakarrik dauzkatenak barnean eta ohikoena erlazio horiez gain 2 Taulan agertzen diren beste hainbat erlazio agertzea da testuetan. Hori dela eta, galdera-erantzun sistemetarako erlazio hauek ere kontuan hartu behar ditugu seinale esanguratsuenak identifikatzeko orduan. Horretarako, aurrez aipatutako teknikarekin saiatu gaitezke, baina metodo horrek arazo bat planteatzen du; atal honetan azaldu bezala, koherentziazko testuek zuhaitz egitura bat osatzen dute eta zuhaitz egitura hauetan oso litekeena da segmentu bat beste baten nukleoa izanda, segmentu bera —edo segmentu multzoa— aldi berean beste segmentu baten satelitea izatea eta horrela, errekurtsibitatea egotea.

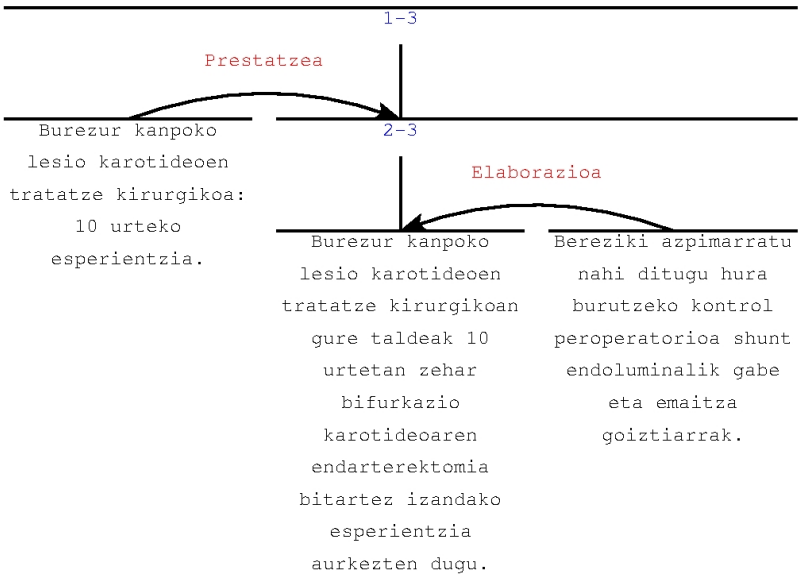
Errekurtsibitatearen adibide bat 2 Irudian ikus daiteke. Testua segmentatuta modu honetara geldituko litzateke:

(7) Burezur kanpoko lesio karotideoen tratatze kirurgikoa: 10 urteko esperientzia.

(8) Burezur kanpoko lesio karotideoen tratatze kirurgikoan gure taldeak 10 urtetan zehar bifurkazio karotideoaren endarterektomia bitartez izandako esperientzia aurkezten dugu.

(9) Bereziki azpimarratu nahi ditugu hura burutzeko kontrol peroperatorioa shunt endoluminalik gabe eta emaitza goiztiarrak.

Adibide honetan ikus dezakegunez, (7) segmentuan dagoen testua, (8) segmentuko testuaren zehaztapen bat da, hau da ELABORAZIOA erlazioa gertatzen da ezkerreko unitatea eskuineko unitatearen satelite bat delako, eta aldiz eskuineko unitatea ezkerreko unitatearen nukleoa. 2 Irudian ikus daitekeen bezala (8) eta (9) segmentuek osatzen duten erlazio hori era berean (7) segmentuaren satelite dela ikus daiteke PRESTATZEA erlazio bat eratuz (7) segmentuaren eta beste bien segmentu-multzoaren artean. Honek, beraz,



Irudia 2: Errekurtsibitatea dagoen zuhaitz-egituraren adibidea

errekurtsibitatea dakar; izan ere, ELABORAZIOAren nukleoa den unitatea, beste erlazio baten satelitea da eta testu hori errepikatu egingo da datu-basean eskuragarri dauden erlazioak analizatzean.

Hau dela eta, lan honetan errekurtsibitateari aurre egin diogu inausketa metodo baten bitartez. Bestalde, erlazioak bereizteko balio duten patroiak aurkitzen ditugunean, patroi horien fidagarritasuna ebaluatu dugu; izan ere, ebaluazioa egin ondoren lortzen ditugun emaitzak onargarriak ez diren bitartean, patroiak hobetu behar ditugu. Corpusetik erauzitako seinaleak corpus berean etiketatuko ditugu eta kontuan izango dugu errore-tarte bat egongo dela. Horrela, patroi fidagarriak lortuko ditugu eta horiek beste corpus etiketatu batean finkatu beharko genituzke (baina hori lan honen helburutik kanpo dago).

2 Aurrekariak

Lan honen helburua IXA taldeko[7] partaideek garatutako programak eta *RhetDB* tresnak ematen dituen emaitzak baliatuta, handik eskuratutako seinaleen bidez erlazio bereizketa hobeak lortzea da aplikazio ahalik eta parametrizagarriena garatuz.

Horretarako, dagoeneko zuhaitz erretorikoetako seinaleak etiketatuta ditugu, eta hori, *RSTTool* (O'Donell, 2000) eta *Rhetorical Data-Base*(*RhetDB*) tresnekin egiten da (Iruskieta, 2014). Hau lortzeko, lehenengo testuak automatikoki segmentatu (Iruskieta eta Zapirain, 2015) eta Unitate Zentrala(*UZ*) detektatzen da (Iruskieta et al., 2015). Ondoren,

[7] http://ixa.si.ehu.es/Ixa

RSTTool-ekin erlazioak etiketatzen dira, eta, bukatzeko, *RhetDB*-k testuetako erlazioak etiketatu eta etiketaturiko patroiak ere multzokatzen ditu. Taboada eta Das-en (2007) lanari jarraituz, etiketatzaile batek (E1) adierazgarrienak iruditzen zaizkion elementuak seinaletzat hartu ditu Euskal RST *Treebank*ean. Ondoren, etiketatzaileak izan dituen arazoen berri ematen du eta seinaleen adibideak prestatzen dituzte beste bi etiketa etiketatzaileek (E2 eta E3), erlazio hauek etiketatzeko. Etiketatzaile batentzat diskurtso-markatzaileak ez dute zertan esanguratsu izan behar beti erlazioetan, baina hala etiketatzea erabaki da, beraien diskurtso-markatzaileen eginkizuna erlazio-egitura esplizitu egitea delako.

Erlazio-seinale hauek begiratzean, ordea, ikus daiteke seinaleen arteko anbiguotasunak eta seinale berbera erlazio ezberdinetako seinaletzat hartzen dituela etiketatzaileak. 3 Taulan ikus daitezke etiketatzaileak etiketatutako seinale anbiguoak.

Seinale anbiguoak			
Seinalea	Agerpen kopurua	Seinalea	Agerpen kopurua
eta	34	-tzeaz gain	4
-nez	15	beraz	4
-tuz	11	ez	4
baina	11	-nez gero	4
bait-	10	bada ere	4
ba-	10	are gehiago	4
bestalde	9	azkenik	4
era berean	8	-lako	3
izan ere	8	horren ondorioz	3
gainera	6	horretarako	3
berriz	5	-larik	3
alde batetik	5	halaber	3
-ta	5	-tzeko orduan	2
	Guztira	**182**	

Taula 3: Euskal RST Treebankeko seinale anbiguoak

Aldiz, erlazio jakin batean zehazki eta bi aldiz baino gehiagotan etiketatzen diren erlazio seinaleak 4 Taulan ikus daitezke. Lehen zutabean seinalea, bigarren zutabean, agerpen kopurua eta hirugarrengoan zein RSTko erlazioetan agertzen den ikus dezakegu.

Seinale ez anbiguoak		
Seinalea	Agerpen kopurua	Erlazioa
-t(z)eko (helburuarekin)	33	Helburua
erabili(z)	16	Metodoa
helburu(a)	10	Helburua
-tzean	9	Zirkunstantzia
ondoren	8	Sekuentzia
adibidez	7	Elaborazioa
hala ere	6	Kontzesioa
-ela eta	5	Kausa
arazo	5	Arazo-soluzioa
izan arren	4	Kontzesioa
-tu ondoren	4	Zirkunstantzia
-nean	4	Zirkunstantzia
nahiz eta	3	Kontzesioa
lortutako emaitzek baieztatzen dute	3	Interpretazioa
hau da	3	Birformulazioa
1.	3	Lista
Guztira	**123**	

Taula 4: Euskal RST Treebankeko seinale ez anbiguoak eta dagokien erlazioa

Guztira, corpus osoko 784 erlaziotik 305 soilik seinalatu dira, hau da erabilitako corpusaren % 38,9 bakarrik seinalatzen dira eta 305 horietatik 123 bakarrik dira erlazio bat seinalearen bidez baieztatzeko gai direnak. Lan honetan aztertzen eta ezberdindu behar diren kausa taldeko (KAUSA,ONDORIOA eta HELBURUA) seinaleetatik HELBURUA eta KAUSA erlaziorako seinale esanguratsuak lortu dira, *-t(z)eko* eta *-ela eta* , hurrenez hurren. Informazio hau, ordea, urria da erlazioak ondo bereizteko.

Horregatik, lan horren ebaluazioa eginda dago (Iruskieta et al., 2016) eta bertatik kausa taldeko patroiak lortu dira harmonizazio prozesu baten bitartez irizpide hau jarraiki:

i) Diskurtso-markatzaileak beti etiketatzea nahiz eta erlazio zehatz batekoak ez izan.

ii) Etiketatzaile bat baino gehiago bat datozen seinaleak etiketatzea.

Irizpide hauek zehazteko urre patroitzat hartzen diren seinaleak zehaztu dira. Seinalearen aurretik, seinalearen kokagunea adierazten da, unitatearen hasieran (*B, Begin*), unitatearen bukaeran (*E, End*), unitatearen erdian (*M, Middle*) edo unitatearen lekuren bat edo gehiagotan (*MM, Multiple*). Har dezagun honako adibide hau:

(10) NUKLEOA: Azken urte hauetan industria desberdinek apostu sendo bat egin dute FeAl aleazioen garapenerako, bereziki aplikazio estrukturaletarako.
SATELITEA: Izan ere, Audi enpresak 1994ean Neckarslum herrian Aluminioaren Zentrua eraiki zuen (duela bost bat urte Aluminioa eta Eraikuntza Arina Diseinurako Zentrua izenez berbaitaiatua), bertan konpaniaren ikonoa den A8 autoa sortuz.

(10) Adibideko erlazio hau N-S norantzako —lehenik nukleoa, gero satelitea— KAUSA erlazio bat litzateke, diskurtso markatzaile edo seinale garbi bat baitauka. Satelitearen hasieran *izan ere* agertzean, KAUSA erlazioa izateko aukerak handiak dira eta horregatik (11) Adibidean ikus daiteke nola lortu patroia:

(11) NUKLEOA: ∅ ; ez dugu baldintza berezirik nukleorako.
SATELITEA: B izan ere ; satelitearen hasieran (*begin*), *izan ere* agertu behar du.

Formatu hau erabiliz ikusiko ditugu Iruskieta et al.ek (2016) argitaratutako patroiak:

a) KAUSA erlazioaren patroiak 5 Taulan. Hemen bi norantzako seinaleak (S-N eta N-S) aurkitu daitezke.

b) HELBURUA erlazioaren patroiak 6 Taulan. Kasu honetan era bi norantzako seinaleak ikus daitezke.

c) ONDORIOA erlazioaren patroiak 7 Taulan. Kausa taldeko beste bi erlazioekin ez bezala, seinaleak N-S norantzan bakarrik daude, baina beste bietan baino seinale eta patroi gehiago lortzen dira.

KAUSAREN PATROIAK			
N-S		**S-N**	
Nukleoa	Satelitea	Satelitea	Nukleoa
∅	M bait-	E -nez	∅
∅	E bait-	M -nez	∅
M interesgarri	E bait-	E -en eraginez	∅
∅	B izan ere	E -nez gero	∅
∅	E eraginda	E -nez gero	B horretarako
∅	M eraginda	M -nez gero	B
∅	M arrazoia	e -ela eta	∅
∅	MM eta arrazoi ... horretarako	E -ela bide	M efektu
E arrazoiengatik	MM -gatik ... -lako ... -gatik	E -ela tarteko	B izan ere
∅	M -elakoan	E -lako	∅
∅	B -en erroan	M emaitza	B eta
∅	M -en ondorioz	M emaitza	∅
∅	E -rekin bat dator	E bait-	B horren ondorioz
		∅	B horren ondorioz
		E -gatik	∅
		E -teagatik	∅
		E eragile izan	B honegatik
		∅	B horregatik
		∅	MM horrek ... ekarri

Taula 5: KAUSA erlazioaren patroiak

HELBURUAREN PATROIAK			
N-S		**S-N**	
Nukleoa	Satelitea	Satelitea	Nukleoa
∅	E -tz?eko	E -t[z]eko	∅
∅	M -tzeko	E -tzeko	MM helburu ... ... lortzeko
M -tzeko	∅	MM -teko ... -teko ... -tzeko	∅
∅	E lortzeko	E -t[z]eko helburuarekin	∅
∅	MM helburu ... -tzea	E -t[z]eko asmoz	∅
∅	M helburu	E -tzeko asmoarekin	∅
∅	MM helburua ... -tea ... ... tzea	E helburuak lortzeko	∅
∅	E -tea ... helburua	B xede hori iristeko	∅
∅	MM -tzea ... helburu ... ... -tea	E -tera	∅
∅	MM -tzeko ... helburu ... ... lortzeko	E -tu nahian	∅
∅	M -tzea	E dadin	∅
∅	MM -tzea ... -tzeko		
∅	E -tzeko asmoz		
E -tzeko asmoz	∅		
∅	MM asmoa ... -tzea		
∅	E dezagun		
∅	E dezaten		
∅	M daitezen		
∅	E burutu nahi izan dugu		
∅	E ikertu nahi dugu		
∅	M betebehar		
∅	M genuke		

Taula 6: HELBURUA erlazioaren patroiak

ONDORIOAREN PATROIAK (N-S)			
Nukleoa	Satelitea	Nukleoa	Satelitea
∅	B ondorioz	∅	M ekar bait-
∅	MM ondorio ... -ri begira	∅	MM bada ... ekarri
∅	B ondorioa	∅	B eta
∅	B -en ondorioz	∅	B eta horrela
∅	B eta ... -en ondorioz	∅	B horrela bada
∅	B eta ondorioz	∅	B honela
∅	MM era honetan ... lortu	∅	MM eta ... aurkitu
m aztertu	MM erakusten ... lortu ... eragin	∅	M aurkitu
∅	MM hori ... lortuz gero	∅	MM eta ... esan nahi du
∅	MM horrela ... lortu	∅	M inplikatzen
∅	M lortu	∅	MM datuek ... adierazten
∅	B lortutako emaitza	∅	B beraz
∅	MM eta emaitzak ... lortu	∅	M beraz
M emaitza	∅	∅	E -larik
∅	M emaitza	∅	M -lako
∅	B emaitza	∅	B hau dela eta
∅	MM emaitza ... erdietsi	∅	B horrenbestez
∅	MM -en emaitzak ... baieztatu ...	∅	E orduan
	... eta ... prebalentzia ... erakutsi	∅	MM aldi berean ... sortzen
∅	E sortuz	∅	M frogatu denez
∅	M sortuz	∅	M eragiteaz gain
∅	E -tuz	∅	MM hartara ... eragina
∅	M dakar	∅	E -raziz
∅	B horien artean	∅	M korrelazioan jarri
∅	M -en bitartez jakin	∅	M areagotu

Taula 7: ONDORIOA erlazioaren patroiak

Taula hauek irizpide hartuta, egingo dugu kausa taldeko erlazio hauen bereizketa, bai beraien artean, eta, bai beste erlazio guztiekin bereiztean ere.

Honek, ordea, ez digu balio 3 Ataza egiteko, hots, sentimendu analisian aplikatzeko EBALUAZIOA eta INTERPRETAZIOA erlazioen arteko bereizketa egiteko. Horretarako ere RhetDB tresnak erlazio horietarako lortutako patroiak eskuragarri izan ditugu eta 8 eta 9 Tauletan ikus daitezke seinale hauek, berriro ere seinaleak norantzaren arabera banatuz.

HAP masterra

INTERPRETAZIOAREN PATROIAK			
N-S		S-N	
Nukleoa	Satelitea	Satelitea	Nukleoa
∅	M emaitzek iradokitzen dute	B egia da	∅
∅	B ondorioz, ez dirudi	M oso goitik	∅
∅	B eta hau kontuan hartzeko[a]	B eta egia esan ez luke	∅
∅	B lortutako emaitzek baieztatzen dute		
∅	MM erdietsiriko emaitzen arabera ... uste dugu		
M emaitza positiboak	M irizpideen arabera kasu honek		
∅	MM -ri esker ... arrakasta irizpidetzat ... nabarmen hobetu		
∅ E balorazioari dagozkie	B eta arrakasta irizpidetzat jo B hobetu beharreko		
∅	MM eta horrek ... emaitzen alderaketa zailtzen du		
∅ ∅ ∅ ∅ ∅	E izan daiteke E bailebiltzan B horrek esan nahi du M interpretatu B eta hauxe garrantzi handikoa		
∅	M emaitzak osatzen eta ulertzen dira		
∅	MM era honetan ... lor daiteke		

Taula 8: INTERPRETAZIOA erlazioaren patroiak

EBALUAZIOAREN PATROIAK (N-S)	
Nukleoa	Satelitea
∅	B ezta hurrik eman ere
∅	M zailtasunak eta lorpenak
∅	M garrantzitsua
∅	M kontuan hart-
∅	E zalantzari tokirik utzi gabe
∅	MM beraz ... oso aproposatzat kalifikatuak
∅	M eta horrek ... abantaila garrantzitsua
∅	M berebiziko garrantzia
∅	M etorkizun handikoak
∅	M oso egokia

Taula 9: EBALUAZIOA erlazioaren patroiak

Lan honen inguruan badaude beste hiru tresna ere antzeko atazak egiten dituztenak.

i) Codra (Joty et al., 2015): RST teoriari jarraitutako diskurtsorako *parser* bat da, baina lan honetan ez bezala, honek *Machine Learning*(*ML*)-eko teknikak erabiltzen ditu, hala nola Markov-en modelo ezkutuak*(Hidden Markov Models)* edota eredu probabilistikoak. Online erabiltzeko moduan dago sarean demoa[8].

