



IKER
GAZTE
NAZIOARTEKO
IKERKETA EUSKARAZ

IV. IKERGAZTE NAZIOARTEKO IKERKETA EUSKARAZ

2021eko ekainaren 9, 10 eta 11a
Gasteiz, Euskal Herria

ANTOLATZAILEA:
Udako Euskal Unibertsitatea (UEU)

GIZARTE ZIENTZIAK ETA ZUZENBIDEA

**Ezeztapena identifikatzeko
Murriztapen Gramatikako
erregelak sentimenduen
analisiaren testuinguruan**

*Jon Alkorta, Koldo Gojenola eta
Mikel Iruskieta*

169-176 or.
<https://dx.doi.org/10.26876/ikergazte.iv.02.21>



Ezeztapena identifikatzeko Murriztapen Gramatikako erregelak sentimenduen analisiaren testuinguruan

Alkorta, J., Gojenola, K. eta Iruskieta, M.

HiTZ Hizkuntza Teknologiaiko Euskal Zentroa - Ixa,
Euskal Herriko Unibertsitatea (UPV/EHU), Manuel Lardizabal Ibilbidea, 1, 20018 Donostia, Gipuzkoa

jon.alkorta@ehu.eus

Laburpena

Lan honetan, ezeztapena identifikatzeko Murriztapen Gramatikako erregelak aurkezten dira sentimenduen analisia izeneko Lengoaia Naturalaren Prozesamenduko arlorako. Ezeztapenak bere irismen-eremuko hitzek edota hitz-multzoek duten informazio subjektiboan eraldatu dezakete eta beharrezkoa da hori identifikatzea. Hemen, Murriztapen Gramatika ingurunean erregelak nola sortu ditugun azaltzen dugu eta erregelen ebaluazioaren berri ere ematen dugu. Emaitzen arabera, erregelak oro har ondo identifikatzen dituzte ezeztapenaren irismen-eremuko hitzak.

Hitz gakoak: sentimenduen analisia, ezeztapen-markak, Murriztapen Gramatika, erregelak

Abstract

In this work, the rules of Restriction Grammar for identifying rejection are presented for the area of Natural Language Processing called Sentiment Analysis. Negation can transform the subjective information contained in words or phrases in its scope and it is necessary to identify this. This work aims to explain how we created the rules in the Restriction Grammar environment and we also report on the evaluation of the rules. The results show that the rules generally identify well the words in the scope of the negation.

Keywords: *sentiment analysis, negation marks, Constraint Grammar, rules*

1. Sarrera eta motibazioa

Sentimenduen analisia izeneko arloaren helburua hitzek edota esaldiek eta testuek duten informazio subjektiboa eraztea da. Lengoaia Naturalaren Prozesamendua izeneko alorraren baitan kokatzen da eta hainbat ataza barne hartzen ditu: i) esaldia edo testua objektiboa edo subjektiboa den sailkatzea, ii) entitateen (adibidez, telefono mugikorrak) ezaugarriei edo aspektuei loturiko informazio subjektiboa eraztea. Baina badaude ataza gehiago ere eta horietako bat testu subjektiboek balorazio positiboa edo negatiboa egiten duten identifikatzea da.

- (1) Mugikor hori atsegin dut.
- (2) Urak 100 gradutan irakiten du.

Sentimenduen analisisiko atazak aplikatuz, (1) adibidea esaldi subjektibo bat dela esan daiteke eta (2) adibidez, aldiz, esaldi objektiboa. Bestetik, (1) adibidean, balorazio positiboa egiten da *mugikor* hitzari dagokionez. Hiru hurbilpen bereizi ohi dira ataza horretarako eta sentimenduen analisisiko beste atazetarako: i) ezagutzan oinarritutako teknikak: balorazio egiten duten hitzen zerrenda bat abiapuntu izan ohi da; ii) teknika estatistikoak: ikasketa automatikoa eta ikaskuntza sakona, besteak beste eta iii) teknika hibridoak: aurreko bi teknikak uztartzen dituztenak dira.

Esan bezala, ezagutzan oinarritutako teknikan baloratzen duten hitzen zerrenda bat erabiltzen da. Hitz horiek zenbakizko balio bat izaten dute duten informazio subjektiboa adierazteko. 1. taulan ikus daitekeenez, hitz bakoitzak -5 eta $+5$ arteko zenbakizko balio bat edo sentimenduen balentzia bat du. Zeiniek hitzek balorazio positiboa (+) edo negatiboa (−) egiten duten adierazten dute. Zenbakiak, berriz, hitzen balorazioaren intentsitatea adierazten dute, zerotik urrundu ahala, hitzaren balorazioaren intentsitatea gero eta handiagoa da. Gure kasuan, euskarazko hitzei zenbaki horiek esleitzeko SO-CAL tresnako sentimenduen lexikoa (Taboada *et al.*, 2011) euskaratu dugu.