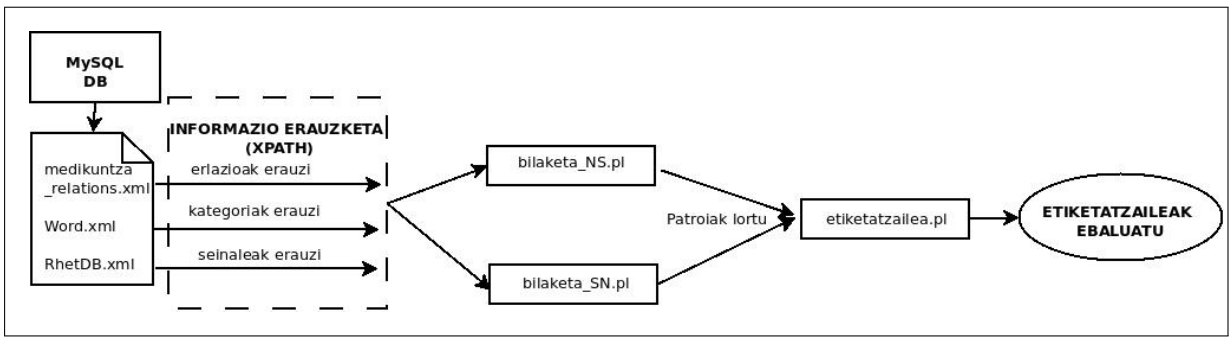
ii) Dizer (Galani et al., 2011): RST teoriari jarraituta sortutako *parser* bat da portugeserako. Lan honetan bezala, testuak segmentatzen dira eta adierazpen erregularren bidez funtzionatzen du batez ere eta online erabiltzeko moduan dago[9]. Lan honetan egiten diren programak *parser* honen antza hartzen duela esan daiteke.

iii) DiSeg 2.0 (da Cunha et al., 2012): Gaztelaniazko diskurtso segmentatzaileak esaldi mailako diskurtso-erlazioak etiketatzen ditu. Testua idatzi eta erregela lexiko eta sintaktikoen bidez testua segmentatuta itzultzen du satelite eta nukleoak identifikatu. Tresna hau ere online erabiltzeko aukera dago[10].

3 Metodologia

Proiektuaren definizioa eta aurrekariak azaldu ondoren, atal honetan lan hau egiteko erabili den metodologia, datu eta prozesuak azaltzen dira. Azalpen hau hurrengo puntutan banatzen da; informazio-iturriak (3.1), informazio erauzketa (3.2) eta garatutako programak (3.3). Azken honetan bi programa egin direnez, beste bi ataletan azalduko dira, bilatzailea (3.3.1) eta etiketatzailea (3.3.2) programak, hurrenez hurren.

Erabili dugun metodologiaren eskema orokorra 3 Irudian ikus dezakegu:



Irudia 3: Metodologia eskema orokorra

[8]Ikus hemen: http://109.228.0.153/Discourse_Parser_Demo/
[9]Ikus hemen: http://www.nilc.icmc.usp.br/dizer2/
[10]Ikus hemen: http://diseg2.termwatch.es/

Lan hau egiterako orduan, beraz, prozesu hau jarraitu dugu:

i) *MySQL* datu-basetik 3 taula eskuratzen ditugu *XML* formatuan.

ii) Taula hauek erabiliz informazioa erauzten dugu *XPATH* erauzketa-tresna erabilita.

a) Medikuntza_relations.xml egitura erabiliz, nahi ditugun erlazioak erauzi ditugu.

b) Word.xml egitura erabiliz, forma bakoitza eta bere kategoria erauzi dugu.

c) RhetDB.xml egitura erabiliz, *RhetDB* tresnak etiketatzen dituen erlazio bakoitzaren seinaleak erauzi ditugu.

iii) Erauzitako fitxategi hauek erabiliz patroi bilaketak egiten dituzten bi PERL programa —*bilaketa_NS.pl* eta *bilaketa_SN.pl*— garatu ditugu, nukleartasunaren arabera egikaritzeko.

iv) Patroiak bilatu ondoren, patroi horiek etiketatu gabeko erlazioetan bilatu eta etiketatzen duen erlazio etiketatzaile bat egin dugu, *etiketatzailea.pl* izenekoa.

v) Nahi ditugun patroiak emanda, etiketatzailearen emaitzak begiratu eta erlazio etiketatzailea ebaluatu dugu, estatistika neurriak kalkulatuz.

3.1 Informazio-iturriak

Proiektua definitzerako orduan aipatu bezala, lan hau egiteko erabili den testu-corpus multzoa Euskal RST *Treebank*etik (Iruskieta et al., 2013) eskuratutakoa da. Banku-datu hau dagoeneko MySQL zerbitzari batean biltegiratuta dago 1 Taulan ageri diren laburpen testuekin. Bertan hainbat taula daude zutabe ezberdinak dituztenak eta horietako taula batzuk baliatu ditugu lan hau egiterako orduan:

1. Medikuntza_relations: Bertan medikuntza testu guztietako segmentuak eta beraien erlazioak zerrendatzen dira. Erlazio horiek aurrez jada definituta daude eta honako zutabeak dituzte besteak beste:

- segment_id: Segmentuaren satelitea da.
- rel_type: Zein erlazio mota den esaten da.
- segment_parent: Zuhaitz egituran erlazioaren gurasoa zein den adierazten da
- norantza: Erlazioa nolakoa den adierazten da gezien bidez; Nukleoaniztuna(↔),N-S norantzako nukleobakarra(←) edota S-N norantzako nukleobakarra(→).

2. Word: Taula honetan, hitz bakoitzaren ezaugarri morfosintaktikoak ageri dira eta, besteak beste, honako zutabeez biltegiratuta dago:

- forma: Hitzaren forma adierazten da.
- kat: Hitzaren kategoria adierazten da.

3. RhetDB: Taula honetan berriz, RhetDB tresnak etiketatutako seinaleak gordetzen dira:

- rel_type: Zein erlazio mota den esaten da.
- non: Seinalea EDU-aren zein kokalekutan dagoen adierazten da.

- testua: Seinalearen testua agertzen da.

Hiru taula hauen zutabe esanguratsuenak bakarrik erakutsiko ditugu azalpena sinplifikatzeko, lan hau egitean *MySQL* datu-baseko eremu hauek erabili baititugu.

Taula hauek *XML* formatuko egituran deskargatu ditugu datu-basetik gerora modu lokalean fitxategi hauek aztertu, informazioa erauzi eta informazio-erauzketa horrekin lortutako datuak, egindako aplikazioekin bateratzeko.

3.2 Informazio erauzketa

Informazio-iturriei dagokienez, (3.1) atalean azaltzen diren fitxategiak ditugu, hots, medikuntza_relations.xml, word.xml eta RhetDB.xml. 3 taula hauek xml formatua oso antzekoa dute, izan ere, automatikoki egindako esportazio baten ondorioz honako egitura daukate:

```
<database name="diskurtsoa">
<!-- Table taula izena -->
<table name="TaulaIzena">
    <column name="Zutabe1">Balioa</column>
    <column name="Zutabe2">Balioa</column>
    <column name="Zutabe3">Balioa</column>
</table>
</database>
```

Hau honela hurrengo pausoa informazio hori erauztea izango da. Horretarako demagun honako informazio hau daukagula.

```
<database>
        <table name="word">
    <column name="dok">TERM23.txt.lemlnk.xml</column>
    <column name="id">w5$\emptyset$</column>
    <column name="forma">teknologiko</column>
    <column name="lema">teknologiko</column>
    <column name="kat">ADJ</column>
    <column name="azp">ARR</column>
    <column name="sent">sent2</column>
    <column name="edu">7</column>
    <column name="sent_pre">sent1</column>
    <column name="sent_post">sent3</column>
    <column name="pos">49</column>
</table>
<table name="word">
    <column name="dok">TERM23.txt.lemlnk.xml</column>
    <column name="id">w51</column>
    <column name="forma">zabala</column>
    <column name="lema">zabal</column>
```

```
            <column name="kat">ADJ</column>
            <column name="azp">ARR</column>
            <column name="sent">sent2</column>
            <column name="edu">7</column>
            <column name="sent_pre">sent1</column>
            <column name="sent_post">sent3</column>
            <column name="pos">5$\emptyset$</column>
        </table>
    </database>
```

XML formatua duten fitxategi guztiak egituratuta daude, etiketa batekin ireki eta beste batekin ixten direlarik. Informazio asko pilatzen da bertan baina gure kasuan taulako eremu jakin batzuk nahi ditugu. Gure lanean, *forma* eta *kat* informazioa eskuratu nahi dugu eta horretarako *XPATH*[11] tresna erabiltzea erabaki dugu. Tresna hau hain zuzen egituraren bidezko bilaketak ahalbidetzen dituen lengoaia bat da. *XML* egitura nodoz jositako zuhaitz gisara interpretatzen du eta eskatzen den araberako bilaketa bat egiten du emaitza modu egituratuan itzuliz. Aurreko adibidearekin jarraituz, forma eta kat zutabeak eskuratzeko *XPATH* bidez *database* barneko *table* barruko *column name* atributua *forma* edo *kat* izatea da. Horretarako agindu hau egikaritu dugu:

```
/database/table/column[@name="forma" or @name="kat"]
```

Honela, taulako sarrera bakoitzeko forma eta kategoria eskuratuko genituzke guk nahi dugun informazioa erauziz. Lan honetan *XPATH* bidezko informazio erauzketa hau, libxml2[12] tresnaren bidez komando lerrotik egikaritu dugularik. Aurreko adibideari jarraituz *XPATH* galdera honela jarriko dugu, terminalean egikaritzeko.

```
xmllint --xpath '/database/table/column[@name="forma" or @name="kat"]'
word.xml
```

Antzeko komandoak erabiliz 3.1 Atalean azaldutako medikuntza_relations.xml fitxategitik N-S norantzako segmentu pareak eskuratu nahi baditugu esaterako, terminalean ondorengo komandoa egikarituko genuke eta honekin 4 Irudian ikus daitekeen emaitza lortuko dugu.

```
xmllint --xpath '/database/table/column[@name="norantza" and text()="NS"]/..'
medikuntza_relations.xml
```

[11] http://www.w3schools.com/xsl/xpath_intro.asp
[12] http://xmlsoft.org/html/index.html

Irudia 4: N-S norantza daukaten eta aurreko aginduak lortzen dituen sarreretako bi

Oso gauza antzekoa egin genezake S-N norantzako segmentuak bilatzeko edota erlazioen araberako bilaketa bat egiteko. Adibidez, KAUSA, ONDORIOA eta HELBURUA erlazioak ez diren beste erlazio guztiak nahiko bagenitu, komando hau exekutatuko dugu:

```
xmllint --xpath '/database/table/column[@name="rel_type" and text()!="kausa"
and text()!="ondorioa" and text()!="helburua"]/..' medikuntza_relations.xml
```

Antzeko aginduak egikarituta ondorengo fitxategiak eskuratu ditugu.

i) NS.xml: N-S norantza daukaten medikuntza_relations.xml taulako sarrera guztiak (5 Irudian bi sarrera ikus ditzakegu).

ii) SN.xml: S-N norantza daukaten medikuntza_relations.xml taulako sarrera guztiak.

iii) NN.xml: N-N norantza daukaten medikuntza_relations.xml taulako sarrera guztiak.

Modu honetara, erlazio guzti hauek banatuta gelditzen zaizkigu nukleartasunaren arabera. Hau oso garrantzitsua da erlazio bakoitzaren seinaleak nukleartasunaren arabera banatzen direlako eta, nahiz eta gure hurbilpenean seinaleak ia beti satelitean agertu, norantza zein den begiratu behar dugu.

Informazio erauzketaren hurrengo pausoa erauzi berri ditugun *XML* fitxategi horien gainean egingo dugu. 1 Atalean proiektua definitzean aipatu bezala, hiru ataza nagusi egin ditugu; KAUSA, ONDORIO eta HELBURUA beraien artean bereizi (1 Ataza), hiru erlazio hauek beste erlazio guztiekin bereizi (2 Ataza) eta azkenik INTERPRETAZIOA eta EBALUAZIOA ezberdindu sentimendu analisirako (3 Ataza).

HAP masterra

1 Ataza gauzatzeko norantza bakoitzeko —hau da fitxategi bakoitzeko— KAUSA, ONDORIOA, HELBURUA erlazioen sarrerak eskuratuko ditugu eta ondoren sarrera horietan gehien interesatzen zaizkigun *segment_id* eta *segment_parent* zutabeak eskuratuko ditugu, aurrez ikusi dugun XPATH galderen bidez. Modu honetan 5 Irudiko emaitza lortu dugu.

```
<column name="segment_id">Horrek ematen dio, horrek baino ez, sinesgarritasun apur bat goitik beherako kaosari.</column>

<column name="segment_parent">Delirium Tremensak jotako morroi baten buruan, beste inorenean ezin liteke halako gezur, eldarnio eta
gorrikeria tendentziosorik egosi.</column>

<column name="segment_id">Badugu liburu honen egileak eta biok gauza bat komunean, biok ala biok Zumarraga herriarekin zerikusia izatea,
alegia.</column>

<column name="segment_parent">Horrek eta nire ogibidearen betebeharrak autorea Larrepetit-en karietara solastatzea posible egin zidaten.</
column>

<column name="segment_id">Gordintasuna eta irudi indartsuak baliatuz, Agirrek protagonistaren eldarnioa eta ezinegona kutsatzen dizkigu.</
column>

<column name="segment_parent">Izan ere, minaren espresio fisikoa hain da erreala, non sentitu ere egin daitekeen, arnasbiderik gabeko
pasarte itogarrietan, intentsitateari mailarik gorenean eusten dioten horietan, goia joz.</column>

<column name="segment_id">Irakurtzen ari nintzela zalantzan aritu bainaiz ez ote zen izango nire gaitasun ahulagatik,</column>

<column name="segment_parent"> barne egituraren kohesio falta zela eta,</column>

<column name="segment_id">Irakurtzen ari nintzela zalantzan aritu bainaiz ez ote zen izango nire gaitasun ahulagatik, barne egituraren
kohesio falta zela eta,edo liburuak berak duen konplexutasunagatik, pasarte ugari behin baino gehiagotan irakurri behar izan ditudala,</
column>

<column name="segment_parent">bidean zerbait ahaztu dudanaren sentsazioa neukalako.</column>

<column name="segment_id">Arestian aipatutako horregatik guztiagatik atsegin ditut, txarragoak edo hobeak,</column>

<column name="segment_parent"> inoren letrekin ez dabiltzan kantariak.</column>
```

Irudia 5: N-S norantzako kausa erlazioen segmentu pareen zerrenda

5 Irudi honek daukan formatuarekin egin dugu lan garatu diren aplikazio guztietan eta fitxategi hauek programen parametroak jarri behar direnez, formatu hau behar beharrezkoa da.

```
<column name="segment_id">1.erlazioari dagokion segmentua</column>
<column name="segment_parent">1.erlazioari dagokion segmentu gurasoa</column>

<column name="segment_id">2.erlazioari dagokion segmentua </column>
<column name="segment_parent">2.erlazioari dagokion segmentu gurasoa</column>

<column name="segment_id">3.erlazioari dagokion segmentua </column>
<column name="segment_parent">3.erlazioari dagokion segmentu gurasoa</column>
                                        |
                                        |
                                        |
<column name="segment_id">n.erlazioari dagokion segmentua </column>
<column name="segment_parent">n.erlazioari dagokion segmentu gurasoa</column>
```

Formatu hau egikarituta dagoen adibidea 6.1 Eranskinean ikus dezakegu.

Etiketa horiek erabiltzeak badu bere justifikazioa: kontuan hartu behar da erlazioen norantzaren arabera segmentu umea eta bere gurasoa aldatu egiten direla. Hau horrela, N-S norantzako erlazioetan *segment_id* nukleoa izango da eta *segment_parent* satelitea, S-N norantzako erlazioetan aldiz alderantziz gertatzen da, *segment_id* satelitea izanik eta *segment_parent* berriz, nukleoa. Ezberdintasun hau aintzat hartu behar dugu, izan ere, garatutako programei parametro bezala pasatzen zaizkion seinaleak zein norantzatan diren definitu beharko baitizkiogu. Hau nola egin hurrengo azpi-atalean (3.3) adieraziko dugu. Horrez gain, etiketak erabiltzeko beste arrazoi bat paragrafo anitzeko segmentuak erabiltzea da. Segmentu hauek zuriuneak izan ditzakete paragrafo jauzia egitean, eta garatutako programek jakin beharra daukatenez noiz hasi eta bukatzen den segmentu bakoitza, etiketa hauek erabiltzen dira segmentuak identifikatzeko.

Informazio-erauzketa honen bidez gainera, beste bi *XML* egitura ere erauzi ditugu. Alde batetik *word.xml* fitxategitik, hitz eta kategoria formak lortu ditugu (6.2 Eranskinean ikus daiteke emaitza). Bestetik, *RhetDB.xml* fitxategitik erlazio bakoitzaren seinaleak eta kokalekua ere erauzi ditugu.

3.3 Garatutako programak

Lan honetan bi programa edota aplikazio garatu ditugu *PERL*[13] lengoaian programatuta. Programazio lengoaia hau erabiltzea erabaki dugu testu-fitxategiak aztertu eta manipulatzeko bereziki sortutako lengoaia delako (Astigarraga et al., 2009) eta adierazpen erregularrak modu sinple batean erabiltzeko aukera ematen duelako. Alde batetik, erlazio zerrenda bat eta seinale zerrenda etiketatu bat pasatu ondoren, seinale horiek bilatzen dituen bilatzaile bat egin dugu, eta, bestetik, erlazio zerrenda etiketatu gabe bat eta seinale zerrenda bat pasata, seinale horien arabera erlazioak etiketatzen dituen etiketatzaile bat eraiki dugu.

Gainera, proiektua definitzerako orduan (ikus 1 Atala) aipatu bezala, programa horiei ahalik eta erabilgarritasun gehien emateko programa parametrizatua izateari eman zaio lehentasuna eta gaur egun kalkulu orriak asko erabiltzen direnez hizkuntzaren prozesamenduan, emaitzak automatikoki kalkulu orrietan ere idazten dituzte egindako bi programek[14].

3.3.1 Bilatzailea

Garatu den lehenengo programa seinale bilatzaile bat izan da eta fitxategi batzuk pasatu behar zaizkio parametro modura:

1. Lehen parametroan aztertu nahi den erlazio zerrenda ematen da. Honen adibidea, 6.1 Eranskinean ikus daiteke.

2. Bigarren parametroan bezala bilatu nahi diren patroiak ematen dira. Honen adibidea, 6.3 Eranskinean ikus daiteke.

[13] https://www.perl.org/

[14] Horretarako, ordea, beharrezkoa dugu Excel modulu bat instalatzea 6.10 Eranskinean azaltzen diren pausoak jarraituz.