Hitza lexikoian	Balentzia
bikain	+5
on	+2
eskas	-1
txar	-3

1. Taula: Lexikoiko hitz zerrenda bat eta hitzen zenbakizko balio subjektiboa.

Ezeztapenak subjektibitatea duten hitzen sentimenduen balentzian aldaketak sor ditzake. Ezeztapenak, sentimenduen analisiaren ikuspegitik balentzia aldatzaile bat izanik, zenbakizko balio bat du (± 4) eta honelako eraldaketak sortzen ditu balentziadun hitzetan:

(3) Antzezlan hori ez₋₄ da [bikaina₊₄] izan. = 0

(4) Ikusi dugun filma ez₊₄ da [txarra₋₃] izan. = +1

(3) adibidean, *bikaina* hitza ageri da (+4) balioa duena eta ezeztapenaren eraginpean dagoenez, eta ezeztapen-markaren -4 balioa jasotzen duenez, esaldiaren sentimenduen balentzia 0 da. Aldiz, (4) adibidean, *txarra* adjektiboak -3 sentimenduen balentzia du eta ezeztapenaren eraginpean, ezeztapen-markaren +4 balioa jasota, esaldiaren sentimenduen balentzia +1 izatera pasatzen da.

Ezeztapenak sentimenduen analisisan eta, bereziki, hitzen eta esaldien sentimenduen balentzian duen eragina kontuan hartuta, lan honen helburua ezeztapen-marken eragin-eremua identifikatzea izango da, horretarako Murriztapen Gramatikako (Karlsson *et al.*, 2011) erregelak erabiliz.

Kontuan hartu behar da, ezeztapenean hiru kontzeptu direla garrantzitsuak baina guk, lan honetan, bi kontzepturekin egingo dugu lan:

- Ezeztapen-markak. Aditzek eta izenek ezeztapena adieraz dezakete (adibidez, *ekidin* edo *ukatu*) baina guk adberbioek, aditzek, postposizioek, menderagailuek eta juntagailuek adierazten duten ezeztapena landu dugu. (3) eta (4) adibideetan, beltzez adierazitakoak dira ezeztapen-markak.
- Ezeztapenaren irismena. Ezeztapen-markek eraldatzen duten testu-zatiei deritze. (3) eta (4) adibideetan, azpimarkatuta ageri dena da ezeztapenaren irismena.
- Ezeztapenaren fokoa. Ezeztapenaren irismenean gehien edo indar handienez ezeztatutako elementua edo testu-zatia da. (3) eta (4) adibideetan, kortxete bidez adierazitakoak dira ezeztapenaren fokoa. Guk lanean kontzeptu hau ez dugu landuko; izan ere, beharrezkoa da lehenik ezeztapenaren irismena ondo identifikatzea, fokoa identifikatzen hasi aurretik.

2. Aurrekariak eta ikerketaren helburuak

Aurretik esan bezala, ezeztapena bezalako fenomenoak lantzea oso garrantzitsua sentimenduen analisisan, baldin eta tresna automatikoek testuetako subjektibitatea ondo sailkatzea edo identifikatzea nahi badugu. Hori dela eta, ikergaiatzat ezeztapena duten lanak asko dira eta gainera, hizkuntza askotan.

Ezeztapena identifikatzeko gehien erabili izan den tekniketako bat hitz-zaku eredia (ingelesez, *bag-of-words*) izan da. Kasu honetan, hitz bakoitzak subjektibitateari loturiko zenbakizko balio jakin bat du. Pang *et al.*-ek (2002) eta Wiegand *et al.*-ek (2010) ezeztapen hitza gehitzen dute hitz guztietan ezeztapen-markaren lehen puntuazio ikurra agertu arte. Modu horretan, Wiegand *et al.*-ek (2010) "I do not NOT like NOT this NOT new NOT Nokia NOT model" moduko egiturak sortzen dituzte.

Beste teknika bat testuinguko balentzia aldatzaileena (ingelesez, *contextual valence shifter*) da. Teknika honen hastapenak Polanyi eta Zaenen-ek (2006) egin zituzten eta zenbait hitzek (hala nola, intentsifikatzaileek, ezeztapenak edota diskurtso-egiturak ere) hitzek duten sentimenduen balentzian eragiten dutela aldarrikatzen dute. Li eta Wu-k (2010) hitz positiboen eta negatiboen zerrendak eta aldatzaileentzako zerrenda multzoa proposatzen dituzte lehenik (tartean, ezeztapenarena). Ondoren, aldatzaile hauen irismena *n*- moduan definitu dute hitz positibo edo negatiboen aurretik.