3. Hirugarren parametroan edonolako izen bat pasatzen zaio, kalkulu-orria izen horrekin sortzeko.

Programa honen funtzionamendua hurrengo urratsetan gauzatzen da:

i) 1. parametroan dauden segmentuak prozesatu eta datu-egitura batean gordetzen dira nukleo-satelite bikoteak.

ii) 2. parametroan dauden patroiak, nukleo-satelite bikoteetan bilatzen dira bertan adierazten den kokalekuaren arabera.

iii) Nukleo-satelite bikoteak patroiren batekin parekatze positiboa lortzen duenean, patroi horren asmatze kopuruaren kontagailua gehitzen da.

iv) Emaitza hauek irteera estandarrean pantailaratu eta 3. parametroan adierazitako izenarekin kalkulu-orri bat sortzen da emaitzak erakutsiz.

Hortaz, programa honi lehen parametro bezala pasatzen den erlazio zerrenda (5 Irudian ikusten den modura) eta bigarren argumentu bezala bilatu nahi den seinale zerrenda eman behar diogu. Seinale horiek diskurtso-unitateetan bilatu beharreko patroi gisara interpretatzen ditu programak eta formatu honetan idatziko ditugu lerro batean[15].

```
[B/E/M/MM][Segment\_id segmentuko patroia] * [B/E/M/MM][Segment\_parent
segmentuko patroia]
```

Segmentuen bilaketa patroiak '*' banatzaile baten bitartez banatzen dira eta patroia idatzi aurretik non aurkitu behar den adierazi behar da, segmentuaren hasieran (B, *begin*), segmentuaren erdian (M, *middle*), segmentuaren bukaeran (E,*end*) edota segmentuaren leku bat baino gehiagotan (MM,*multiple*).

Patroiak aldiz adierazpen erregular gisa idatz daitezke, baina beti ere kontuan hartuz honako baldintzak:

(1) Gidoia (-) aurkitzen badu, hitz baten aurrizki edo atzizki gisara interpretatzen du, gidoia non dagoen arabera.

(2) Hiru puntu jarraian (...) aurkitzen baditu, aurreko hitzaren eta hurrengo hitzaren artean zerbait egongo dela suposatzen du. Adierazpen erregularrak erabili nahi izanez gero, hiru puntuak '.*' adierazpenaren baliokide direla esan daiteke.

(3) Kakotx ([ ]) artean jarriz gero barnean hitzen kategoria bat sartu behar da eta kategoria horretako edozein hitz baliozkotuko du patroiak.

(4) Diskurtso-unitate horrek ez badu patroirik zero '$\emptyset$' zenbakia jartzen da.

[15]Ohartu *segment_id* eta *segment_parent* testuak ezberdinak direla N-S norantza daukaten erlazioetan eta S-N norantza dauzkate erlazioetan.

Adibide sinple bat probatuz, demagun N-S norantza daukaten erlazioetan ONDORIOAren seinaleetan erabiltzen diren hainbat seinale bilatu nahi direla, 10 Taulan ikus ditzakegunak esaterako:

Adibidea	Nukleoa	Satelitea
1	∅	MM hartara ... eragina
2	∅	E -raziz
3	∅	B [DET] artean
4	M aztertu	MM erakusten ... lortu ... eragin

Taula 10: Patroien adibideak

Kasu horretan lehenengo nukleoa eta gero satelitea zehazten denez (ikus 6.3 Eranskina), '*' banatzaileaz ezkerrekoa nukleoarentzako patroia *segment_id* testuetan bilatzekoa eta eskuinaldekoa satelitearentzako patroia izango da, *segment_parent* testuetan bilatzekoa. Lehenengo adibidean, satelitean '*hartara*' seinalea bilatuko du eta beste zerbaiten ondoren '*eragina*' izeneko forma bilatuko du. Lehen parametro bezala pasatako testuko *segment_parent* batean patroia parekatzen bada, patroi horren asmatzea —*pattern matching*— gertatuko da. Bigarren adibidean aurrekoan bezala nukleoan ez du ezer bilatuko baina oraingoan satelitearen balizko patroirako '*E*' ezarri denez, patroia satelitearen amaieran aurkitzen saiatuko da, hots, satelitearen amaieran '*-raziz*' atzizkia duen satelite testurik baden begiratzen du. Hirugarren adibidean berriro ere nukleoa ezikusi egiten du programak eta oraingoan '*B*' aurkitu duenez, satelitearen hasieran determinante kategoria duen hitz formaren bat eta ondoren '*artean*' hitza dagoen segmentuak parekatzen ditu. Laugarren adibidean berriz, aurreko hirurak ez bezala honek bi baldintza ditu, nukleoaren segmentuaren erdialdean '*aztertu*' hitza bilatzen du eta nukleoan kasu hori gertatzen den erlazioetan satelitean '*erakusten ... lortu ... eragin*' patroia bilatzen du. Nukleoko patroiaren eta satelitearen patroiaren parekatzea positiboa baldin bada, orduan bat datozela kontatuko du.

Programa honek, beraz, patroien fitxategia lerroz lerro irakurtzen du eta lehen parametro bezala pasatako erlazioetan banan banan begiratzen ditu ea parekatze positiboa egon den (ikus 6.4 Eranskina) eta hala egon bada, patroi horren kontagailua inkrementatzen da. Gainera, asmatze kopurua kontatu eta hirugarren parametro bezala pasatzen den izenarekin automatikoki kalkulu orri bat sortu eta bertan idazten da satelite eta nukleo patroi bakoitzak izan duen parekatze positibo kopurua (ikus 6.5 Eranskina). Bilatzaile honek eskuratzen dituen emaitza guztiak, emaitzak atalean (4 Atala) ikus ditzakegu.

Programa honek abantaila nagusi hauek ditu: 1) erlazio motekiko guztiz independentea da, 2) adierazpen erregularrak erabil daitezke seinaleen patroiak definitzerakoan eta 3) adierazpen erregularrak ezagutzen ez dituztenentzat '...' eta '-' bidez patroiaren malgutasuna lortzen da eta kategoria lexikalak sartzeko aukerak emaitza hobeak lortzen ditu. Parametrizagarritasun honekin, parametro bezala pasatako edozein erlazio zerrenda hartu eta EDU-etan dauden seinaleak hartu ondoren, bilatu eta kontatu nahi dira bertan programaren erabiltzailearen xedea edonolakoa izanik.

Gure kasuan, seinale edo patroi hauek *RhetDB* programak etiketatu eta ematen dizkigu eta berauek 5 Taulatik hasi eta 9 Taula arte ikus ditzakegu 2 Atalean. Patroi horiek zehazterako orduan, ordea, hainbat erabaki hartu ditugu:

i) Aditz laguntzaileak orokortzea; era honetan, *'dute'*, *'dugu'*, *'daukagu'* eta horrelakoetan, hitzak baliokidetzat hartzen ditugu, hitz forma bera ez baita seinale esanguratsua.

ii) 1.pertsona eta 3.pertsona bateratzea; demagun *'gauzatu da'* seinalea daukagula, eta hori bilatzea nahi dugula, baina gaur egun oso ohikoa da 3.pertsonaz gain 1.pertsona pluralean ere idaztea eta *'gauzatu dugu'* bere baliokidetzat har dezakegu.

iii) Testuetan askotan agertzen diren *stop-word* hitzak saihestea, ez baitute izaten normalean balio semantiko handirik.

iv) Sinonimoak erabili, esanahi bera edo antzekoa duten hitzak bateratuz. Adibidez, *'erdietsi'*, *'eskuratu'* edo *'lortu'* aditzak erlazionatu, eta, *RethDB* tresnak *'emaitza ... erdietsi'* patroia etiketatu badu, guk garatutako programak *'emaitza ... eskuratu'* edota *'emaitza ... lortu'* ere bilatzea.

v) Orokorrean patroi hobekuntzarako baliagarri den edozein aldaketa, kategorien erabilera, zehaztapena, hitzak aukeran jartzea, hala nola. Hau guztia identifikatu nahi dugun erlazio motaren menpe izango dugu.

Hala ere, lan honetan aplikazio hau helburu zehatz batera begira erabili dugu, alde batetik KAUSA, ONDORIOA eta HELBURUA erlazioen seinaleen eta bestetik EBALUAZIOA eta INTERPRETAZIO seinaleen fidagarritasuna ikusteko, koherentzia daukaten testu mota guztietan. Honako prozesua jarraitu da:

i) Analizatu nahi den erlazioaren seinaleak eskuratu eta fitxategi bat sortu seinale horien patroiak aurrez aipatutako formatuan jarriz.

ii) Programari erlazio horretako testuak eman eta seinale bakoitza zenbat aldiz agertzen den ikusten da automatikoki kalkulu-orri batean jarriz asmatze kopuruak.

iii) Prozesu bera errepikatzen joan konparatu nahi diren beste erlazio guztiekin, erlazio horietako testuak parametroa jarriz, baina beti ere erlazio seinale bakoitzarekin. Azken finean, parametro horren funtsa beste erlazioetan ez daudela bilatzea da.

iv) Seinale horien konparaketa taula bat eskuratzen da eta erlazio horretarako etiketazio prozesurako seinale esanguratsuenak lortzen dira.

Honek, ordea, 1 Atalean azaldutako arazo bat ekartzen digu, errekurtsibitatearena hain zuzen ere. Hori dela eta, 2 Irudiko arazo hori ekiditeko teknika bat baliatu dugu. Segmentu parea analizatzen doan heinean, programak begiratuko du ea unean analizatzen dagoen segmentu zatia ordurarte prozesatutako testuen azpi-katea den, eta, hala bada, segmentu

hori ez analizatzea erabakitzen du aplikazioak. Azken finean, zuhaitzaren inausketa bat egiten ari gara, dagoeneko erabilitako adarrak moztuz. Metodo honen eraginkortasuna agerian geratzen da emaitzak (ikus 4 Atala) ikustean.

Behin patroi zerrenda eta patroi bakoitzaren maiztasunak izanda, zein patroi diren esanguratsu eta fidagarriak ebatzi behar dugu. Horretarako honako irizpide hauek jarraituko ditugu lehentasunaren arabera:

i) Maiztasun handienak dituzten patroiak aukeratuko ditugu, beti ere nahi ez dugun erlazioetan askotan agertzen ez badira.

ii) Patroi zehatzak aukeratuko ditugu, beste erlazioetan aurkitzea zailak izango direnak. *MM* kokalekua edo unitate anitzak dituzten patroiak, testuetan ohikoak ez diren hitzak dituztenak.

iii) Unitatearen erdian agertzen diren seinaleak soilik dituzten patroiak ekidingo ditugu, horiek aukera handiago baitute, nahi ez dugun erlazioetan agertzeko, batez ere arruntak den *amaitu* hitza.

iv) Erlazio mota kontuan izanik beharrezko ikusten ditugun irizpideak hartuko ditugu, baina, horretarako, erlazio hauek ondo ezagutu behar ditugu.

3.3.2 Etiketatzailea

Garatu den bigarren programa, berriz, erlazioen etiketatzaile bat izan da, eta aurrekoa bezala ahalik eta erabilgarri eta parametrizagarriena egin da. Bilatzailearen gisara, aplikazio honi ere 3 parametro pasa behar zaizkio:

1. Lehen parametroan etiketatu nahi den erlazio zerrenda ematen da. Honen adibidea, 6.6 Eranskinean ikus daiteke.
2. Bigarren parametroan bezala etiketatu nahi diren erlazioaren izenak eta bere patroiak ematen dira. Honen adibidea, 6.7 Eranskinean ikus daiteke.
3. Hirugarren parametroan edonolako izen bat pasatzen zaio, emaitzak idazten dituen kalkulu-orria izen horrekin sortzeko.

Programa honen funtzionamendua hurrengo urratsetan gauzatzen da:

i) 1. parametroan dauden segmentuak eta erlazio bakoitzaren norantza prozesatu ondoren, datu-egitura batean gordetzen dira nukleo-satelite bikoteak eta beraien norantzak.

ii) Nukleo-satelite nukleo-satelite bikote hauetan bilatzen dira 2. parametroan pasatako patroiak.

iii) Nukleo-satelite bikoteak patroiren batekin parekatze positiboa lortzen duenean, patroi horri dagokion erlazio hori etiketatzen da.

HAP masterra

iv) Emaitza hauek irteera estandarrean pantailaratu eta 3. parametroan adierazitako izenarekin kalkulu-orri bat sortzen da emaitzak erakutsiz.

Beraz, berriro bi fitxategi pasatu behar zaizkio programari. Lehenengoa, aurreko programaren modura erlazio zerrenda bat da, baina kasu honetan erlazioaren norantza ere adierazi beharko zaio; izan ere, etiketazio prozesua ezberdina da nukleartasun ordenaren arabera. Fitxategi honek honako formatua izango du:

```
<column name="segment_id">1.erlazioari dagokion segmentua</column>
<column name="segment_parent">1.erlazioari dagokion segmentu gurasoa</column>
<column name="norantza">1.erlazioari dagokion norantza (NS/NN/SN)</column>

<column name="segment_id">2.erlazioari dagokion segmentua </column>
<column name="segment_parent">2.erlazioari dagokion segmentu gurasoa</column>
<column name="norantza">2.erlazioari dagokion norantza (NS/NN/SN)</column>

<column name="segment_id">3.erlazioari dagokion segmentua </column>
<column name="segment_parent">3.erlazioari dagokion segmentu gurasoa</column>
<column name="norantza">3.erlazioari dagokion norantza (NS/NN/SN)</column>
                                        |
                                        |
                                        |
<column name="segment_id">n.erlazioari dagokion segmentua </column>
<column name="segment_parent">n.erlazioari dagokion segmentu gurasoa</column>
<column name="norantza">n.erlazioari dagokion norantza (NS/NN/SN)</column>
```

Formatu honen erabileraren adibidea 6.6 Eranskinean ikus dezakegu.

Bigarren parametro bezala pasatako fitxategiak berriz, erlazioaren izena eta bere ustezko seinale fidagarriak izango ditu. Erlazio bat baino gehiago etiketatu nahi bada, zuriune batez banatu behar dira sistemak jakin dezan erlazio bakoitzaren seinale edo diskurtso-unitateak zeintzuk diren. Horretaz gain, aurrez aipatutako moduan satelite eta nukleoaren ordenak berebiziko garrantzia dauka, seinale ziurrak identifikatzerako orduan eta, beraz, fitxategi hauetan seinalearen ordena —S-N edo N-S— adierazi behar da lerro hasieran eta ondoren patroiak adierazi, 6 Irudian ikusten dugun moduan.

```
KAUSA
NS * 0 * B izan ere
NS * 0 * MM -gatik ... -lako ... -gatik
NS * 0 * MM eta arrazoi ... horretarako
NS * 0 * E -rekin bat dator
SN * E -nez (gero)? * 0
SN * M -nez * 0
SN * E -ela eta * 0
SN * E -ela bide *  M efektu
SN * B horren ondorioz * 0
SN * E eragile [adl] * B honegatik

ONDORIOA
NS * 0 * b ondorioz |
NS * 0 * MM ondorio ... -ri begira
NS * 0 * B ondorioa
NS * 0 * B -en ondorioz
NS * 0 * MM era honetan ... lortu
NS * 0 * MM erakusten ... lortu ... eragin
NS * 0 * B eta horrela
NS * 0 * M emaitza
NS * 0 * B beraz
NS * 0 * M aurkitu
NS * 0 * MM eta ... esan nahi [adl]
NS * 0 * MM aldi berean ... sortzen
NS * 0 * E -raziz
NS * 0 * B [det] artean
NS * 0 * M korrelazioan [adi]
NS * 0 * M -en bitartez jakin
```

Irudia 6: KAUSA eta ONDORIO erlazioaren patroien adibidea

Lan honetan erabili ditugun bi patroi fitxategiak 6.7 Eranskinean ikus ditzakegu. Alde batetik, kausa taldekoak aztertu daitezke 6.7.1 Eranskinean. Bestetik, INTERPRETAZIOA eta EBALUAZIOA erlazioen etiketazio patroiak ikus ditzakegu 6.7.2 Eranskinean.

Programa honek bigarren parametro bezala pasatako sateliteko eta nukleoko patroiak bilatzen ditu, etiketatu nahi diren erlazioetan, eta patroi horietakoren bat bilatzen badu, lehen lerroan adierazitako erlazio izenarekin etiketatzen du (ikus adibidea 6.8 Eranskinean). Gainera, bilatzailearen moduan automatikoki hirugarren parametro bezala pasatako izenarekin kalkulu orri bat sortzen du erlazioaren nukleoaren segmentua, satelitearen segmentua eta etiketatu duen erlazioaren izena idatziz, etiketatu duen kasuetan. Honen adibidea 6.9 Eranskinean ikus dezakegu. Honez gain, sistemak analizatu dituen erlazio kopurua eta horietako zenbat etiketatu diren ere esaten digu.

3.4 Ebaluazio tresnak

Azkenik, etiketatze prozesua gauzatu ondoren, ebaluazioa egin behar diogu etiketatzaileari. Erlazio jakin bat identifikatzen duten patroi zerrenda fidagarria den edo ez ebatzi nahi dugu. Hori ikusteko lehen pausoa corpus berdina erabilita etiketazioa egitea izango da eta benetako datuekin alderatu. Horretarako honako faktoreak hartuko ditugu kontuan:

i) Identifikatu nahi dugun erlazioa zenbat aldiz lortu dugun modu zuzenean begiratuko dugu erlazioaren asmatze-tasa eskuratuz.

ii) Identifikatu nahi dugun erlazioa zenbat aldiz etiketatu den oker ikusiko dugu, erroretasa kalkulatzeko.

iii) Erabilitako patroi zerrendak daukan fidagarritasuna ikusteko, doitasuna (*precision*), estaldura (*recall*) eta F-neurria (*F-score*) estatistika neurriak aterako ditugu.

Hemen ateratako emaitzak onargarriak ez diren bitartean, patroiak hobetu behar ditugu; izan ere, corpus berberetik ateratako seinaleak corpus berdinean etiketatzean, teorian behintzat emaitza onak eman behar ditu, nahiz eta beti errore-tarte bat egongo den. Behin patroi fidagarriak lortuta, hurrengo pausoa corpus berri batekin probatzea izango genuke eta corpus hori etiketatuta lortzen ditugun emaitzak ikertzea. Horretarako, corpuseko erlazioak dagoeneko eskuz edo erdi-automatikoki etiketatuak egotea lagungarri da.