Erlazio semantikoetan oinarritutako teknika (ingelesez, *semantic relations*) ere badago. Erlazio semantikoek kontzeptuen edo esanahien arteko erlazioa zehazten dute (adibidez antonimoa, sinonimoa, homonimoa, etab.) eta

hori ezeztapenarekin ere erabil daiteke. Wiegand *et al.* (2010) lanean, esaldien azterketarako perpausaren egitura linguistikoa erabiltzen da eta perpausaren balorazioa bere atalen (izen sintagma, aditz sintagma eta hizkera zatiak) balorazioaren mende egongo balitz bezala lantzen dute. Hori horrela izanik, hitzen ezeztatzeak esaldi oso baten edo haren zati batzuen balorazioa aldatuko balu bezala lantzen dituzte ezeztapena eta bere irismena.

Azken teknika erlazioetan eta dependentzietan (ingelesez, *relations and dependency based*) oinarritutakoa da. Perpaus bateko hitzen eta mendekotasun sintaktikoen arteko erlazio gramatikalek testu-harremanak erazten laguntzen dute. Hori baliatuz, esaterako Jai eta al.-ek (2008) termino (ezeztapena barne) desberdinen eragin-eremua identifikatzen saiatu dira Stanford Parser zuhaitza erabiliz.

Gure lanari dagokionez, sentimenduen analisisian ezeztapena lantzeko Murriztapen Gramatika erabiliko dugu. Hurbilpen honek lotura du, bai testuinguruko balentzia aldatzaileekin, baita erlazio eta dependentziekin ere. Gure lanak lotura du balentzia aldatzaileekin kontuan hartzen dugulako, hitz zerrendak erabiltzen ditugulako eta ezeztapenaren irismena mugatzen saiatu garelako. Bestetik, erlazio eta dependentziekin ere badu lotura lanak hitzen eta sintagmen arteko erlazioak kontuan hartzen ditugulako Murriztapen Gramatika erabiliz ezeztapen-marken irismena edo fokua mugatu nahi dugunean.

3. Metodologia

Jarraian, gure lana garatzeko erabilitako metodologia azalduko dugu. Lehendabizi, lan honen marko teorikoari, hau da, Murriztapen Gramatikari dagozkion azalpenak emango ditugu. Ondoren, erregelak Murriztapen Gramatikan sortzeko prozedura nolakoa izan den azalduko dugu. Azkenik, sortutako erregela horiek nola ebaluatu ditugun aipatuko dugu.

3.1. Marko Teorikoa

Ezeztapen-markak eta euren irismen-eremua identifikatzeko erregelako sortzeko erabili dugun marko teorikoa Murriztapen Gramatika (ingelesez, *Constraint Grammar, CG*) (Karlsson *et al.*, 2011) da. Lengoaia Naturalaren Prozesamenduan, desanbiguazio lexikoa egiteko eta esaldien azaleko azterketa sintaktikoa egiteko erabiltzen da. Bere ezaugarriak dagokienez, hiru aipa daitezke:

- Edozein testu aztertzeke helburua duen formalismo hau hizkuntzarekiko independentea da eta analisi morfologikoan oinarritzen da.
- Murriztailea da; izan ere, desanbiguazio-erregelak deitutakoak ditu analisi morfologikoaren eta funtzio sintaktikoen desanbiguazioa egiteko eta testuinguru jakin batean zuzenak ez diren ahalik eta interpretazio gehienak kentzeko.
- Islapen-erregela edo mapaketa-erregela deiturikoak ere baditu eta euren egitekoa etiketatze morfologikoan eta sintaktikoan dauden hutsuneak betetzea da, horretarako testuingurua kontuan hartuz.

Beraz, gramatikan erregelak idazten direnean eta gramatika konpilatzen denean, gramatikak aztertutako testuko hitzetan etiketak utziko ditu. Etiketak askotarikoak izan daitezke (lema, inflexioa, eratorpena edota funtzio sintaktikoa), nahiz eta gure kasuan, ezeztapen-markak eta euren irismen-eremua izango diren. Bukatzeko, aipatu behar da gramatikak erregelen bidez interpretazio linguistiko automatikoak gehitu, ezabatu edo aukeratzeko aukera ere ematen duela, betiere esaldiaren testuingurua kontuan hartuta.