3.5 Emaitzak ebaluatzeko fitxategiak

Informazio erauzketa (3.2 Atala) eta programak garatu (3.3 Atala) ondoren hiru katalogo sortu ditugu: *NS* eta *SN* izenekoak norantza bakoitzerako fitxategiak gordetzeko, eta *Etiketatzailea* izeneko katalogoa etiketatzailearekin.

a) *NS* katalogoaren barruan fitxategi hauek daude:

NS

NS.xml:	NS norantzako erlazio guztiak
KAUSA_NS.xml	Kausa eta N-S diren erlazio guztiak
KAUSA_NS.txt	Kausa eta N-S diren erlazioen segmentuak
ONDORIOA_NS.xml	Ondorioa eta N-S diren erlazio guztiak
ONDORIOA_NS.txt	Ondorioa eta N-S diren erlazioen segmentuak
HELBURUA_NS.xml	Helburua eta N-S diren erlazio guztiak
HELBURUA_NS.txt	Helburua eta N-S diren erlazioen segmentuak
EBALUAZIOA_NS.xml	Ebaluazioa eta N-S diren erlazio guztiak
EBALUAZIOA_NS.txt	Ebaluazioa eta N-S diren erlazioen segmentuak
INTERPRETAZIOA_NS.xml	Interpretazioa eta N-S diren erlazio guztiak
INTERPRETAZIOA_NS.txt	Interpretazioa eta N-S diren erlazioen segmentuak
kategoria:	Forma bakoitzaren kategoria zerrenda
bilaketa_NS.pl	**N-S Patroien bilaketa egiten duen Perl programa**
SEINALEAK	
NS_kausa.txt	NS Kausaren patroiak
NS_ondorioa.txt	NS ondorioaren patroiak
NS_helburua.txt	NS helburuaren patroiak
NS_ebaluazioa.txt	NS ebaluazioaren patroiak
NS_interpretazioa.txt	NS interpretazioaren patroiak

b) Aldiz, *SN* katalogoaren barruan honakoak daude:

SN

SN.xml:	S-N norantzako erlazio guztiak
KAUSA_SN.xml	Kausa eta S-N diren erlazio guztiak
KAUSA_SN.txt	Kausa eta S-N diren erlazioen segmentuak
ONDORIOA_SN.xml	Ondorioa eta S-N diren erlazio guztiak
ONDORIOA_SN.txt	Ondorioa eta S-N diren erlazioen segmentuak
HELBURUA_SN.xml	Helburua eta S-N diren erlazio guztiak
HELBURUA_SN.txt	Helburua eta S-N diren erlazioen segmentuak
EBALUAZIOA_SN.xml	Ebaluazioa eta S-N diren erlazio guztiak
EBALUAZIOA_SN.txt	Ebaluazioa eta S-N diren erlazioen segmentuak
INTERPRETAZIOA_SN.xml	Interpretazioa eta S-N diren erlazio guztiak
INTERPRETAZIOA_SN.txt	Interpretazioa eta S-N diren erlazioen segmentuak
kategoria:	Forma bakoitzaren kategoria zerrenda
bilaketa_SN.pl	**S-N patroien bilaketa egiten duen PERL programa**
SEINALEAK	
SN_kausa.txt	SN Kausaren patroiak
SN_ondorioa.txt	SN ondorioaren patroiak
SN_helburua.txt	SN helburuaren patroiak
SN_ebaluazioa.txt	SN ebaluazioaren patroiak
SN_interpretazioa.txt	SN interpretazioaren patroiak

c) *Etiketatzailea* katalogoaren kasuan, berriz, bi fitxategi ditugu soilik, programa eta kategorien fitxategia.

Etiketatzailea

kategoria:	Forma bakoitzaren kategoria zerrenda
etiketatzailea.pl	**Erlazioak etiketatzen dituen PERL programa**

4 Emaitzak

Atal honetan, gure laneko helburuak betetzeko lortu ditugun emaitzak azaldu eta ondorioztatuko ditugu hurrengo eginkizunak irizpide hartuta:

a) KAUSA, ONDORIOA eta HELBURUA erlazioak bereiztea elkarren artean (4.1 Atala).

b) KAUSA, ONDORIOA eta HELBURUA erlazioak bereiztea beste erlazio guztiekin (4.2 Atala).

c) EBALUAZIOA eta INTERPRETAZIOA erlazioak bereiztea (4.3 Atala).

Horretarako, 3.5 Atalean ikusten diren dokumentuak lortu ditugu, eta, hortik, erabili dugun corpusean norantza bakoitzeko zenbat erlazio dauden jakin dezakegu. Kopuru hauek 11 Taulan ikus ditzakegu.

	ERLAZIOAK		
	N-S	S-N	GUZTIRA
KAUSA	44	44	88 erlazio
ONDORIOA	81	0	81 erlazio
HELBURUA	110	35	135 erlazio
EBALUAZIOA	132	2	134 erlazio
INTERPRETAZIOA	61	1	62 erlazio
BESTEAK1	1047	507	1554 erlazio
BESTEAK2	1089	586	1675 erlazio

Taula 11: Euskal RST *Treebank*-en corpusean dauden erlazio kopurua

Beraz, kontuan izan behar dugu ondoren ikusiko ditugun emaitzetan agertuko diren ehunekoak 11 Taulan adierazten diren erlazio kopuruen araberakoak direla. Adibidez, % 50 agertzen bada KAUSA_NS erlazioetan, horrek 22 (44/2) aldiz gertatu dela esan nahi du, aldiz, EBALUAZIOA_NS erlazioan azaltzen bazaigu, horrek 66 (132/2) aldiz agertzen dela pentsatuko dugu. Adierazgarria da baita ere, ez dagoela S-N norantzako ONDORIOA erlaziorik eta orokorrean S-N norantzako erlazioak askoz gutxiago direla.

Horretaz gain, argi izan behar dugu 11 Taulako BESTEAK1 eta BESTEAK2 taldeak bereiztea. BESTEAK1 erlazio taldeak KAUSA, ONDORIOA eta HELBURUA ezik, beste erlazio guztiak hartzen ditu barnean. BESTEAK2 taldean, aldiz, EBALUAZIOA eta INTERPRETAZIOA ezik, beste erlazio guztiak sartzen dira. RST-ko erlazio guzti hauek 2 Taulan ikus ditzakegu.

4.1 Kausa taldeko erlazioen bereizketa elkarren artean

Garatu dugun bilatzailea martxan jarrita, lehen atazan KAUSA, ONDORIOA eta HELBURUA bereizi ditugu.

4.1.1 KAUSAren patroiak

Horretarako, *bilaketa_NS.pl* programaren bidez, KAUSA erlazioko eta N-S norantzako patroiak bilatzeko eskatu diogu erlazio bakoitzeko. Honen emaitzak 12 Taulan ikus ditzakegu:

KAUSAREN PATROIAK (N-S)		**Patroi parekatzea (%)**		
Nukleoa	Satelitea	Kausa	Ondorioa	Helburua
∅	M bait-	15,91	3,70	4,55
∅	E bait-	11,36	1,23	0,91
M interesgarri	E bait-	2,27	0,00	0,00
∅	B izan ere	9,09	0,00	0,00
∅	E eraginda	2,27	0,00	0,00
∅	M eraginda	2,27	0,00	0,00
∅	M arrazoia	2,27	1,23	0,00
∅	MM eta arrazoi ... horretarako	2,27	0,00	0,91
E arrazoiengatik	MM -gatik ... -lako ... -gatik	2,27	0,00	0,00
∅	M -elakoan	2,27	0,00	0,00
∅	B -en erroan	2,27	0,00	0,00
∅	M -en ondorioz	2,27	1,23	0,91
∅	E -rekin bat dator	2,27	0,00	0,00

Taula 12: Kausa taldeko erlazioen patroi parekatzea KAUSA N-S seinaleetan

Espero bezala eta 12 Taulan ikus dezakegun moduan, KAUSA identifikatzeko erabili nahi ditugun patroi guztiek KAUSA erlazioetan asmatzea lortzen dute eta patroi bakoitza gutxienez behin agertzen da. Horretaz gain, *bait-* patroiak pisu handia duela ikusten da, satelitearen bukaeran edo erdian KAUSA erlazioen % 27,27tan agertzen baita. Hala ere, bi hauen artean konparatzen badugu, ikus dezakegu seinale hau unitatearen amaieran gehiagotan agertzen dela ONDORIO eta HELBURU erlazioetan eta beraz, *bait-* unitatearen erdian agertzeak KAUSAren seinale fidagarria izateko aukera txikitzen digu.

Nukleartasun ezberdinari dagokionez, S-N norantzako patroiak ere begiratu ditugu, oraingoan *bilaketa_SN.pl* programa erabilita. Programa horren exekuzioak ateratzen dituen emaitzak 13 Taulan azter ditzakegu. Taula honetan ONDORIOA erlazioaren daturik ez daukagu, izan ere 11 Taulan ikusi dugun moduan, lehenengo satelitea eta gero nukleoa datorren S-N ondorio erlaziorik ez dago bertan.

KAUSAREN PATROIAK (S-N)		**Patroi parekatzea (%)**	
Satelitea	Nukleoa	Kausa	Helburua
E -nez	∅	8,51	0,00
M -nez	∅	10,64	0,00
E -en eraginez	∅	0,00	0,00
E -nez gero	∅	2,13	0,00
E -nez gero	B horretarako	0,00	0,00
M -nez gero	B	2,13	0,00
E -ela eta	∅	14,89	0,00
E -ela bide	M efektu	2,13	0,00
E -ela tarteko	B izan ere	0,00	0,00
E -lako	∅	2,13	0,00
M emaitza	B eta	2,13	0,00
M emaitza	∅	4,26	0,00
E bait-	B horren ondorioz	0,00	0,00
∅	B horren ondorioz	2,13	0,00
E -gatik	∅	2,13	0,00
E -teagatik	∅	0,00	0,00
E eragile [ADL]	B honegatik	2,13	0,00
∅	B horregatik	2,13	0,00
∅	MM horrek ... ekar-	2,13	0,00

Taula 13: Kausa taldeko erlazioen patroi parekatzea KAUSA S-N seinaleetan

Emaitzei begiratuz gero, argi ikus dezakegu HELBURUA erlazioekin ez dela inoiz parekatzen KAUSA S-N patroi bakar batek ere, eta bereizketa hau egiterakoan teorian % 100eko fidagarritasuna du. Hala ere, kontuan izan behar dugu patroi bakoitzaren maiztasuna KAUSA erlazioetan eta 13 Taulan ageri den bezala, satelitearen unitatearen amaieran edo erdian *-nez* atzizkiarekin hitzak daudenean —hots, unitatearen hasieran agertzen ez denean— KAUSA S-N erlazio guztien % 19,15ean agertzen dela. Gainera, *-ela eta* seinalearekin amaitzen denean ere (% 10etik gorako maiztasuna), bi patroi hauek KAUSA S-N erlaziorako patroi esanguratsua izateko aukerak handitzen ditu. Honez gain, kontuan izan behar dugu badaudela teorian KAUSA erlazioaren patroiak direnak (*RhetDB* tresnak hala etiketatu ditu) baina hemen patroiak parekatzean asmatzerik izan ez dituenak[16].

[16]Honen arrazoia etengabeko datu-base aldaketa da eta nahiz eta *RhetDB* programak patroi hauek etiketatu, ez da corpus guztiz berdina erabili eta horregatik eskuratzen ditu gure bilatzaileak patroi parekatze batzuetan hutsegiteak.

4.1.2 ONDORIOAren patroiak

KAUSA erlazioa aztertuta, ONDORIOA aztertuko dugu, eta erlazio honen alderdirik esanguratsuena, erlazio guztiak norantza berberekoak direla da; izan ere, zuhaitzaren ezkerrean nukleoa dago corpusean ditugun 81 erlazioetan eta, beraz, ez dago S-N norantzako ONDORIO erlazio bakar bat ere. Erlazio honen patroiak 14 eta 15 Tauletan ikus ditzakegu:

ONDORIOAREN PATROIAK (N-S)		**Patroi parekatzea (%)**		
Nukleoa	Satelitea	Kausa	Ondorioa	Helburua
∅	B ondorioz	0,00	2,47	0,00
∅	MM ondorio ... -ri begira	0,00	1,23	0,00
∅	B ondorioa	0,00	1,23	0,00
∅	B -en ondorioz	0,00	1,23	0,00
∅	B eta ... -en ondorioz	0,00	0,00	0,00
∅	B eta ondorioz	0,00	1,23	0,00
∅	MM era honetan ... lortu	0,00	1,23	0,91
m aztertu	MM erakusten ... lortu ... eragin	0,00	0,00	0,00
∅	MM hori ... lortuz gero	0,00	1,23	0,00
∅	MM horrela ... lortu	0,00	1,23	0,00
∅	M lortu	2,27	11,11	4,55
∅	B lortutako emaitza	0,00	1,23	0,00
∅	MM eta emaitzak ... lortu	0,00	0,00	0,00
M emaitza	∅	2,27	3,70	0,91
∅	M emaitza	0,00	7,41	3,64
∅	B emaitza	0,00	3,70	0,00
∅	MM emaitza ... ... (erdietsi/eskuratu/lortu)	0,00	4,94	1,82
∅	MM -en emaitzak ... baieztatu ... ... eta ... prebalentzia ... erakutsi	0,00	0,00	0,00
∅	E sortuz	0,00	0,00	0,00
∅	M sortuz	2.27	2,47	0,00
∅	E -tuz	0,00	2,47	0,00
∅	M dakar	0,00	1,23	0,00
∅	M ekar bait-	0,00	1,23	0,00
∅	MM bada ... ekarri	0,00	1,23	0,00
∅	B eta	2,27	16,05	1,82
∅	B eta horrela	0,00	1,23	0,00
∅	B horrela bada	0,00	0,00	0,00
∅	B honela	0,00	1,23	0,00
∅	MM eta ... aurkitu	0,00	2,47	0,00
∅	M aurkitu	0,00	4,94	0,00
∅	MM eta ... esan nahi [ADL]	0,00	1,23	0,00
∅	M inplikatzen	0,00	2,47	0,00
∅	MM datuek ... adierazten	0,00	1,23	0,00
∅	B beraz	0,00	2,47	0,00
∅	M beraz	0,00	4,94	0,91

Taula 14: Kausa taldeko erlazioen patroi parekatzea ONDORIOA N-S seinaleetan (1)

ONDORIOAREN PATROIAK (N-S)		**Patroi parekatzea (%)**		
Nukleoa	Satelitea	Kausa	Ondorioa	Helburua
∅	E -larik	0,00	2,47	0,91
∅	M -lako	11,36	12,35	4,55
∅	B hau dela eta	0,00	1,23	0,00
∅	B horrenbestez	0,00	1,23	0,00
∅	E orduan	0,00	1,23	0,00
∅	MM aldi berean ... sortzen	0,00	1,23	0,00
∅	M frogatu denez	0,00	1,23	0,00
∅	M eragiteaz gain	0,00	1,23	0,00
∅	MM hartara ... eragina	0,00	0,00	0,00
∅	E -raziz	0,00	2,47	0,00
∅	B [DET] artean	0,00	1,23	0,00
∅	M korrelazioan [ADI]	0,00	2,47	0,00
∅	M -en bitartez jakin	0,00	1,23	0,00
∅	M areagotu	0,00	0,00	0,00

Taula 15: Kausa taldeko erlazioen patroi parekatzea ONDORIOA N-S seinaleetan (2)

Bi taula hauetan ikus dezakegun lehen kontua da patroi kopurua handia dela beste kausa taldeko erlazioekin konparatzen badugu. Emaitza hauek begiratuz gero, honakoa esan dezakegu:

1. 14 Taulan ohar gaitezke *lortu* seinalea satelitearen erdian agertzen denean 81 erlazio horietatik % 11,11 kasutan gertatzen dela. Hala ere, kontuz ibili behar gara; izan ere, HELBURUA erlazioan seinaleek fidagarritasun gutxiago izan ohi dute, unitatearen luzera handia denean, satelite hori edonon agertzeko arriskua handiagotzen delako. Gauza bera gertatzen da *emaitza* erdian dagoenean edota 15 Taulako *-lako* atzizkia daukaten hitzekin, ONDORIOA ez den beste erlazio mota batzuetan ere ager daitekeelako.

2. Hau ikusita, kausa taldeko KAUSA eta HELBURUA erlazioekin ez bezala, ONDORIOAn seinale esanguratsuak aurkitzea zailagoa dela esan dezakegu anbiguotasun handiagoa baitago, eta emaitzei begiratuz gero patroi ziurrak aurkitzea zaila da.

3. Nahiz eta kausa taldean nahiko ondo bereizi ONDORIOA, beste erlazioekin sartzean arazoak emango dizkigu oso hitz errepikatua —*stop word*— baita *eta* hitza euskara hizkuntzan eta esaterako SEKUENTZIA erlazioan asko erabiltzen da *eta* juntagailua.

4.1.3 HELBURUAren patroiak

Atal honetan berriz kausa taldeko beste bi erlazioekin bezala, HELBURUA erlazioaren patroiak parekatzearen emaitzak aztertuko ditugu. Horretarako, berriro N-S norantzako kausa taldeko erlazioak hartu eta gure bilatzailea martxan jarrita, azkenean, 12 Taulako patroi-frekuentziak aurki ditzakegu.