3.2. Erregelak sortzea

- 1- **Euskarazko ezeztapen-marken zerrenda osatu.** Lehenik eta behin, euskararen ezeztapen-marken zerrenda osatu dugu eta zerrenda hori Altuna *et al.* (2017) eta Altuna *et al.* (1985) lanetan oinarrituta osatu dugu. Honako hauek izan dira ikerketarako erabili ditugun ezeztapen-markak: *ez, ezin, gabe, ezik, salbu, ezta* eta *ezean*.
- 2- **Ezeztapen-marka hori duten esaldiak erauzi corpusetik.** Gero, Euskarazko Iritzi Corpuseko (Alkorta *et al.*, 2016) 96 testutan, ezeztapen-markaren bat duten esaldiak bilatu ditugu eta horiekin esaldi-zerrenda bat osatu dugu, 2. taulan ikus daitekeen moduan. Guztira 359 ezeztapen-marken instantzia lortu dugu.
- 3- **Esaldietako hitzei sentimenduen balentziak esleitu.** Ondoren, 320 esaldi horietako hitzei nahiz esaldi osoei sentimenduen balentzia esleitu diegu. Hori egiteko, hitzei kategoria gramatikala edota marka lexikala

TESTUA	ITEMA
EGU23	Ez da espero euri asko egitea, baina giroa hezea eta freskoa izango da.
EGU23	Asko ez du egingo, baina zerbait bai.
EGU24	Zumaian ez du elurrik egingo baina elur-kota 400-500 metrora jaitsiko da, beraz kontuz errepideetan.
EGU24	Oraintxe bertan ematen du Frantzia iparraldera iritsiko dela eta hortik haize indartsua sortuko du gure lurraldean, baina ez da izango lehengokoa bezain indartsua.
EGU25	Abuztua hasi berri dugun honetan, ez dugu urte sasoi honetarako espero izan ohi dugun eguraldirik izango.
EGU25	Goizean berriz aterri mantenduko du eta ez du ia euririk egingo.

2. Taula: Ezeztapen-markak aztertzeako sortutako azpicorpusaren zati bat.

esleitzen dien *Eustagger* tresnari (Aduriz *et al.*, 2003) *Sentitegi* (Alkorta *et al.*, 2018) sentimenduen hiztegiko hitzak gehitu dizkiogu; lexikoian hitzak kategoria gramatikalean oinarrituta baitaude.

- (5) Pogostkinak ezin [hobeki]₊₂ atera zituen. (MUS20)
- (6) [Irabazi]₊₂ ezinik jarraitzen du Eibarrek. (KIR17)
- (7) Ikuspuntu [politikotik]₋₁ ez ezik, [ekonomikotik]₊₃ ere Greziak esperantza ekarri du Europako hegoaldeko beste herrietara, tartean Euskal Herrira. (POL08)

- 4- **Corpuseko esaldiak banan-banan eskuz aztertu.** Esaldietako ezeztapen-markak identifikatu ondoren, haren irismena zein den adierazi dugu. Irismen-eremuko sentimendu-balentzian nahiz esaldi osokoa utzi duen ondorioa aintzat hartuta, esaldia bera eta haren ezeztapen-marka multzokatu egin ditugu.

- (8) Pogostkinak [ezin hobeki]₊₂ atera zituen. (MUS20)
- (9) [Irabazi]₊₂ ezinik jarraitzen du Eibarrek. (KIR17)
- (10) [Ikuspuntu politikotik]₋₁ ez ezik, [ekonomikotik]₊₃ ere Greziak esperantza ekarri du Europako hegoaldeko beste herrietara, tartean Euskal Herrira. (POL08)

Adibidez, (8) adibidean ageri den ezeztapen-markak (*ezin*) *hobeki* adjektiboaren balentzia indartzen du. Aldiz, (9) adibidean, *ezinik* ezeztapen-markak *irabazi* aditzaren balentzia ahultzen du. Azkenik, (10) adibidean, *ez ezik* egiturak sentimenduen balentziaren ikuspegitik ez du eraginik.

- 5- **Erregelak eskuz prestatuta utzi Murriztapen Gramatikan sortu aurretik.** Urrats honetan, ezeztapen-markak eta haien irismena identifikatzeko erregelak eskuz sortu ditugu, corpuseko adibideetan oinarrituz. Erregela horietako batzuk 3. taulan ikus daitezke. Erregelak eskuz sortzeko sistematikoki ezeztapen-marken inguruan ageri diren hitzen gramatika-kategoriak identifikatu ditugu, betiere, ezeztapena agertzen den proposizioak. (11) adibidean oinarrituz, PM [ezin (+ ad. lag.) + ad. nag.] PM erregela sortu dugu. Aldiz, (12) adibidearekin, PM [ad. nag. + ezin] PM erregela sortu dugu. Hori egin dugu corpuseko adibide guztiekin, abiapuntu izan diren erregela horiek orokortuz¹.

- (11) (...) [ezin da baztertu]₋₁ ekaitz zaparradaren bat izatea. (EGU35)
- (12) [Irabazi]₊₂ ezinik jarraitzen du Eibarrek (...) (KIR22)

¹Erregelek honako informazioa biltzen dute: i) kakeok [] ezeztapen-markaren irismena adierazten dute, ii) parentesiek () sintagma horren agerpena hautazkoa dela adierazten dute, iii) *letra etzanak* ezeztapen-marka edo egitura lexikalizatuen osagaiak adierazten dituzte, iv) barrak /, azkenik, ezeztapen-markaren irismenean, aukeran aipaturiko osagai ezberdinek daudela adierazten du, v) ISk eta ASk izen- eta aditz-sintagma adierazten dituzte (ageri diren beste elementuak: adjektiboa, adberbioa, aditza, aditz laguntzailea eta sintagma dira), vi) puntuazio-markaren murriztapena (PM) da (helburua erregeletan ezeztapen-marken irismena esaldi barnean kokatzea eta beste esaldietako elementuak ez hartzea da).

Adb.	Ezeztapen-marka	Erregelen egitura
(1)	ezin	PM <i>ezin</i> + [adjektiboa/adberbioa] (+ atzizki konp.) PM
(5)	ezin	PM [ad. nag. + <i>ezin</i>] PM PM [<i>ezin</i> (+ ad. lag.) + ad. nag.] PM
(10)	ez	PM [IS] <i>ez ezik</i> PM

3. Taula: Sentimendu-balentzian eragin ezberdinak dituzten ezeztapenak identifikatzeko proposatu diren erregetako batzuk.

- 6- **Eskuz sortutako erregelak Murriztapen Gramatikaren ingurunera egokitu.** Aurreko urratseko erregeletan oinarrituz, erregelak Murriztapen Gramatika formalismora egokitu eta ebaluatu ditugu. 1. irudian erregelak nolakoak diren ikus daiteke. Erregela horiek honela osatuta daude:

```
LIST PUNUAZIOA = PUNT.PUNT PUNT.KOMA PUNT.BI.PUNT PUNT.GALD PUNT.ESKL PUNT.HIRU PUNT.PUNT.KOMA;
LIST EZ = 'ez';
LIST BESTERIK = 'beste';

# (2)
# besterik ez egiturak
MAP (!besterikezHAS) TARGET (DET) IF (0C BESTERIK) (1C EZ);

(...)
```

1. Irudia: Ezeztapen-markak eta beren irismen-eremua identifikatzeko erregelen adibide bat Murriztapen Gramatika formalismoan.

- LIST zerrendak. Bertan hitz jakinak edo bestelako elementuak, puntuazio-markak adibidez, zerrendatzen dira. 1. irudian, puntuazio-markak (LIST_PUNUAZIOA) eta ezeztapen-markak (LIST_EZ, LIST_BESTERIK) zerrendatu ditugu.
- MAP komandoa (islapen-erregela). Hauen bidez erregeletan adierazitako egiturak corpusean bilatzen dira eta aurkitzen badira, etiketa bat esleitzen zaio corpusean bilatu den egiturari. Ezaugarri hauek osatuta dago erregela:
 - * Esleituko den etiketa. Etiketak aurretik ! ikurra du, 1. irudian.
 - * Etiketa esleituko zaion hitza. Adibidez, *!besterikezHAS* etiketa determinatzaile bati (DET) esleituko zaio.
 - * Erregelaren baldintzak. Esaterako, *!besterikezHAS* etiketa esleitzeko, 0 posizioan BESTERIK zerrendak, hau da, *beste* hitzak, (0C BESTERIK) egon behar du eta 1 posizioan EZ zerrendak (kasu honetan, *ez* ezeztapen-markak), erregelak funtziona dezan.

Murriztapen Gramatika formalismoan idatzirik erregela bakoitzak bere etiketa uzten du corpuseko le-roan. 2. irudian, esaterako, Murriztapen Gramatikaren sortu ditugun erregelek bi hitzi etiketa ezarri diete (*!gabeAUR* etiketa *harrotu* hitzean eta *!gabe* etiketa *gabe* hitzean).

3.3. Erregelak ebaluatzea

Gramatikak corpuseko hitzetan etiketak utzi dituela begiztatu ondoren, gramatikako erregelako ebaluatzeke urrats hauek burutu ditugu:

```
''<, >'' <<PUNT.KOMA>>
PUNT.KOMA
''< batere >''
'' batere '' ADB ARR ZERO w39,L-A-ADB-ARR-13,lsfi48 @ADLG %SINT
''< harrotu >'' S:278/0
'' harrotu '' ADI SIN PART NOTDEK w40,L-A-ADI-SIN-38,lsfi49 @-JADNAG %ADIKAT S:278 !gabeAUR
''< gabe >'' S:141/0
'' gabe '' ADB ARR ZERO w41,L-A-ADB-ARR-14,lsfi50 @KM> %SINT S:141 !gabe
''<$.>'' <<PUNT.PUNT>>
PUNT.PUNT
```