HELBURUAREN PATROIAK (N-S)		**Patroi parekatzea (%)**		
Nukleoa	Satelitea	Kausa	Ondorioa	Helburua
∅	E -tz?eko	2,27	0,00	49,09
∅	M -tzeko	0,00	11,11	23,64
M -tzeko	∅	9,09	16,05	19,09
∅	E lortzeko	0,00	0,00	1,82
∅	MM helburu ... -tzea	0,00	1,23	8,18
∅	M helburu	0,00	2,47	11,82
∅	MM helburua ... -tea ... ... tzea	0,00	0,00	0,91
∅	E -tea ... helburua	0,00	1,23	2,73
∅	MM -tzea ... helburu ... ... -tea	0,00	0,00	0,91
∅	MM -tzeko ... helburu ... ... lortzeko	0,00	0,00	0,00
∅	M -tzea	4,55	11,11	12,73
∅	MM -tzea ... -tzeko	0,00	0,00	4,55
∅	E -tzeko asmoz	0,00	0,00	6,36
E -tzeko asmoz	∅	0,00	0,00	1,82
∅	MM asmoa ... -tzea	0,00	0,00	0,91
∅	E (dezagun/dezaten)	0,00	1,23	1,82
∅	M daitezen	2,27	0,00	1,82
∅	E burutu nahi izan [ADL]	0,00	0,00	0,91
∅	E [ADI] nahi [ADL]	0,00	0,00	0,91
∅	M betebehar	2,27	0,00	0,91
∅	M genuke	0,00	0,00	0,91

Taula 16: Kausa taldeko erlazioen patroi parekatzea HELBURUA N-S seinaleetan

Iruskieta et al -ek (2016) esaten duten bezala HELBURUA erlazioa identifikatzeko patroirik onena satelitearen amaieran *-tzeko* edo *-teko* atzizkia daukaten hitzak agertzen direnean da, HELBURU erlazio guztien ia erdietan gertatzen baita kasu hori. Horretaz gain, *-tzeko*, *-tzea* edota *helburu* seinaleak erdian agertzen badira, HELBURUA erlazioetan askotan agertzen dira, baina ONDORIOAn gertatzen zen bezala, kasu honetan ere unitatearen erdian dauden patroiak beste erlazioetan gertatzeko probabilitatea handiagoa da eta, beraz, ez dira oso seinale fidagarriak.

Aldiz, nukleartasuna aldatzen badugu eta S-N norantzari begiratzen badiogu, antzeko seinaleak ikus ditzakegu 17 Taulan ageri den bezala[17].

HELBURUAREN PATROIAK (S-N)		**Patroi parekatzea (%)**	
Satelitea	Nukleoa	Kausa	Helburua
E -t[z]eko	∅	2,13	65,71
E -tzeko	MM helburu ..... lortzeko	0,00	0,00
MM -teko ... -teko ... -tzeko	∅	0,00	0,00
E -t[z]eko helburuarekin	∅	0,00	8,57
E -t[z]eko asmoz	∅	0,00	2,86
E -tzeko asmoarekin	∅	0,00	0,00
E helburuak lortzeko	∅	0,00	0,00
B xede hori iristeko	∅	0,00	2,86
E -tera	∅	0,00	0,00
E -tu nahian	∅	0,00	2,86
E dadin	∅	0,00	2,86

Taula 17: Kausa taldeko erlazioen patroi parekatzea HELBURUA S-N seinaleetan

Kasu honetan ere, N-S norantzaren antzeko ondorioa atera dezakegu eta satelitearen amaieran *-tzeko* edo *-teko* agertzeak HELBURUA erlazioaren oso seinale esangura da, % 65eko maiztasunean gertatzen baita. Horretaz gain, ikus dezakegu, seinale horri *helburuarekin* hitza itsasten badiogu bukaeran kasuen % 8,57ean hori gertatzen dela. Hau honela, argi dago norantza edozein delarik ere, *-tzeko* edo *-teko* hitzak unitatearen amaieran agertzeak HELBURU erlazioa beste erlazioekiko bereizteko laguntza handia eman diezagukeela.

Laburbilduz, kausa taldeko hiru erlazioen patroi esanguratsuak lortzen saiatu gara beraien artean ezberdinduz. Ikusi dugunez, KAUSA eta HELBURUA erlazioak bereizteko patroi esanguratsuak egon daitezke, baina ONDORIOAren kasuan ez da lan erraza erlazio hau identifikatzea.

4.2 Kausa taldeko erlazioen bereizketa beste erlazioekin

Aurreko atalean (Ikus 4.1 Atala) kausa taldeko hiru erlazioak elkarrengandik bereizi ditugu. Errealitatean, ordea, ez da ohikoa KAUSA, ONDORIOA eta HELBURUA erlazioak bakarrik dituzten testuak aurkitzea eta normalean beste erlazio batzuekin nahasten dira. Hortaz, beraien artean bereizteaz gain, beste erlazio hauek guztiekin ere bereizi beharko ditugu kausa taldeko erlazioak identifikatzeko patroi fidagarriak lortu nahi baditugu.

Bereizketa hau egiteko aurreko metodologia bera erabili eta kausa taldeko erlazioen patroiak, kausa taldekoak ez diren 1554 erlazioekin parekatu ditugu. Metodo honek, ordea,

[17]Gogoratu norantza honetako ONDORIO erlaziorik ez dagoela eta horregatik ez dela agertzen erlazio horren daturik.

arazo bat dakarkigu; izan ere, agertzen diren emaitzak ez dira onak. Honen adibidea, 18 Taulan ikus dezakegu.

KAUSAREN PATROIAK (N-S)		**Patroi parekatzea (%)**			
Satelitea	Nukleoa	Kausa	Ondorioa	Helburua	Besteak
0	M bait-	15,91	3,70	4,55	11,56
0	E bait-	11,36	1,23	0,91	2,96
M interesgarri	E bait-	2,27	0,00	0,00	0,00
0	B izan ere	9,09	0,00	0,00	1,24
0	E eraginda	2,27	0,00	0,00	0,00
0	M eraginda	2,27	0,00	0,00	0,48
0	M arrazoia	2,27	1,23	0,00	1,43
0	MM eta arrazoi ... horretarako	2,27	0,00	0,91	0,19
E arrazoiengatik	MM -gatik ... -lako ... -gatik	2,27	0,00	0,00	0,00
0	M -elakoan	2,27	0,00	0,00	0,76
0	B -en erroan	2,27	0,00	0,00	0,00
0	M -en ondorioz	2,27	1,23	0,91	0,86
0	E -rekin bat dator	2,27	0,00	0,00	0,00

Taula 18: KAUSA N-S seinaleak erlazio guztietan

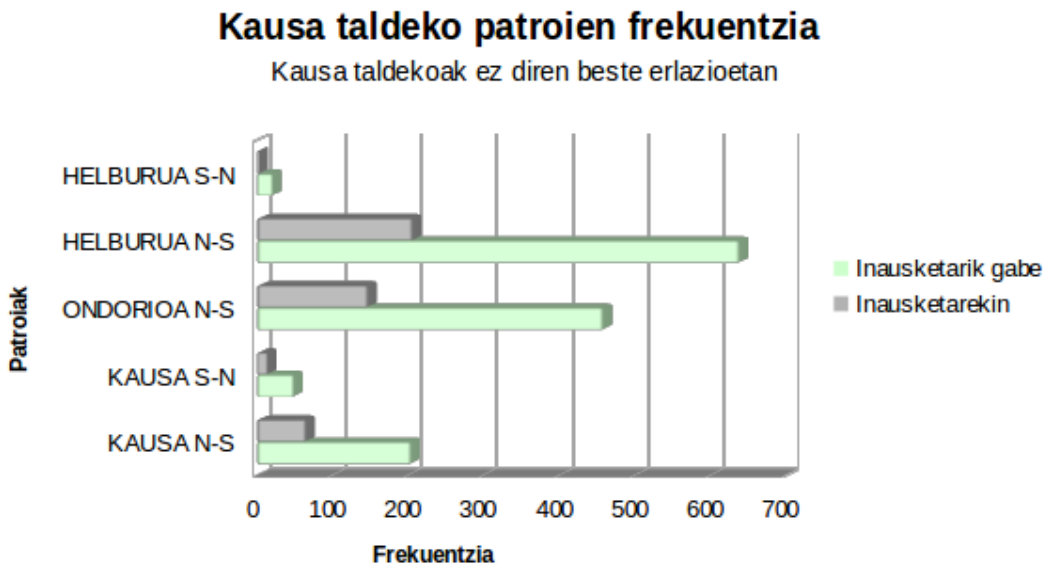
4.1.1 Atalean adierazi dugun bezala, N-S norantzako KAUSA erlazioen patroi esanguratsuenak honakoak direla erabaki dugu:

a) *bait-* aurrizkia daukaten hitzak satelitearen amaieran edo erdian egotea. Bien artean KAUSA S-N erlazioen artean ia % 27ko frekuentzia dute.

b) Satelitearen hasieran *izan ere* agertzea.

Bi patroi hauek beste erlazioetan gertatzen diren ikusten badugu, ikus dezakegu a) kasuan % 14,52n gertatzen dela beste erlazioetan eta horrek patroiari fidagarritasun asko kentzen dio KAUSA erlazioak, patroi horrekin identifikatu nahi baditugu. b) kasuan ere beste erlazioen % 1,24an gertatzen dela ikus dezakegu. Honek badu bere justifikazioa; izan ere, erabili dugun metodologian corpuseko erlazioak aztertu ditugu, baina proiektuaren definizioan (ikus 1 Atala) ikusi bezala, zuhaitz egituratik aterata daude erlazio hauek guztiak eta, ondorioz, errekurtsibitatea egon daiteke, esaterako EDU bat beste baten nukleoa denean, baina era berean, bien artean osatzen den erlazioa beste erlazio baten satelitea izan daiteke. Hori dela eta, metodologian beste teknika bat proposatu dugu, inausketarena hain zuzen. Horretarako, dagoeneko aztertuta ditugun segmentuak ez ditugu analizatuko eta honek errekurtsibitatearen arazoa murrizten digu, emaitza hobeak atereaz.

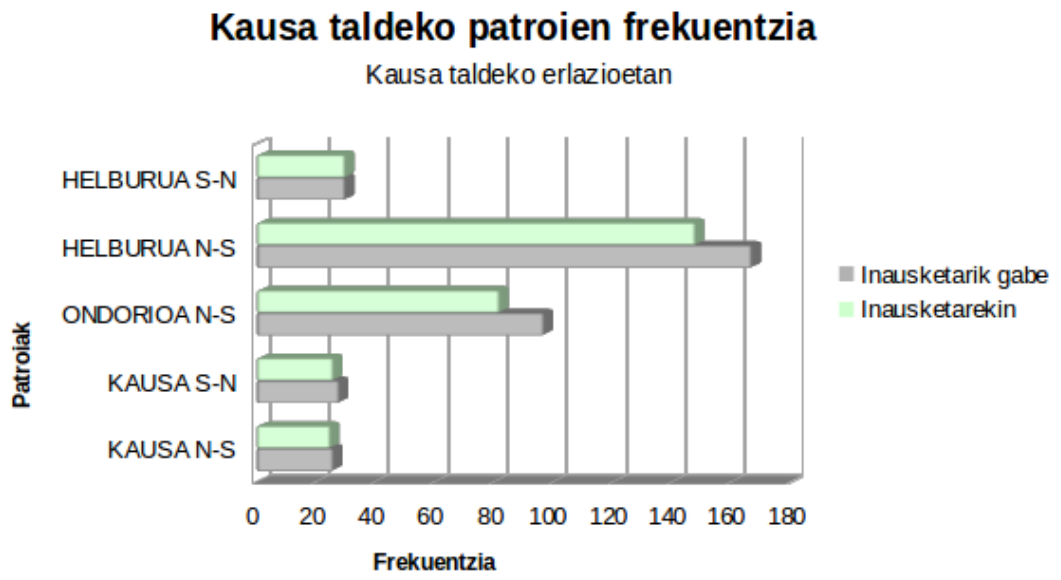
Metodo honekin, kausa taldeko erlazioak besteekin konparatuta bereizketa hobea egiteko aukera ematen digu, baina ez hori bakarrik, kausa taldeko erlazioen artean patroi parekatzeak gutxi edo bat ere ez ditu galtzen, eta, beraz, bereizketa hau egiteko inausketa prozesu hau teknika ona dela erabaki dugu. Horren adibide da, 7 Irudian ikus daitekeen grafiko hau.



Irudia 7: Inausketa prozesuaren garrantzia (1)

Grafikoan, ardatz horizontalean, KAUSA, HELBURUA eta ONDORIOA erlazioak ez diren erlazioetan gertatu den patroi maiztasuna ikus dezakegu kausa taldeko erlazioen patroiak parekatuz. Barra berdez margotuta inausketarik egin gabeko maiztasuna eta barra grisez margotuta aldiz inausketa egindakoak lortzen ditugun emaitzak. Bertan argi ikus daiteke inausketa egin gabe eta inausketa eginda lortzen dugun patroi parekatze positiboen frekuentzia ezberdina dela erlazio eta norantza guztietan, eta horrela, errekurtsibitatearen arazoari aurre egitea lortu dugu.

Teknika horrek, ordea, patroi bakoitzaren frekuentzia txikitu dezake kausa taldeko erlazioetan ere; izan ere, hor ere errekurtsibitatea egon daiteke. Hori dela eta, kausa taldeko erlazioekin konparaketa berdina egin dugu 8 Irudian ikus dezakegun moduan eta erlazio bakoitzeko erlazio horretan seinaleen maiztasun ezberdintasuna zenbatekoa den begiratu dugu.



Irudia 8: Inausketa prozesuaren garrantzia (2)

Ikus daitekeenez, inausketa prozesuak HELBURUA N-S erlazioetan eta ONDORIOA N-S erlazioetan patroi parekatze positiboen maiztasuna txikiagotzen du. Hala ere, maiztasun galtze hori oso txikia da 7 Irudian aztertu dezakegunarekin konparatzen badugu eta patroi esanguratsuenak aukeratzeko orduan ez digu eragiten.

Hau ikusita, ondoren etiketaziorako erabiliko diren patroi esanguratsuenen aukeraketa, inausketaren teknikaren arabera lortutako emaitzetatik eskuratzea erabaki dugu, eta hauek atal honetan datozen tauletan berdez margotuta adieraziko ditugu ondorengo hiru azpiataletan:

a) KAUSA patroiak inausketarekin (4.2.1 Atala).
b) ONDORIO patroiak inausketarekin (4.2.2 Atala).
c) HELBURU patroiak inausketarekin (4.2.3 Atala).
d) Etiketatze prozesuaren emaitzak (4.2.4 Atala)

4.2.1 KAUSA patroiak inausketarekin

3.3.1 irizpide-zerrendari jarraituz, KAUSAren N-S norantzako eta S-N norantzako patroien parekatze eta aukeraketa hauek ikus daitezke 19 eta 20 Tauletan hurrenez hurren.

KAUSAREN PATROIAK (N-S)		**Patroi parekatzea (%)**			
Nukleoa	Satelitea	Kausa	Ondorioa	Helburua	Besteak
0	M bait-	13,64	3,70	2,73	3,53
0	E bait-	11,36	1,23	0,91	1,05
M interesgarri	E bait-	2,27	0,00	0,00	0,00
0	B izan ere	9,09	0,00	0,00	0,67
0	E eraginda	2,27	0,00	0,00	0,00
0	M eraginda	2,27	0,00	0,00	0,10
0	M arrazoia	2,27	1,23	0,00	0,29
0	MM eta arrazoi ... horretarako	2,27	0,00	0,91	0,10
E arrazoiengatik	MM -gatik ... -lako ... -gatik	2,27	0,00	0,00	0,00
0	M -elakoan	2,27	0,00	0,00	0,19
0	B -en erroan	2,27	0,00	0,00	0,00
0	M -en ondorioz	2,27	1,23	0,00	0,29
0	E -rekin bat dator	2,27	0,00	0,00	0,00

Taula 19: KAUSA N-S erlazioarentzat patroiak inausketarekin

19 Taulan KAUSA N-S erlazioarentzat aukeratu ditugun patroiak ikus ditzakegu margotuta. Aukeraketa hau egiterako orduan patroi hauek KAUSA erlazioetan frekuentzia handia izatea bilatu da batez ere, baina horrez gain, kausa taldekoak ez diren beste erlazioetan frekuentzia txikia izatea ere nahi dugu. Adibidez, *izan ere* edota *bait-* aurrizkia satelitearen amaieran egotea erabakiorra da aukeraketa egiterako orduan, beste erlazioetan gehienez % 1,23 inguruko frekuentzia soilik daukalako, baina *bait-* seinalea satelitearen erdian dagoenean KAUSA ez den beste erlazioetan, patroi horren frekuentzia handiagotzen da eta hori ezin dugu seinale fidagarritzat hartu.

Horietaz gain, frekuentzia txikiagoa daukaten baina beste erlazioetan agertzen ez diren patroiak ere kontuan hartu ditugu eta ahalik eta patroi zehatzenak direnean (esaterako nukleoaren erdian *arrazoiengatik* hitza eta satelitean *gatik ... -lako ... -gatik* sekuentzia agertzean) patroiak fidagarritzat hartu ditugu, hain patroi zehatzak besteetan aurkitzeko probabilitatea txikia delako.

KAUSAREN PATROIAK (S-N)		Patroi parekatzea (%)		
Satelitea	Nukleoa	Kausa	Helburua	Besteak
E -nez	∅	8,51	0,00	0,99
M -nez	∅	8,51	0,00	1,18
E -en eraginez	∅	0,00	0,00	0,00
E -nez gero	∅	2,13	0,00	0,00
E -nez gero	B horretarako	0,00	0,00	0,00
M -nez gero	B	0,00	0,00	0,00
e -ela eta	∅	14,89	0,00	0,00
E -ela bide	M efektu	2,13	0,00	0,00
E -ela tarteko	B izan ere	0,00	0,00	0,00
E -lako	∅	2,13	0,00	0,00
M emaitza	B eta	2,13	0,00	0,00
M emaitza	∅	4,26	0,00	0,00
E bait-	B horren ondorioz	0,00	0,00	0,00
∅	B horren ondorioz	2,13	0,00	0,00
E -gatik	∅	2,13	0,00	0,20
E -teagatik	∅	0,00	0,00	0,00
E eragile [ADL]	B honegatik	2,13	0,00	0,00
∅	B horregatik	2,13	0,00	0,00
∅	MM horrek ... ekar-	2,13	0,00	0,39

Taula 20: KAUSA S-N erlazioarentzat patroiak inausketarekin

20 Taulako aukeraketei begiratuz gero, ikus dezakegu KAUSA S-N erlazioetan patroirik fidagarriena satelitean *-ela eta* karaktere katearekin amaitzen denean gertatzen dela eta gainera beste inongo erlazioetan ez dago kasu bakar bat ere hori gertatzen dena. Modu bertsuan, *-nez* seinalea satelitearen erdian edota bukaeran egonda seinale fidagarria dirudi. Aukeratutako beste guztiak, aldiz, nahiz eta frekuentzia txikiagoa izan KAUSA S-N erlazioetan, patroi zehatzak eta besteetan agertzen ez zaizkigunak direlako aukeratu ditugu.

4.2.2 ONDORIOA patroiak inausketarekin

Modu berean, ONDORIOA erlazioaren patroien parekatzeak ikus ditzakegu 21 eta 22 Tauletan. Gogoratu ez dagoela ONDORIOA erlaziorik S-N norantzan.