2. Irudia: Murriztapen Gramatika formalismoan sortutako erregelek esleitutako ! etiketa.

Erregela		Kopurua		Kopurua
LIST motakoak				24
MAP motakoak				73
	Orokorrak	3	EDO/EDOTA/ALA EZ	4
	Egitura lexikalizatuak	10	EZ... EZ...	4
	EZIN	18	EZ	14
	SALBU	4	GABE	3
	IZAN_EZIK	5	EZTA	9
	EZ_EZIK	1		

4. Taula: Murriztapen Gramatikan sortutako zerrenden eta komandoen kopurua.

- 1- Euskarazko Iritzi Corpusa (Alkorta *et al.*, 2016) prozesatu erregelak ebaluatu ahal izateko. Murriztapen Gramatika formalismoan oinarritzen den euskarazko desanbiguatzaile morfosintaktikoa erabili dugu (Aduriz *et al.*, 1997) erabiliko diren 48 testu horiek prozesatzeko eta Murriztapen Gramatika formalismora ekartzeko.
- 2- Murriztapen Gramatika formalismoan sorturiko 73 erregelak corpusaren test zatiko 144 testuetan aplikatu.
- 3- Gramatikak aplikaturiko erregelak ebaluatu. Erregelek corpusean esleitutako etiketak ebaluatu ditu etiketatzailerak. Etiketatzailerak hiru etiketa erabili ditu: ETIK_ONDO (erregelaren etiketa zuzena denean), ETIK_FALTA (corpuseko hitzak etiketa behar zuenean eta erregelak jarri ez dionean) eta ETIK_GAIZKI (erregelak corpuseko hitzari etiketa jarri dionean eta hitzak etiketa behar ez duenean).
3. irudian, hiru hitz daude eta hirurek egon beharko lukete etiketatuta. Hiruretatik bi (*harrotu* eta *gabe*) etiketatuta daude eta ETIK_ONDO etiketa jarri zaie, lerroen hasieretan. Baina, hori ez da hala *batere* hitzaren kasuan eta hitz horri ETIK_FALTA etiketa jarri zaio.

```

"<, >"<PUNT.KOMA>"
PUNT.KOMA
ETIK.FALTA "< batere >"
"batere" ADB ARR ZERO w39,L-A-ADB-ARR-13,lsfi48 @ADLG %SINT
ETIK.ONDO "< harrotu >" S:278/0
"harrotu" ADI SIN PART NOTDEK w40,L-A-ADI-SIN-38,lsfi49 @-JADNAG %ADIKAT S:278 !gabeAUR
ETIK.ONDO "< gabe >" S:141/0
"gabe" ADB ARR ZERO w41,L-A-ADB-ARR-14,lsfi50 @KM> %SINT S:141 !gabe
"<$.>"<PUNT.PUNT>"
PUNT.PUNT

```

3. Irudia: Erregelak ebaluatzeko etiketatzailerak hitzei esleitutako etiketak.

4. Emaitzak

4.1 azpiatalean, Murriztapen Gramatikan sortu ditugun zerrenden eta komandoen berri eman dugu; bai, kopuruari dagokionez, baita euren ezaugarriei buruz ere. 4.2 azpiatalean, erregela horien ebaluazioaren berri eman dugu.

4.1. Lana

Lan honetan, ezeztapen-markak eta euren irismen-eremua identifikatzeko garatu diren zerrenden eta komandoen kopurua 4. taulan ikus daiteke. Alde batetik, 24 LIST zerrenda eratu ditugu. Bertan, ezeztapen-marka guztiak, puntuazio-markak eta juntagailuak ageri dira zerrendatuta. Bestetik, berriz, 73 MAP komando ere sortu ditugu.

Komando gehienak ezeztapen-marka bakoitzaren irismen-eremua identifikatzeko sortu dira. Horrela, EZIN ezeztapen-markaren irismen-eremua identifikatzeko 18 komando erabili dira eta EZ ezeztapen-markarenak identifikatzeko, berriz, 14. Beste ezeztapen-marken kasuan, euren irismeneko hitzak identifikatzeko erabili den komandoen kopurua baxuagoa da. Egitura lexikalizatuak² deitutakoak identifikatzeko, berriz, 10 komando erabili dira eta beste 3 komando orokor ere sortu ditugu.

²Lan honetan egitura lexikalizatuak deitzen diogu ezeztapen-marka bat maiztasun handiz beste hitz jakin batekin agertzen diren egiturei. Horren adibide dira: *dударik gabe*, *zalantzarik gabe*, *baino ez*, etab.

Eragina	Doitasuna	Estaldura	F1	Hitzak (instantziak): ✓/✗/±
Indartu	1,00	1,00	1,00	4 (2): 4/0/0
Ahuldu	0,93	0,80	0,86	1.050 (192): 784/63/203
Eraginik ez	0,97	1,00	0,98	36 (19): 35/1/0
Guztira	0,93	0,80	0,86	1.090: 823/64/203

5. Taula: Emaitzak ezeztapen-markek sentimendu-balentzian eragiten dutenaren ikuspegitik.

	Doitasuna	Estaldura	F1	Hitzak (instantziak): ✓/✗/±
Ezeztapen-markak	1,00	0,96	0,98	195 (195): 188/0/7
Egitura lexikalizatuak	0,96	1,00	0,98	28 (16): 27/1/0
Irismena	0,91	0,75	0,82	867 (195): 601/63/196
Guztira	0,93	0,80	0,86	1.090: 823/64/203

6. Taula: Emaitzak ezeztapenari loturiko elementu-motaren ikuspegitik.

4.2. Ebaluazioa

Ebaluazioa bi modutan egin dugu. Alde batetik, ezeztapenak balentziadun hitzetan eta irismen-eremuan duen eraginaren arabera ebaluazioa egin dugu (5. taula). Bestetik, ezeztapenari loturiko elementu-motaren arabera ebaluazioa (6. taula) ere egin dugu.

5. taulako datuen arabera³, erregelen doitasuna 0,90tik gorakoa da eta estaldura ere ona da, ezeztapen-markak balentzia ahultzen duen kasuan izan ezik (0,80). Horrenbestez, F1 0,95tik gorakoa da ezeztapen-markak irismeneko balentzia aldatu ez edo indartzen duen kasuetan eta 0,86koa ezeztapen-markak irismen-eremuko balentzia ahultzen duen kasuetan.

6. taulan, berriz, ezeztapenarekin lotutako hiru elementuen azterketa egiten da. Hiru elementuak identifikatzeko doitasuna 0,90tik gorakoa da. Hala ere, estalduraren kasuan, ezeztapen-markak eta egitura lexikalizatuak identifikatzean 0,98koa da baina irismena identifikatzerakoan, nabarmen baxuagoa da, 0,75koa hain zuzen ere. Hori dela eta, ezeztapen-marketan eta egitura lexikalizatueta F1a 1etik gertu dago eta irismen-eremuaren kasuan 0,82koa da.

Errore-analisiak erregelen osaketari buruzko informazioa eman digu. irismen-eremuko 196 hitz ez direla identifikatu ikusi da. Horren atzean puntuazio-markaren murriztapena dagoela ohartu gara; izan ere, zeharkako kalte batzuk sortzen ditu: esaldiren batean zerrendaketa bat dagoenean, eta zerrendaketa hortako elementuak koma bidez bereizita daudenean, esaterako.

5. Ondorioak

Lan honetan, sentimenduen analisiaren arloan ezeztapen-markak eta euren irismen-eremua identifikatzeko erregelak garatu eta ebaluatu ditugu Murritzapen Gramatika erabiliz. Corpusetik ezeztapen-markak dituzten esaldiak lortu eta ezeztapen-marka horiek euren irismen-eremuko sentimenduen balentzian izan dezaketen eraginak (balentzia indartu, ahuldu edo eraginik ez izan) identifikatu ostean, eskuz erregelak batzuk osatu ditugu eta ondoren, horiek Murritzapen Gramatikaren formalismora eraman ditugu. Azkenik, erregelak ebaluatu ditugu eta emaitzek adieraz dutenez, erregelak oro har ondo dabilta, nahiz eta ezeztapen-marken irismen-eremua identifikatzeko zailtasun gehiago dituzten, puntuazio-markaren murriztapena tarteko.

6. Etorkizunerako planteatzen den norabidea

Etorkizunera begira, bi helburu ditugu. Alde batetik, erregelak are zehatzagoak izatea nahiko genuke eta horregatik, corpus handiago batean oinarrituz, sorturiko erregelak probatu eta hobetu nahiko genituzke; batez ere, ezeztapen-marken irismen-eremua identifikatzeko dauden zailtasunengatik. Beste alde batetik, erregela hauei era-

³5. taulan, eskuineko zutabearen, hitzei eta instantziei loturiko datuak jarri ditugu. Lehenik eta behin, hitz-kopurua agertzen da, ondoren, parentesi artean, instantzia-kopurua eta ondoren, ezeztapenari loturiko hitzak ondo, gaizki edo ez diren asmatu. Ikusten den moduan, hitz eta instantzia-kopuruak ezberdinak dira. Ezeztapen-marken kasuan, hitz bat instantzia bat da. Irismenaren kasuan, instantziako hitz-kopurua ezberdina da. Guztira 192 irismen-eremu daude corpusean eta horiek 1.050 hitz dituzte. Azkenik, egitura lexikalizatueta, hitz batetik hiru hitzerainokoa izan daiteke instantzia bat, betiere egitura lexikalizatu horren ezaugarrien arabera.

bilera praktikoa eman nahiko genieke eta testuen subjektibitatea aztertzen duen euskarazko tresna bat garatu eta erregela hauek bertan integratu nahiko genituzke.