ONDORIOAREN PATROIAK (N-S)		**Patroi parekatzea (%)**			
Nukleoa	Satelitea	Kausa	Ondorioa	Helburua	Besteak
∅	B ondorioz	0,00	2,47	0,00	0,10
∅	MM ondorio ... -ri begira	0,00	1,23	0,00	0,00
∅	B ondorioa	0,00	1,23	0,00	0,00
∅	B -en ondorioz	0,00	1,23	0,00	0,00
∅	B eta ... -en ondorioz	0,00	0,00	0,00	0,00
∅	B eta ondorioz	0,00	1,23	0,00	0,00
∅	MM era honetan ... lortu	0,00	0,00	0,00	0,00
m aztertu	MM erakusten ... lortu ... eragin	0,00	0,00	0,00	0,00
∅	MM hori ... lortuz gero	0,00	1,23	0,00	0,00
∅	MM horrela ... lortu	0,00	1,23	0,00	0,00
∅	M lortu	2,27	8,64	1,82	0,96
∅	B lortutako emaitza	0,00	0,00	0,00	0,00
∅	MM eta emaitzak ... lortu	0,00	0,00	0,00	0,00
M emaitza	∅	2,27	3,70	0,91	0,86
∅	M emaitza	0,00	4,94	2,73	0,86
∅	B emaitza	0,00	3,70	0,00	0,10
∅	MM emaitza ... ... (erdietsi/eskuratu/lortu)	0,00	2,47	0,91	0,10
∅	MM -en emaitzak ... baieztatu ... ... eta ... prebalentzia ... erakutsi	0,00	0,00	0,00	0,00
∅	E sortuz	0,00	0,00	0,00	0,00
∅	M sortuz	2,27	2,47	0,00	0,10
∅	E -tuz	0,00	2,47	0,00	0,76
∅	M dakar	0,00	1,23	0,00	0,19
∅	M ekar bait-	0,00	1,23	0,00	0,00
∅	MM bada ... ekarri	0,00	1,23	0,00	0,00
∅	B eta	2,27	14,81	1,82	2,96
∅	B eta horrela	0,00	1,23	0,00	0,00
∅	B horrela bada	0,00	0,00	0,00	0,00
∅	B honela	0,00	1,23	0,00	0,10
∅	MM eta ... aurkitu	0,00	2,47	0,00	0,67
∅	M aurkitu	0,00	4,94	0,00	0,96
∅	MM eta ... esan nahi [ADL]	0,00	1,23	0,00	0,10
∅	M inplikatzen	0,00	1,23	0,00	0,19
∅	MM datuek ... adierazten	0,00	1,23	0,00	0,00
∅	B beraz	0,00	1,23	0,00	0,00
∅	M beraz	0,00	2,47	0,00	0,48

Taula 21: ONDORIOA N-S erlazioarentzat patroiak inausketarekin (1)

ONDORIOAREN PATROIAK (N-S)		Patroi parekatzea (%)			
Nukleoa	Satelitea	Kausa	Ondorioa	Helburua	Besteak
∅	E -larik	0,00	2,47	0,91	0,48
∅	M -lako	11,36	11,11	4,55	3,34
∅	B hau dela eta	0,00	0,00	0,00	0,00
∅	B horrenbestez	0,00	1,23	0,00	0,00
∅	E orduan	0,00	1,23	0,00	0,10
∅	MM aldi berean ... sortzen	0,00	1,23	0,00	0,10
∅	M frogatu denez	0,00	1,23	0,00	0,00
∅	M eragiteaz gain	0,00	1,23	0,00	0,00
∅	MM hartara ... eragina	0,00	0,00	0,00	0,00
∅	E -raziz	0,00	2,47	0,00	0,10
∅	B [DET] artean	0,00	1,23	0,00	0,10
∅	M korrelazioan [ADI]	0,00	2,47	0,00	0,00
∅	M -en bitartez jakin	0,00	1,23	0,00	0,10
∅	M areagotu	0,00	0,00	0,00	0,29

Taula 22: ONDORIOA N-S erlazioarentzat patroiak inausketarekin (2)

ONDORIOAren kasuan ere patroi fidagarrienak aukeratu eta margotu ditugu 21 eta 22 Tauletan. Kasu honetan, seinale ugari eta anbiguoak daude, eta, beraz, aukeratu ditugun seinale ia gehinetsuenak beste erlazioetan ia agertzen ez direnak eta zehatzak aukeratu ditugu. Hala ere, erlazio honetan badaude hiru patroi-frekuentzia handiagoa duten patroiak:

1. Satelitearen erdian *lortu* daukan patroia. Patroi hau 4.1.2 Atalean azaldu dugu anbiguoa zela, HELBURUA N-S erlazioetan ia % 5eko frekuentziara iristen zelako, baina inausketa eginda eta errekurtsibitatea kenduta, 21 Taulan ikus dezakegu % 1,82ko frekuentziara jaitsi dela eta horregatik patroi honen aukeraketa egitea erabaki dugu.

2. Satelitearen amaieran *eta* agertzen den patroia. Ia % 15eko frekuentziarekin, baina hau ez aukeratzea erabaki dugu. Honen arrazoia, kausa taldeko 3 erlazioetan frekuentzia nahikoa altua daukala da eta, gainera, *eta* hitza euskara hizkuntzan oso erabilia denez, ez da fidagarria honelako seinale bat bere baitan soilik identifikatzeko erabiltzea, nahiz eta hitz hori satelite hasieran agertu behar duen patroiak parekatze positiboa lortzeko.

3. 22 Taulan ikus dezakegu *-lako* atzizkia satelitearen erdian dagoen patroia. ONDORIOAn frekuentzia handia dauka, baina ezin dugu fidagarritzat hartu, beste erlazio guztietan ere frekuentzia handia baitauka, batez ere, KAUSA N-S erlazioetan % 10eko patroi parekatzea baino handiagoa lortzen da.

4.2.3 HELBURUA patroiak inausketarekin

Kausa taldeko erlazioekin amaitzeko, azkenik, HELBURUA erlazioaren emaitzak eta patroi aukeraketak ikus ditzakegu 23 eta 24 Tauletan.

HELBURUAREN PATROIAK (N-S)		**Patroi parekatzea (%)**			
Nukleoa	Satelitea	Kausa	Ondorioa	Helburua	Besteak
∅	E -tz?eko	2,27	0,00	49,09	1,34
∅	M -tzeko	0,00	8,64	20,00	6,59
M -tzeko	∅	9,09	16,05	9,09	6,30
∅	E lortzeko	0,00	0,00	1,82	0,00
∅	MM helburu ... -tzea	0,00	0,00	5,45	0,10
	M helburu	0,00	1,23	9,09	0,38
∅	MM helburua ... -tea ... ... tzea	0,00	1,64	0,91	0,00
∅	MM -tea ... helburua	0,00	1,23	2,73	0,10
∅	MM -tzea ... helburu ... ... -tea	0,00	0,00	0,00	0,00
∅	MM -tzeko ... helburu ... ... lortzeko	0,00	0,00	0,00	0,00
∅	M -tzea	4,55	7,41	10,00	3,63
∅	MM -tzea ... -tzeko	0,00	0,00	3,64	0,48
∅	E -tzeko asmoz	0,00	0,00	5,45	0,00
E -tzeko asmoz	∅	0,00	0,00	0,00	0,19
∅	MM asmoa ... -tzea	0,00	0,00	0,91	0,00
∅	E (dezagun/dezaten)	0,00	1,23	1,82	0,10
∅	M daitezen	0,00	0,00	1,82	0,19
∅	E burutu nahi izan [ADL]	0,00	0,00	0,91	0,10
∅	E [ADI] nahi [ADL]	0,00	0,00	0,91	0,00
∅	M betebehar	2,27	0,00	0,91	0,10
∅	M genuke	0,00	0,00	0,91	0,10

Taula 23: HELBURUA N-S erlazioarentzat patroiak inausketarekin

HELBURUA N-S erlazioan 23 Taulan ikus dezakegu *-tzeko* edo *-teko* atzizkia satelitearen amaieran agertzen bada probabilitate oso handian daude HELBURU erlazio bat izateko eta hori da erlazio honetan daukagun seinalerik esanguratsuena. Horretaz gain, inausketa prozesuak satelitearen erdian *helburu* hitza ONDORIO erlazioetan agertzea gutxitzen duenez, hau ere patroi esanguratsutzat hartzea erabaki dugu. Aukeratu ditugun beste patroiak, aldiz, patroi zehatzagoak dira eta beste erlazioetan gutxi agertzen direnak.

HELBURUAREN PATROIAK (S-N)		Patroi parekatzea (%)		
Satelitea	Nukleoa	Kausa	Helburua	Besteak
E -t[z]eko	∅	2,13	65,71	0,59
E -tzeko	MM helburu ..... lortzeko	0,00	0,00	0,00
MM -teko ... -teko ... -tzeko	∅	0,00	0,00	0,00
E -t[z]eko helburuarekin	∅	0,00	8,57	0,00
E -t[z]eko asmoz	∅	0,00	2,86	0,00
E -tzeko asmoarekin	∅	0,00	0,00	0,00
E helburuak lortzeko	∅	0,00	0,00	0,00
B xede [DET] iristeko	∅	0,00	2,86	0,00
E -tera	∅	0,00	0,00	0,00
E -tu nahian	∅	0,00	2,86	0,00
E dadin	∅	0,00	2,86	0,00

Taula 24: HELBURUA S-N erlazioarentzat patroiak inausketarekin

HELBURUA S-N erlazioaren patroiak inausketa bidez aztertutako emaitzak 24 Taulan ikus ditzakegu. N-S norantzarekin gertatzen den moduan, satelitea *-tzeko* eta *-teko* atzizkiekin amaitzean dago probabilitate handiena HELBURU bezala identifikatzeko. Aukeratu ditugun besteak, HELBURU erlazioetan patroi parekatze positiboak izan dituzten eta beste erlazioetan ez dauden patroiak izan dira.

4.2.4 Erlazio etiketatzailearen emaitzak

19 eta 24 taulen artean berdez margotuta ageri diren patroiak erabilita, *etiketatzailea.pl* programa jarri dugu martxan eta horrela erlazio bakoitzean bere balizko patroiek duten fidagarritasuna aztertu nahi dugu.

Kausa taldeko erlazioen etiketazio prozesuaren ondoren, erlazio bakoitzaren asmatze-tasa begiratu dugu eta hauek 25 Taulan ikus ditzakegu.

Etiketatutako erlazioak			
Patroiak	Zuzen etiketatuta	Erlazioak	Asmatze-tasa
KAUSA N-S	14	44	% 31,82
ONDORIOA N-S	23	81	% 28,40
HELBURUA N-S	81	110	% 73,64
KAUSA S-N	16	44	% 36,36
HELBURUA S-N	29	35	% 82,86

Taula 25: Kausa taldeko erlazioen etiketatze egokiak

25 Taulan ikus dezakegunez, asmatze-tasa txikiena espero bezala ONDORIOA erlazioaren patroiarekin lortu dugu, kausa taldeko erlazioen artean bereizteko zailena berau baita, baina KAUSA erlazioarekin ere nahiko emaitza kaxkarrak lortu ditugu. Hala ere, honen arrazoia izan daiteke aztergai genituen kausa erlazioak urriak direla eta bilaketa prozesuan

ez dugu lortu % 50eko frekuentziarik guztira inongo norantzan. Aldiz, HELBURUA seinalea ondo etiketatu dela esan dezakegu; izan ere, % 70etik gorako asmatze-tasa izan du, N-S norantzan 110etik 81etan detektatuz eta S-N norantzan 35etik 29 alditan detektatuz, emaitza oso onak emanez.

Honetaz gain, ordea, erlazio horiek zenbat aldiz etiketatu diren modu okerrean ere beharrezkoa da ebaluazioa egin ahal izateko. Datu horiek 26 Taulan azter ditzakegu.

Beste erlazioak			
Patroiak	Oker etiketatuta	Erlazioak	Errore-tasa
KAUSA N-S	15	1047(N-S)	% 1,43
ONDORIOA N-S	58	1047(N-S)	% 5,54
HELBURUA N-S	61	1047(N-S)	% 5,83
KAUSA S-N	21	507(S-N)	% 4,14
HELBURUA S-N	11	507(S-N)	% 2,17

Taula 26: Kausa taldeko erlazioen etiketatze okerrak

26 Taulan ikus daiteke guztietan ere % 6ko errore tasara ez dela iristen, baina adierazgarria da, HELBURUA N-S norantzan oso ondo etiketatzen zela uste izanda, HELBURUA ez den 60 erlazioetan modu okerrean etiketatu dela. Kontuan izan behar da kausa taldeko erlazioak ez direnak guztira 1554 direla eta kopuru hori txikitzat har daitekeela, baina horrek adierazten digu HELBURUA N-S 20 kasutik behin gaizki etiketatzen dela, eta hori hobetzea garrantzitsua litzateke. Aldiz, KAUSA N-S eta HELBURUA S-N erlazioek errore tasa txikiagoa dute, % 1,43 eta % 2,17koa direlarik datu zehatzak.

25 eta 26 Taulak ikertuko bagenitu, modu sinple batean lortu genitzake estatistika-neurri hauek; doitasuna (*precision*), estaldura (*recall*) eta f-neurria (*f-score*). Neurri hauek 27 Taulan ikus ditzakegu:

Eraginkortasuna			
Patroiak	Doitasuna	Estaldura	F-neurria
KAUSA N-S	0,48	0,32	0,38
ONDORIOA N-S	0,28	0,28	0,28
HELBURUA N-S	0,57	0,74	0,64
KAUSA S-N	0,43	0,36	0,40
HELBURUA S-N	0,73	0,83	0,77

Taula 27: Kausa taldeko etiketatzaileen estatistika neurriak

27 Taulan ikusten den moduan, doitasun, estaldura eta F-neurririk onenak HELBURUA S-N etiketatzaileak ematen ditu, izan ere, asmatze-tasarik handiena eta errore-tasarik txikiena berak ematen ditu. Ondoren, alderantzizko norantzako HELBURUA N-S etiketatzailea da eraginkorrena eta, beste aldean, ONDORIOA erlazioaren etiketatzailea daukagu, etiketatze desegokiena egiten duena, 0,28ko balioarekin hiru estatistika neurrietan.

Ondorioz, etiketatzailerik onena HELBURUA erlazioarena dela esan dezakegu, izan ere, bilatzaileak patroi esanguratsuak lortzen ditu eta gainera nahikoa ongi bereizten da RST-ko beste erlazioekin. Hori lortzea oso garrantzitsua da gure lanerako, galdera-erantzunetan garrantzi handia daukalako helburua zehazteak. Aldiz, KAUSA eta ONDORIOA identifikatzeko oraindik lana gelditzen da eta beste teknika batzuk erabili beharko dira, esaterako, korreferentziazko patroiak cluster bikoteetan erabiliz (Rutherford eta Xue, 2014).

4.3 EBALUAZIOA eta INTERPRETAZIOAren bereizketa

Azken ataza (ikus 3 Ataza), EBALUAZIOA eta INTERPRETAZIO erlazioak bereiztea izango da gerora sentimendu analisian aplikatzeko. Horretarako, 4.2.1, 4.2.2 eta 4.2.3 Ataletan egin dugun modura, atal honetan ere inausketarekin lortutako emaitzak erakutsiko ditugu eta erlazio bakoitzeko berdez margotuko ditugu EBALUAZIOA eta INTERPRETAZIOA erlazioen patroirik esanguratsuenak aztertu:

a) EBALUAZIO patroiak inausketarekin (4.3.1 Atala).
b) INTERPRETAZIO patroiak inausketarekin (4.3.2 Atala)
c) Etiketatze prozesuaren emaitzak (4.3.3 Atala).

4.3.1 EBALUAZIOAREN patroiak inausketarekin

28 Taulan ikus ditzakegu EBALUAZIOA erlazioaren patroi parekatzeak erlazio horretan, INTERPRETAZIOAN eta beste erlazio guztietan. Margotuta ageri direnak berriz, patroi esanguratsuak direla ebatzi dugu 3.3.1 Atalean ikusi dugun irizpide-zerrendaren arabera.

EBALUAZIOAREN PATROIAK (N-S)		**Patroi parekatzea(%)**		
Nukleoa	Satelitea	Ebaluazioa	Interpretazioa	Besteak
∅	B ezta hurrik eman ere	0,76	0,00	0,00
∅	M zailtasunak eta lorpenak	0,00	0,00	0,00
∅	M garrantzitsua	0,76	0,00	0,18
∅	M kontuan hart-	0,76	0,00	0,18
∅	E zalantzari tokirik utzi gabe	0,76	0,00	0,00
∅	MM beraz ... oso aproposatzat kalifikatuak	0,76	0,00	0,00
∅	M eta [DET] ... abantaila garrantzitsua	0,00	0,00	0,00
∅	M berebiziko garrantzia	0,76	0,00	0,09
∅	M etorkizun handikoak	0,76	0,00	0,09
∅	M oso egokia	0,76	0,00	0,09

Taula 28: EBALUAZIOA N-S seinaleak INTERPRETAZIOA eta beste erlazioetan

EBALUAZIOAren kasuan, patroi esanguratsutzat aukeratu ditugunak, patroiaren parekatze positiboren bat daukaten eta beste erlazioetan agertzen ez direnak izan dira. EBALUAZIOA erlazioan seinale gutxi daude (kontuan izan 132 erlazio daudela N-S norantzakoak) eta beraz, oso frekuentzia txikia daukate 28 Taulan ikusten diren seinaleetan. Hau horrela, seinale hauekin EBALUAZIOA identifikatzea ez dela nahikoa izango iragarri dezakegu.

4.3.2 INTERPRETAZIO patroiak inausketarekin

Beste erlazio guztien antzera, INTERPRETAZIOA erlazioaren patroi parekatzea eta etiketaziorako egin den aukeraketa 29 Taulan ikus dezakegu N-S norantzarako eta 30 Taulan S-N norantzarako. Ohartu S-N norantzan interpretazio erlazio bakarra existitzen dela corpusean, beraz, seinalea ere bakarra eskuratu da.