Erreferentziak

- Aduriz, Itziar, Izaskun Aldezabal, Inaki Alegria, J Arriola, Arantza Diaz de Ilarraza, Nerea Ezeiza, eta Koldo Gojenola. 2003. Finite state applications for basque. In *EACL'2003 Workshop on Finite-State Methods in Natural Language Processing*, 3–11. Citeseer.
- , José María Arriola, Xabier Artola, Arantza Díaz de Ilarraza, Koldo Gojenola, eta Montse Maritxalar. 1997. Morphosyntactic disambiguation for Basque based on the Constraint Grammar Formalism. *Proceedings of Recent Advances in NLP (RANLP97), Tzigov Chark (Bulgary)* 282–288.
- Alkorta, Jon. 2019. Sentimenduen analisi automatikorantz: oinarritzko baliabideen sorkuntza eta hizkuntza maila ezberdinetako balentzia-aldatzaileen identifikazioa. *Doktoretza-tesia*.
- , Koldo Gojenola, eta Mikel Iruskieta. 2016. Creating and evaluating a polarity-balanced corpus for basque sentiment analysis. In *IWoDA16 Fourth International Workshop on Discourse Analysis. Santiago de Compostela, September*, volume 29.
- , Koldo Gojenola, eta Mikel Iruskieta. 2018. Sentitegi: Semi-manually created semantic oriented basque lexicon for sentiment analysis. *Computación y sistemas* 22.1295–1306.
- Altuna, Begoña, María Jesús Aranzabe, eta Arantza Díaz de Ilarraza. 2017. Euskarazko ezeztapenaren tratamendu automatikorako azterketa. In *Iñaki Alegria, Ainhoa Latatu, Miren Josu Ormaetxebarria eta Patxi Salaberri (arg.), II. IkerGazte, Nazioarteko Ikerketa Euskaraz: Giza Zientziak eta Arteak, 127-134, Udako Euskal Unibertsitatea (UEU), Bilbo*.
- Altuna, Patxi, Pello Salaburu, Patxi Goenaga, María Pilar Lasarte, Lino Akasolo, Miren Azkarate, Piarres Charriton, Andolin Eguskita, Jean Haritschelhar, Alan King, Jose Maria Larrarte, Jose Antonio Mujika, Beñat Oyharçabal, eta Karmele Rotaetxe. 1985. Euskal Gramatika Lehen urratsak (EGLU) II. *Euskaltzaindiko Gramatika batzordea, Euskaltzaindia, Bilbo*.
- Esuli, Andrea. 2008. Automatic generation of lexical resources for opinion mining: models, algorithms and applications. In *Acm sigir forum*, volume 42, 105–106. ACM New York, NY, USA.
- Karlsson, Fred, Atro Voutilainen, Juha Heikkilae, eta Arto Anttila. 2011. *Constraint Grammar: a language-independent system for parsing unrestricted text*, volume 4. Walter de Gruyter.
- Li, Nan, eta Desheng Dash Wu. 2010. Using text mining and sentiment analysis for online forums hotspot detection and forecast. *Decision support systems* 48.354–368.
- Pang, Bo, Lillian Lee, eta Shivakumar Vaithyanathan. 2002. Thumbs up? sentiment classification using machine learning techniques. *arXiv preprint cs/0205070*.
- Polanyi, Livia, eta Annie Zaenen. 2006. Contextual valence shifters. In *Computing attitude and affect in text: Theory and applications*, 1–10. Springer.
- Taboada, Maite, Julian Brooke, Milan Tofiloski, Kimberly Voll, eta Manfred Stede. 2011. Lexicon-based methods for sentiment analysis. *Computational linguistics* 37.267–307.
- Wiegand, Michael, Alexandra Balahur, Benjamin Roth, Dietrich Klakow, eta Andrés Montoyo. 2010. A survey on the role of negation in sentiment analysis. In *Proceedings of the workshop on negation and speculation in natural language processing*, 60–68.

7. Eskerrak eta oharrak

- Lan hau *Sentimenduen analisi automatikorantz: oinarritzko baliabideen sorkuntza eta hizkuntza maila ezberdinetako balentzia-aldatzaileen identifikazioa* (Alkorta, 2019) tesitik eratorrita dago.
- Lan hau *DeepReading: Mining, Understanding, and Reasoning with Multilingual Content* (RTI2018-096846-B-C21) (MCIU/AEI/FEDER,UE) proiektuaren barruan finantziatu da.