INTERPRETAZIOAREN PATROIAK (N-S)		**Patroi parekatzea(%)**		
Nukleoa	Satelitea	Ebaluazioa	Interpretazioa	Besteak
∅	M emaitzek iradokitzen [ADL]	0,00	1,64	0,00
∅	B ondorioz, ez dirudi	0,00	1,64	0,00
∅	B eta [DET] kontuan hartzeko[a][k]	0,00	0,00	0,00
∅	B lortutako emaitzek baieztatzen [ADL]	0,00	1,64	0,00
∅	MM erdietsiriko emaitzen arabera ... uste [ADL]	0,00	1,64	0,00
M emaitza positiboak	M irizpideen arabera kasu [DET]	0,00	0,00	0,00
∅	MM -ri esker ... arrakasta irizpidetzat ... nabarmen [ADI]	0,00	1.64	0,00
∅	B eta [IZE] irizpidetzat jo	0,00	0,00	0,00
E balorazioari dagozkie	B hobetu beharreko	0,00	0,00	0,00
∅	MM eta DET ... emaitzen alderaketa zailtzen du	0,00	1,64	0,00
∅	E izan daite[z]ke	0,00	1,64	0,00
∅	E bail-	0,00	1,64	0,00
∅	B [DET] esan nahi [ADL]	0,00	0,00	0,00
∅	M interpretatu	0,00	0,00	0,00
∅	B eta [DET] garrantzi handikoa	0,00	1,64	0,00
∅	M emaitzak osatzen eta ulertzen [ADL]	0,00	1,64	0,00
∅	MM era honetan ... ... lor (daiteke/dezakegu	0,00	1,64	0,00

Taula 29: INTERPRETAZIOA N-S seinaleak EBALUAZIOAn eta beste erlazioetan

INTERPRETAZIOA N-S erlazioak identifikatzeko ere seinale gutxi ditugunez, 29 Taulan patroi parekatze positiboa eta beste erlazioetan aurkitzen ez diren patroiak aukeratu ditugu.

INTERPRETAZIOAREN PATROIAK (S-N)		Patroi parekatzea(%)		
Satelitea	Nukleoa	Ebaluazioa	Interpretazioa	Besteak
B egia da	∅	0,00	100,00	0,00
M oso goitik	∅	0,00	0,00	0,00
B eta egia esan ez luke[te]	∅	0,00	0,00	0,00

Taula 30: INTERPRETAZIOA S-N seinaleak EBALUAZIOAn eta beste erlazioetan

INTERPRETAZIOA S-N erlazioetan berriz, norantza horretako erlazio bakarra daukagunez, satelitearen hasieran *egia da* agertzen den patroia bakarrik aukeratu dugu.

4.3.3 Erlazio etiketatzailearen emaitzak

Behin patroiak identifikatu ditugunean, etiketatzailea jarriko dugu martxan EBALUAZIOA eta INTERPRETAZIOA erlazioak identifikatzeko eta ebaluatzeko. Horretarako, bi etiketatzaile sortu ditugu eta asmatze-tasak eskuratu ditugu etiketatze egokiak eta erlazio kopurua begiratuta. Asmatze-tasa hauek 31 Taulan ikus ditzakegu.

Etiketatutako erlazioak			
Erlazioa	Zuzen etiketatuta	Erlazioak	Asmatze-tasa
EBALUAZIOA N-S	7	132	% 5,30
INTERPRETAZIOA N-S	14	61	% 22,95
INTERPRETAZIOA S-N	1	1	% 100,00

Taula 31: EBALUAZIOA eta INTERPRETAZIOA erlazioen etiketatze egokiak

31 Taulan ikus dezakegunez, EBALUAZIOA erlazioa identifikatzea ez da batere erraza eskuragarri ditugun patroiekin (% 5,3ko asmatze-tasa soilik); izan ere *RhetDB* tresnak ematen dizkigun seinaleak gutxi eta ez oso esanguratsuak dira eta, beraz, etiketatze prozesua zaila da. INTERPRETAZIOA erlazioaren kasuan emaitza zerbait hobeak lortzen ditu, baina N-S ordenan ez da iristen % 23ko asmatze-tasara, eta nahiz eta S-N ordenan guztiz asmatu, hori erlazio bakar bat dagoelako soilik gertatzen da eta, beraz, ez da hain esanguratsua.

Etiketatze prozesu honetan huts egin dutenak kontatuko bagenitu berriz, 32 Taulako emaitzak eskuratuko ditugu.

Etiketazio emaitzak			
Erlazioa	Oker etiketatuta	Erlazioak	Errore-tasa
EBALUAZIOA N-S	12	1089(N-S)	% 1,10
INTERPRETAZIOA N-S	11	1089(N-S)	% 1,01
INTERPRETAZIOA S-N	0	586(S-N)	% 0,00

Taula 32: EBALUAZIOA eta INTERPRETAZIOA erlazioen etiketatze okerrak

Etiketatze prozesu okerrak begiratzen baditugu aldiz, errore-tasa oso txikiak agertzen zaizkigu; beraz, argi dago erabiltzen ari garen patroi zerrenda nahiz eta ez izan oso eraginkorra EBALUAZIOA eta INTERPRETAZIOA ondo etiketatzeko, nahikoa etiketatzaile ziurra dela eta oso gutxitan etiketatzen duela oker. Gainera, gogoratu behar gara ataza honen eginkizuna EBALUAZIOA eta INTERPRETAZIO batez ere elkarren artean bereiztea zela eta hori oso ondo egiten du, izan ere, ez dago etiketatzailerik EBALUAZIOA etiketatzen duena INTERPRETAZIOAN eta alderantziz.

Etiketatzaile hau estatistikoki baloratzeko, 33 Taulan ikus ditzakegu estatistika neurriak:

Eraginkortasuna			
Patroiak	Doitasuna	Estaldura	F-neurria
EBALUAZIOA N-S	0,37	0,05	0,09
INTERPRETAZIOA N-S	0,56	0,23	0,33
INTERPRETAZIOA S-N	1,00	1,00	1,00

Taula 33: EBALUAZIOA eta INTERPRETAZIOA erlazioen estatistika neurriak

Estatistika neurri hauek begiratuz gero, oso argi dago EBALUAZIOA erlazioaren etiketatzailea ez dela batere eraginkorra, 0,05eko f-neurria lortzen baitu. Aldiz, INTERPRETAZIOA N-S erlazioen kasuan hobea da neurri hori, baina, hala ere, ez da laurden batera iristen. Azkenik, INTERPRETAZIOA S-N etiketatzailea daukagu eta nahiz eta honek f-neurri maximoa izan, aurrez esan dugun moduan, ez da esanguratsua.

5 Ondorioak eta etorkizuneko lanak

Atal honetan lanaren ondorio orokorra eta etorkizunean egin ditzakegun honen inguruko lanak azalduko ditugu, 5.1 eta 5.2 Ataletan.

5.1 Ondorio orokorrak

Lan honetan egin diren atazak amaituta, honako ondorioak atera ditugu:

1: KAUSA, ONDORIOA eta HELBURUA beraien artean bereizterako orduan ikusi dugu nahiko emaitza onak ateratzen direla, batez ere, HELBURUA eta KAUSA erlazioekin. ONDORIOAren kasuan oraindik eta gehiago zehaztu edo bilatu behar ditugu bere patroiak, baina bertan ere badira seinale esanguratsuak.

2: KAUSA, ONDORIOA eta HELBURUA beste erlazio guztiekin alderatzen badugu, ataza asko zailtzen da. Nahiz eta demostratu erlazioen arteko errekurtsibitatea kenduta emaitzak hobetzen direla, beste erlazioen kopurua askoz ere handiagoa da kausa taldekoekin alderatuta eta horrek jada erroreak egoteko probabilitatea asko handitzen du. Etiketatzaileak erabili ditugunean, ikusi dugu kausa taldeko erlazioak etiketatzea

ez dela batere erraza beste erlazioak sartzen direnean. Hala eta guztiz ere, galdera-erantzun sistemetarako oso baliagarriak diren HELBURUA erlazioa ondo bereizten da beste erlazio guztiekin ere.

3: EBALUAZIOA eta INTERPRETAZIOA erlazioak beraien artean ondo bereizten dira, baina berriro ere corpus falta nabarmena ikusi dugu hemen. Gainera, lortu ditugun EBALUAZIOAren patroiak eskasak dira eta horregatik, etiketatzerako orduan, eman dituen emaitzak ez dira onak. Hala ere, bi erlazio hauetan lortutako seinaleak fidagarriak direla ikusi dugu, oso errore-tasa txikia izan baitute biek ala biek.

Kontuan izan behar dugu RST-ren erlazio erretorikoen eta orokorrean diskurtsoaren gaia nahiko berria dela eta oraindik lan asko egin behar dela arlo honetan. Diskurtso segmentatzailea, unitate zentralaren azterketa eta erlazioen detekzio automatikoan lan ugari egin dezakegu oraindik ere, batez ere euskarazko corpusekin ari garenean lanean.

5.2 Etorkizuneko lanak

Lan honek, hala ere, bide luzea dauka oraindik. Testu kopuru txikiak erabili ditugunez, garrantzitsua litzateke corpus tamaina handitzea sistema hobeto probatzeko eta eraginkortasuna ikusteko. Horretaz gain, ondorengo ataza hauek ere egin ditzakegu koherentziazko erlazioen inguruan:

i) Diskurtso-markatzaileak (juntagailu eta lokailuak) kendu eta identifikatu kausa taldekoak: KAUSA, ONDORIOA eta HELBURUA, elkarren artean bereiziz.

ii) RS3 zuhaitzetan oinarrituta nola egin daitekeen galdera-erantzun esanguratsuak aztertu.

iii) Diskurtso segmentatzailea (Iruskieta eta Zapirain, 2015), unitate zentralaren detektatzailea (Iruskieta et al., 2015) eta kausa-ondorioa-helburua detektatzailea batu eta saiatu galde-erantzun interesgarriak egiten. Euskaraz dauden galdera-erantzun sistema hobetzeko asmoz edo diskurtsoko informazioa sartzeko asmoz.

iv) Diskurtso segmentatzailea (Iruskieta eta Zapirain, 2015), unitate zentralaren detektatzailea eta ebaluazioa-interpretazioa detektatzailea batu eta saiatu testuen polaritatea asmatzen. *Elhuyar Elixa* elementuetan oinarrituriko polaritate tresna hobetzeko asmoz edo diskurtsoko informazioa sartzeko asmoz.

6 Eranskinak

6.1 Bilatzailerako erabiltzen den testuen formatua

bilatzailea_NS.pl eta *bilatzailea_SN.pl* programak erabiltzeko parametro bezala fitxategi bat pasa behar zaio formatu honetan, erlazio hauetan bilatu dezan.

```
<column name="segment_id"> Azterlan honek haur oinaren eta oinzolako
gangaren hazkundea baloratzen du 4 eta 6 urte arteko haurren
artean.</column>

<column name="segment_parent"> Adin horretan hasten da aldaketa hormonalek
agintzen duten garapena.</column>

<column name="segment_id"> Azterlan honek haur oinaren eta oinzolako
gangaren hazkundea baloratzen du 4 eta 6 urte arteko haurren artean.
</column>

<column name="segment_parent"> Adin horretan hasten da aldaketa hormonalek
 agintzen duten garapena.</column>

<column name="segment_id"> izan ere, gaixoen bizi-kalitatea benetan
kaskarra zen.</column>

<column name="segment_parent"> pisu galera desegokia eta etengabeko
gonbitoak zirela tarteko,</column>

<column name="segment_id"> Hona hemen kalkaneo-stop teknika erabiliz
gure zerbitzuan ebakuntza egin diegun haurrek izandako emaitzak.</column>

<column name="segment_parent"> oin malgua izateagatik</column>

<column name="segment_id"> Proportzioak gora egiten du T1c tumoreetan;
</column>

<column name="segment_parent"> izan ere, hirutik bat hautematen da
horrelakoetan.</column>
```

6.2 Forma eta kategoria zerrenda

Word.xml fitxategia erauztean honakoa lortzen dugu, *kategoriak* izeneko fitxategi batean gordez.

```
<column name="forma">Birikietako</column>

<column name="kat">IZE</column>

<column name="forma">tuberkulosiak</column>

<column name="kat">IZE</column>

<column name="forma">tratatzeko</column>

<column name="kat">ADI</column>

<column name="forma">kolapsoterapiapean</column>

<column name="kat">ADJ</column>

<column name="forma">dauden</column>

<column name="kat">ADT</column>

<column name="forma">pazienteen</column>

<column name="kat">ADJ</column>
```

6.3 Bilatzailerako erabiltzen den patroien formatua

bilatzailea_NS.pl eta *bilatzailea_SN.pl* programak erabiltzeko parametro bezala fitxategi bat pasa behar zaio patroi zerrendarekin.

N-S norantzako kasuetan, lehenik nukleoaren patroia eta bigarrenik satelitearena '*' banatzailearen artean banatuz. KAUSA N-S erlazioaren bilaketarako patroien adibidea ikus dezakegu:

```
0 * M bait-
0 * E bait-
M interesgarri * E bait-
0 * B izan ere
0 * E eraginda
0 * M eraginda
0 * M arrazoia
0 * MM eta arrazoi ... horretarako
E arrazoiengatik * MM -gatik ... -lako ... -gatik
0 * M -elakoan
0 * B -en erroan
0 * M -en ondorioz
0 * E -rekin bat dator
```

Aldiz, S-N norantzan ere nahiz eta formatua berdina izan, kontuan izan behar dugu lehendabizi satelitearen patroia agertuko dela eta gero nukleoarena '*' banatzaile batez banatuz, kasu honetan ere. Ondoren, KAUSA S-N patroi bilaketarako erabili den adibidea ikus dezakegu:

```
E -nez * 0
M -nez * 0
E -en eraginez * 0
E -nez gero * 0
E -nez gero * B horretarako
M -nez gero * B horretarako
E -ela eta * 0
E -ela bide * M efektu
E -ela tarteko * B izan ere
E -lako * 0
M emaitza * B eta
M emaitza * 0
E bait- * B horren ondorioz
0 * B horren ondorioz
E -gatik * 0
E -teagatik * 0
E eragile [ADL] * B honegatik
0 * B horregatik
0 * MM horrek ... ekar-
```

6.4 Bilatzaileko emaitzaren irteera estandarra

bilatzailea_NS.pl eta *bilatzailea_SN.pl* programek patroi bakoitzeko zenbat aldiz gertatu den parekatze positiboa kontatzen du eta zein segmentu bikotetan gertatu den ere erakusten du. Hurrengo adibidean ikus dezakegu irteera estandarrean erakusten duen emaitzaren zati bat, KAUSA N-S erlazioaren patroiak HELBURUA N-S testuetan identifikatzean.

```
ANALIZATUTAKO TESTU KOPURUA: 110
**********************
NUKLEO PATROIA: 0 - SATELITE PATROIA:m bait-
**********************
NUKLEOAN:
PATROIA : 0
SEKUENTZIA :
--------------------------------------
SATELITEAN:
PATROIA : m bait-
SEKUENTZIA : helburuak ere bi jomuga zituen: alde batetik, lanbidearen
errealitaterako 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 zalantzari tokirik utzi gabe.
izan ere, itzulpengintza eta interpretaritzako ikasketetan ez dugu
diziplinartekotasun hori baztertu baina, barneratu ere ez dugu erabat
egin (behin betiko indartzeko arrazoi garbia, beharbada).
^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
NUKLEOAN:
PATROIA : 0
SEKUENTZIA :
--------------------------------------
SATELITEAN:
PATROIA : m bait-
SEKUENTZIA : azken hori dela eta, gure asmoa okerreko uste bat betiko
zuzentzea da, ikasleen artean oso zabalduta dagoenez, karreran zehar
ikasi, barneratu eta gainditu behar dituzten ikasgaiak bloke isolatuak
baitira, bihar-etziko lanbidearekin zerikusirik ez dutenak. azken
batean, halako gogoeta teoriko eta praktiko bat egin nahi izan dugu
dokumentazioaren, terminologiaren, fraseologiaren eta
itzulpengintzaren artean behintzat -bestetik ez bada ere nabari den
erabateko loturaren inguruan eta horrelako esperientziek -gure uste
apalean- espainian gaur egun eskura ditugun itzulpengintza-ikasketetan
```

```
eragiten dituzten ondorioen gainean, ondorio akademiko zein
irakaskuntza-mailakoen gainean.

NUKLEOAN:
PATROIA : 0
SEKUENTZIA : lubaki horretakoekin bat egiten duen inor derrigorrezko
enpatia ariketa bat egin beharrean dago
--------------------------------------
SATELITEAN:
PATROIA : m bait-
SEKUENTZIA : garroren testuak dituen bertute literario nabarmenak,
bikainak, goxatzeko asmotan,ikaragarri ondo idatzita baitago
eleberria,hainbat pasartetan hunkigarria ere iristeraino.

**********************
0 - m bait- PATROIAK GUZTIRA: 3
**********************
**********************
NUKLEO PATROIA: 0 - SATELITE PATROIA:e bait-
**********************
NUKLEOAN:
PATROIA : 0
SEKUENTZIA : komunikazio honen gaiak izango dira aurkitutako erronkak,
identifikatutako aukerak eta emandako irtenbideak
--------------------------------------
SATELITEAN:
PATROIA : e bait-
SEKUENTZIA : hizkera espezializatuen terminologia inguru eleanizdun
batean kudeatzeko, zeinetan hizkuntza bat gutxienez hizkuntza
minorizatua baita.

**********************
0 - e bait- PATROIAK GUZTIRA: 1
**********************
**********************
NUKLEO PATROIA: m interesgarri - SATELITE PATROIA:e bait-
**********************
**********************
m interesgarri - e bait- PATROIAK GUZTIRA: 0
**********************
**********************
NUKLEO PATROIA: 0 - SATELITE PATROIA:b izan ere
```

6.5 Bilatzailetik lortzen den kalkulu-orria

bilatzailea_NS.pl eta bilatzailea_NS.pl programak exekutatzean, automatikoki kalkulu-orriak lortzen ditu. Horren adibidea 9 Irudian ikus dezakegu, KAUSA N-S patroiak HELBURUA N-S kalkulu-orria hain zuzen:

	A		B	
1	Nuc_pattern		Sat_pattern	Pattern_match
2		0	m bait-	3
3		0	e bait-	1
4	m interesgarri		e bait-	0
5		0	b izan ere	0
6		0	e eraginda	0
7		0	m eraginda	0
8		0	m arrazoia	0
9		0	mm eta arrazoi ... horretarako	1
10	e arrazoiengatik		mm -gatik ... -lako ... -gatik	0
11		0	m -elakoan	0
12		0	b -en erroan	0
13		0	m -en ondorioz	0
14		0	e -rekin bat dator	0
15	TESTUAK GUZTIRA		110	

Irudia 9: KAUSA N-S patroiak HELBURUA N-S testuekin parekatzean lortzen den kalkulu-orria

6.6 Etiketatzeko erabiltzen den testu formatua

Etiketatzeko orduan erabiltzen dugun *segment_id*, *segment_parent* eta norantza etiketak izan behar ditu. Honako adibidean ikus daiteke formatua:

```
<column name="segment_id">Farmako tuberkulostatikoak agertu arte,
biriketako tuberkulosia tratatzeko erbiltzen ziren bi teknika,
torakoplastia eta pneumotorax terapeutikoa dira.</column>

<column name="segment_parent"> Pneumologoak gaur, arnas
gutxiegitasuna eramaten duten pakipleuritisean, bular kaiolaren
itxuragabetasunean eta eskoliosian dautzan ondoriozko konplikazioei
egin behar die aurre.</column>

<column name="norantza">NS</column>

<column name="segment_id"> Prozedura kirurgikoak duela 45#5 urte
```

```
burutu ziren</column>

<column name="segment_parent"> eta hurrengo hauetan zeutzan: Alde
bateko torakoplastia 13 kasutan (7 eskuineko aldean eta 6
ezkerrekoan); Pneumotoraxa 15 kasutan (7 eskuinekoak eta 8
ezkerrekoak); alde bietako torakoplastia kasu baten
eta torakoplastia eta pneumotoraxen konbinazioa beste kasu baten.</column>

<column name="norantza">NS</column>

<column name="segment_id"> Sei kasutan gaixoek gaueko VMDa hartzen
zuten (bostek BIPAP eta batek bolumetrikoa); sie kasu hauetatik
bostek, VMD instalatu aurretik etxeko oxigenoterapia kronikoa (EOK)
zeramaten.</column>

<column name="segment_parent"> Azpimarratzekoa da gaueko VMDa
hartzen zuten gaixoetariko lauri EOKa kendu ahal izan zitzaiela.</
column>

<column name="norantza">NS</column>

<column name="segment_id"> Gaixo guztiek zeukaten aireztapen
gutxiegitasuna;</column>

<column name="segment_parent"> hamar kasutan butxaketa-motakoa zen
eta gainerakoetan ezbutxa ketakoa edo mistoa zen.</column>

<column name="norantza">NS</column>

<column name="segment_id"> Sei kasutan gaixoek gaueko VMDa hartzen
zuten (bostek BIPAP eta batek bolumetrikoa);</column>

<column name="segment_parent"> sie kasu hauetatik bostek, VMD
instalatu aurretik etxeko oxigenoterapia kronikoa (EOK) zeramaten.</
column>

<column name="norantza">NS</column>
```

6.7 Etiketatzeko erabili diren patroiak

Eranskin honetan ikus ditzakegu gure lanean etiketatzeko erabili ditugun KAUSA, ONDORIOA eta HELBURUAren patroiak (6.7.1 Eranskina) eta EBALUAZIOA eta INTERPRETAZIOA etiketatzeko erabilitako patroiak (6.7.2 Eranskina). Horretarako, erlazioak zuriunez banatzen dira eta lehen lerroan erlazioen izena idatzi behar da. Ondoren, lerro bakoitzean N-S edo S-N norantza den adierazi eta patroiak jartzen dira, guztiak '*' banatzaile baten bitartez bananduz.

6.7.1 Kausa taldeko patroiak

```
KAUSA
NS * 0 * E bait-
NS * 0 * B izan ere
NS * 0 * MM -gatik ... -lako ... -gatik
NS * 0 * MM eta arrazoi ... horretarako
NS * 0 * E -rekin bat dator
SN * E -nez (gero)? * 0
SN * M -nez * 0
SN * E -ela eta * 0
SN * E -ela bide * M efektu
SN * 0 * B horren ondorioz
SN * E eragile [adl] * B honegatik

ONDORIOA
NS * 0 * b ondorioz
NS * 0 * MM ondorio ... -ri begira
NS * 0 * B ondorioa
NS * 0 * B -en ondorioz
NS * 0 * MM era honetan ... lortu
NS * 0 * MM erakusten ... lortu ... eragin
NS * 0 * B eta horrela
NS * 0 * M sortuz
NS * 0 * B beraz
NS * 0 * M aurkitu
NS * 0 * MM eta ... esan nahi [adl]
NS * 0 * MM aldi berean ... sortzen
NS * 0 * E -raziz
NS * 0 * B [det] artean
NS * 0 * M korrelazioan [adi]
NS * 0 * M -en bitartez jakin
```

```
HELBURUA
NS * 0 * E -tz?eko
NS * 0 * M helburu
NS * 0 * MM helburua ... -tea ... tzea
NS * 0 * MM -tea ... helburua
NS * 0 * MM -tzea ... helburu ... -tea
NS * 0 * E -tzeko asmoz
NS * E -tzeko asmoz * 0
NS * 0 * E (dezagun|dezaten)
NS * 0 * M daitezen
NS * 0 * E burutu nahi izan [adl]
NS * 0 * E [adi] nahi [adl]
SN * E -tz?eko * 0
SN * E -tz?eko helburuarekin * 0
SN * E -tz?eko asmoz * 0
SN * B xede [det] iristeko * 0
SN * E -tu nahian * 0
SN * E dadin * 0
```

6.7.2 EBALUAZIOA eta INTERPRETAZIOAren patroiak

```
INTERPRETAZIOA
NS * 0 * M emaitzek iradokitzen [adl]
NS * 0 * B ondorioz, ez dirudi
NS * 0 * B lortutako emaitzek baieztatzen [adl]
NS * 0 * MM -ri esker ... arrakasta irizpidetzat ... nabarmen [adi]
NS * 0 * MM eta [det] ... emaitzen alderaketa zailtzen [adl]
NS * 0 * E izan daitez?ke
NS * 0 MM erdietsiriko emaitzen arabera ... uste [adl]
NS * E balorazioari [adt] * B hobetu beharreko
NS * 0 * E bail-
NS * 0 * B [det] esan nahi [adl]
NS * 0 * M emaitzak osatzen eta ulertzen [adl]
NS * 0 * MM era honetan ... lor (daiteke|dezakegu)
SN * b egia da * 0

EBALUAZIOA
NS * 0 * B ezta hurrik eman ere
NS * 0 * M kontuan hart-
NS * 0 * E zalantzari tokirik utzi gabe
NS * 0 * MM beraz ... oso aproposatzat kalifikatuak
NS * 0 * M berebiziko garrantzia
```

6.8 Etiketatzailearen emaitzaren irteera estandarra

Etiketatzaileak, segmentu bakoitzeko erlazioen patroiekin parekatzeko positiboa gertatu den begiratzen du. Hurrengo adibidean ikus dezakegu irteera estandarrean erakusten duen emaitzaren zati bat KAUSA,HELBURUA eta ONDORIOA erlazioak etiketatzean.

```
SEGMENT_ID: hona hemen kalkaneo-stop teknika erabiliz gure zerbitzuan
ebakuntza egin diegun haurrek izandako emaitzak.
SEGMENT_PARENT : oin malgua izateagatik
NUKLEARTASUNA : ns
ETIKETATUTAKO ERLAZIOA : Ez da aurkitu erlaziorik
-------------------------------------
-------------------------------------
SEGMENT_ID: proportzioak gora egiten du t1c tumoreetan;
SEGMENT_PARENT : izan ere, hirutik bat hautematen da horrelakoetan.
NUKLEARTASUNA : ns
ETIKETATUTAKO ERLAZIOA : kausa
-------------------------------------
-------------------------------------
SEGMENT_ID: eskuratu ditugun datuek (baita alor jakinetako adituek
emandako iritziek ere) adierazten dutenez,
SEGMENT_PARENT : zientzia-alor jakin batean onartuko diren terminoak
ebaluatzeko hierarkia bat ezarri behar da.
NUKLEARTASUNA : sn
ETIKETATUTAKO ERLAZIOA : Ez da aurkitu erlaziorik
-------------------------------------
-------------------------------------
SEGMENT_ID: 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.
SEGMENT_PARENT : izan ere, iritzi ezberdinetako zientzialari serbiarrek
adostasuna lortu dute eta aurreko hamarkadetan ingelesari eman diote
zientzia-komunikaziorako hizkuntza bakarraren estatusa.
NUKLEARTASUNA : ns
ETIKETATUTAKO ERLAZIOA : kausa
-------------------------------------
-------------------------------------
SEGMENT_ID: alde batetik, gero eta indartsuagoa da nazioarteko
harmonizazioa lortu beharra,
SEGMENT_PARENT : ekonomian, politikan eta kultura eta gizarte gaietan
etenik gabe sortzen ari diren loturak eta elkarren arteko trukaketak
```

```
eraginda;
NUKLEARTASUNA : ns
ETIKETATUTAKO ERLAZIOA : Ez da aurkitu erlaziorik
--------------------------------------
--------------------------------------
SEGMENT_ID: terminologiak berak ere, uztartu egin behar ditu joera orokor
horiek, eransten zaizkien beste batzuekin batera, hala nola: teknologien
aurrerakuntza zorabiagarria, zientziak diziplinartekotasunera eta
hiperespezializaziora daramatzan bilakaera, eta informazioa elkarren
artean berehala trukatu beharra.
SEGMENT_PARENT : gizartearekin lotuta dagoen jarduera denez,
NUKLEARTASUNA : ns
ETIKETATUTAKO ERLAZIOA : Ez da aurkitu erlaziorik
```

6.9 Etiketatzailetik lortzen den kalkulu-orria

Etiketatzailea exekutatzean, automatikoki kalkulu-orria sortzen du. Horren adibidea 10 Irudian ikus dezakegu, kausa taldeko erlazioen etiketazioa ikusten delarik gure corpuseko testuetan.

	A	B	C	D
1	Segment_id	Segment_parent	Nuclearity	Relation_tagged
2	41 oinetan (%64,1) emaitza bikainak erdietsi genitu	lehen mailako arretan, erradiologikoki, estatistikari begira h	ns	ondorioa
3	tamainen araberako banaketa honako hau izan zen:	spearman testaren bidez, besapeko gongoila metastasiaren p	ns	ondorioa
4	400 tumoretatik 336 (%84.0) nos kartzinoma dukta	tamainen araberako banaketa honako hau izan zen: 0-5 mm (	ns	ondorioa
5	1996. eta 2004. urteen bitartean gure ospitalean iza	horien artean, 6 kasu baztertu ditugu hainbat arrazoirengatik.	ns	ondorioa
6	vocall proiektua (vocational language learning for le	gutxi erabiltzen eta irakasten diren hizkuntzetan kontzentratz	ns	ondorioa
7	lan hori, madrilgo hezkuntza ministeritzak diruz bab	lehenik, atzizki eta aurrizkien bidezko eratorpen-prozesuak az	ns	ondorioa
8	hitzaldi honek azken hiru urteotan lau unibertsitate	lan hori, madrilgo hezkuntza ministeritzak diruz babesturiko c	ns	ondorioa
9	euskarak bere aldetik, lehenengo eredua baino ez da	lehenik, erromantikoetan aurreko era bietara jokatzen duen de	ns	ondorioa
10	hau hobeto azaltzeko, kontutan hartu behar da hizk	euskarak bere aldetik, lehenengo eredua baino ez dauka esku	ns	ondorioa
11	edozein azalpen teorikok argitu behar dituen bi desb	hau hobeto azaltzeko, kontutan hartu behar da hizkuntza erro	ns	ondorioa
12	hitzaldi honek azken hiru urteotan lau unibertsitate	edozein azalpen teorikok argitu behar dituen bi desberdintasu	ns	ondorioa
13	komunikazioan landuko ditugun alderdiak zehaztu b	definizioz, toponimoa edo izen geografikoa hauxe da: "izen pr	ns	ondorioa
14	katodotik pasarazten den airearen oxigenoak, kanpo	ioi hauek elektrolitotik zehar anodorantz mugitzen dira eta an	ns	ondorioa
15	sofc pilek bi elektrodo dauzkate, katodoa eta anodoa	sofc pilen funtzionamendua oso erraza da: katodotik pasarazt	ns	ondorioa
16	materialen arloan kokatutako ikerlan honek, sofc (so	erregai-pilak, erregaiak energia elektriko zuzen bilakatzen dute	ns	ondorioa
17	candida albicans animalia homeotermoetarako ohizk	legamia honek odolez sakabanatzeko gaitasuna dauka arazo l	ns	ondorioa
18	peroxisomen proliferatzaileek konposatu kimikoen ta	hauen artean, karraskariak bezalako espezie sentikorretan gib	ns	ondorioa
19	azken urte hauetan industria desberdinek apostu se	izan ere, audi enpresak 1994ean neckarslum herrian aluminio	ns	ondorioa
20	diagnositestak eta positibo faltsuak gainbegiratu dit	horien artean, honako hauek dira nabarmentzeko modukoak:	ns	ondorioa
21	azterketa medikoetan, hainbat arrazoirengatik, kasu	41 oinetan (%64,1) emaitza bikainak erdietsi genituen; 22 oin	ns	ondorioa
22	bame kontrolerako mekanismo gisa erabiltzeko eta g	400 tumoretatik 336 (%84.0) nos kartzinoma duktal iragazkor	ns	ondorioa
23	informatika, bulego-lana eta eraikuntzako arloetako b	eta horrek esan nahi du arlo horietako irlanderazko termino be	ns	ondorioa
24	zelanbait, idazketa espezializatua "idazketa teknikoa	horren ondorioz, ez du lortzen hizkuntzaren eredu matematiko	ns	ondorioa
25	lehenik, atzizki eta aurrizkien bidezko eratorpen-pro	ondorioz, bi desberdintasun nagusi aurkitu dira: euskararen b	ns	ondorioa
26	bakoitzak bere aburuen araberako geografiako termi	eta horrela, egoera nahasia da, eta koherentziarik gabea.	ns	ondorioa

Irudia 10: Kausa taldeko erlazioak etiketatzean lortzen den kalkulu-orria

6.10 PERL ingurunean kalkulu-orriak erabiltzeko ezarpen plana

Garatu diren *bilatzailea_SN.pl*, *bilatzailea_NS.pl* eta *etiketatzailea.pl* programak erabiltzeko, beharrezkoa dugu PERL ingurunerako modulu bat instalatzea. Instalatuta ez badago, honako agindua egikaritu beharko dugu gure terminalean.

```
sudo perl -MCPAN -e 'install Excel::Writer::XLSX'
sudo apt-get update
```

Erreferentziak

Aitzol Astigarraga, Koldo Gojenola, Kepa Sarasola, eta Aitor Soroa. *TAPE Testu-analisirako PERL erremintak*. Udako Euskal Unibertsitatea, 2009.

Iria da Cunha, Eric SanJuan, M. Teresa Cabre Juan Manuel Torres Moreno, eta Gerardo Sierra. A symbolic approach for automatic detection of nuclearity and rhetorical relations among intra-sentence discourse segments in spanish. CICLing (1):462–474, 2012.

Erick Galani, Thiago Pardo, Iria da Cunha, Juan Manuel Torres, eta E. San Juan. Dizer 2.0 – an adaptable on-line discourse parser. In *Anais do III Workshop RST e os Estudos do Texto*, pages 1–17, Cuiabá, MT, Brasil, 2011.

Roxana Girju. Automatic detection of causal relations for question answering. In *MultiSumQA '03 Proceedings of the ACL 2003 workshop on Multilingual summarization and question answering*, pages 76–83, Texas, 2003.

Iakes Goenaga, Olatz Arregi, Klara Ceberio, eta Arantza Díaz de Ilarraza. Automatic coreference annotation in basque. In *11th International Workshop on Treebanks and Linguistic Theories*, pages 115–126, Portugal, 2012.

Mikel Iruskieta. *Pragmatikako erlaziozko diskurtso-egitura: deskribapena eta bere ebaluazioa hizkuntzalaritza konputazionalean*. PhD thesis, EHU, 2014.

Mikel Iruskieta eta Beñat Zapirain. Euseduseg: a dependency-based edu segmentation for basque. In *Actas del XXXI Congreso de la Sociedad Española del Procesamiento del Lenguaje Natural*, pages 41–48, Alicante, 2015.

Mikel Iruskieta, María Jesus Aranzabe, Arantza Diaz de Ilarraza, Itziar Gonzalez, Mikel Lersundi, eta Oier Lopez de la Calle. The rst basque treebank: an online search interface to check rethorical relations. In *4th Workshop "RST and Discourse Studies"*, Brasil, 2013.

Mikel Iruskieta, Arantza Diaz de Illarraza, Gorka Labaka, eta Mikel lersundi. The detection of central units in basque scientific abstracts. In *Actas del XXXI Congreso de la Sociedad Española del Procesamiento del Lenguaje Natural*, Alicante, 2015.

HAP masterra

Mikel Iruskieta, Maria Jesus Aranzabe, Arantza Diaz de Illarraza, eta Mikel Lersundi. Kausazko koherentzia-erlazioak seinalatzeko modua aztergai, euskarazko laburpen zientifikoetan. *Euskal Herriko Unibertsitateko Argitalpen Zerbitzua, GOGOA 14, Xabier Arrazola Gogoan (1962-2015)*, Euskal Herriko Unibertsitateko Hizkuntza, Ezagutza, Komunikazio eta Ekintzari buruzko aldizkaria:45–77, 2016.

Shafiq Joty, Giuseppe Carenini, eta Raymond Ng. Codra: A novel discriminative framework for rhetorical analysis. *Journal Computational Linguistics*, Vol. 41:385–435, 2015.

Willian Mann eta Sandra A. Thompson. Rhetorical structure theory: A theory of text organization. 8(3), 1987.

Willian Mann eta Sandra A. Thompson. Rhetorical structure theory: Toward a functional theory of text organization. *Text-Interdisciplinary Journal for the Study of Discourse*, 8 (3):243–280, 1988.

Willian C. Mann eta Maite Taboada. Rst-ren webgunea, 2010. URL http://www.sfu.ca/rst/.

Michael O'Donell. Rsttool 2.4: a markup tool for rhetorical structure theory. *First International Conference on Natural Language Generation INLG '00*, pages 253–256, 2000.

Attapol T. Rutherford eta Nianwen Xue. Discovering implicit discourse relations through brown cluster pair representation and coreference patterns. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, page 645–654, Sweden, 2014.

Catarina Silva eta Bernadete Riveiro. The importance of stop word removal on recall values in text categorization. *International Joint Conference on Neural Networks*, 2003.

Maite Taboada eta Debopam Das. Annotation uppon annotation: Adding signalling information to a corpus discourse relations. Dialogue and discourse:249–281, 2007.