



IKER
GAZTE
NAZIOARTEKO
IKERKETA EUSKARAZ

III. IKERGAZTE

NAZIOARTEKO IKERKETA EUSKARAZ

2019ko maiatzaren 27, 28 eta 29
Baiona, Euskal Herria

ANTOLATZAILEA:
Udako Euskal Unibertsitatea (UEU)

GIZA ZIENTZIAK ETA ARTEA

**Sentimenduen tratamendu
konputazionalerantz: gramatika
maila ezberdinetako sentimendu
balentzia aldatzaileen bila**

*Jon Alkorta, Koldo Gojenola eta
Mikel Iruskieta*

39-46 or.
<https://dx.doi.org/10.26876/ikergazte.iii.01.05>



Sentimenduen tratamendu konputazionalerantz: gramatika maila ezberdinetako sentimendu balentzia aldatzaileen bila

Alkorta, Jon; Gojenola, Koldo eta Iruskieta, Mikel

Euskal Herriko Unibertsitatea (UPV/EHU). Ixa Taldea

jon.alkorta@ehu.eus

Laburpena

Sarean gero eta iritzi-testu gehiago daude, iritzi positiboa edo negatiboa izan dezaketenak, eta horiek automatikoki prozesatzeko beharra sortu da, eskuz egitea lan nekeza delako. Hizkuntza Prozesamenduan, lexikoian oinarritzen diren tresnak daude eta esaldi edo testu bat positiboa edo negatiboa den esaten dute. Lexikoian, hitzek [-5] (balorazio oso negatiboa) eta [+5] (balorazio oso positiboa) arteko sentimendu balentzia dute. Lan honetan, euskarak gramatika maila ezberdinetan dituen testuinguruzko balentzia aldatzaileak identifikatzeko azterketa azaltzen da. Hots, hitzen sentimendu balentzia edo balorazioa aldatzen duten fenomenoak identifikatzen dira. Lexikoian oinarritutako sentimendu analisirako tresnek askotan huts egiten dute balentzia aldatzaileak ez identifikatzeagatik. Hori motibaziotzat hartuta eta *Sentitegi* sentimendu lexikoa eta Euskarazko Iritzi Corpusa baliabidetzat hartuta, fonologiako, morfologiako, sintaxiko eta diskurtsoko balentzia aldatzaileak identifikatu ditugu. Emaitzek erakusten dutenez, euskarazko balentzia aldatzaileek hitz edo sintagmaren sentimendu balentzia indartu edo ahuldu dezakete. Beste batzuetan, ordea, balentzia aldatzailea den hizkuntza fenomenoak sentimendu hitzen balentzian ez du eraginik.

Hitz gakoak: sentimendu analisia, testuinguruzko balentzia aldatzaileak, euskara

Abstract

There are more and more opinion texts with positive or negative subjectivity in the Internet and there is a need to automatically process them because doing this by hand is unfeasible. In language processing, there are tools based on lexicon and they determine if a sentence or text has a positive or negative opinion. In the lexicon, words have sentiment valence between [-5] (very negative) and [+5] (very positive). This work presents a study to identify contextual valence shifters in Basque. That is to say, we identify phenomena that change the sentiment valence or evaluation of words. Many lexicon-based tools for sentiment analysis have a lack of information about valence shifters. With that motivation and using the Basque Opinion Corpus and Sentitegi sentiment lexicon, we have identified phonological, morphological, syntactic and discourse-based sentiment shifters. According to our results, sentiment shifters can strengthen or weaken the sentiment valence of words or phrases in Basque. In some cases, however, the linguistic phenomenon associated to a valence shifter does not have any effect on the sentiment valence of words or phrases.

Keywords: sentiment analysis, contextual valence shifters, Basque

1. Sarrera eta motibazioa

Internetek eta sare sozialek indarra hartu zutenetik, milaka eta milaka iritzi topa ditzakegu Interneten¹. Iritzi horiek informazio garrantzitsua eman dezakete produktu, gertaera edo ekintza baten inguruan; produktua hobetzen lagun dezaketelako, besteak beste. Baina lan nekeza da iritzi guztiak banan-banan biltzea, aztertzea eta horrek iritziak automatikoki prozesatzeko beharra eragin du. Hizkuntza-ezagutzan oinarrituta zenbait baliabide sortu dira iritziek balorazio positiboa edo negatiboa duten adierazten dutenak. Baliabide hauen oinarrian sentimendu balentzia duten hitzez osaturiko lexikoiak daude. Sentimendu balentzia² hitz bati esleitzen zaion zenbakizko balioa da eta hitzaren orientazio semantikoa, berriz, hitzak balorazio positiboa edo negatiboa duen adierazten du. Esaterako, “txar”

¹Euskarazko iritziak hainbat webgunetan topa daitezke. Egunkarietan topa daitezke (Berria, Argia, Naiz) edo baita webgune espezializatuetan ere (Kritiken Hemerroteka, Zinea.eus).

²Ataza hau sentimendu analisia arloan kokatzen denez, artikuluan zehar, balorazio subjektiboa edo subjektibitatea adierazten duten hitzek sentimendu balentzia dutela adieraziko dugu. Aipatzen diren “hurbi”, “ezezkotasun” edota “txikitasun” moduko hitzek ez dute sentimendua adierazten baina terminologia errazteko asmoz, eurek sentimendu balentzia dutela esango dugu.

hitzaren sentimendu balentzia $[-3]$ da, $[-5]$ (balorazio oso negatiboa) eta $[+5]$ (balorazio oso positiboa) arteko eskala batean, beraz, balorazio negatibokoa da. Baina lexikoian oinarritutako sentimendu analisirako tresnek ez dituzte gramatika maila ezberdinetako fenomenoak egoki prozesatzen eta emaitza ez da bere horretan fidagarria.

- (1) **OIASSO MUSEOA.** [Erabiltzailea]. 2018/02/06. BERRIA deskontu sarea.

"Bikotea eta biok Oiasso museoan izan ginen. Bertako langileak oso **jatorrak** eta deskontuarekin inongo **arazorik**_[-3] ez. Agian umeein edo joateko **egoki**agoa delako baina museo bera **ez** zitzaigun **gehi**gi **gustatu**_[+0,5]; informazioa **nahasi**_[-3] **xamarra**, eta batzuetan erdi **ezkutuan** dago gure **gustorako**."_[-0,42]

(1) adibidean, Oiasso museoari buruzko iritzi bat ageri da. Beltzez dauden hitzak sentimendua duten hitzak dira. Hitz horiek "jator", "arazo", "egoki", "gustatu", "nahasi", "ezkutu" eta "gusto" dira. Iritzi-testuei sentimendu balentzia esleitzeko SO-CAL tresnaren (Taboada, 2008) euskarazko bertsioa erabili dugu eta iritzi horretako hitzei eta iritzi-testu osoari sentimendu balentzia esleitu diegu. Testu osoaren sentimendu balentzia $[-0, 42]$ da. Bestalde, SO-CAL tresnak "arazo", "gustatu" eta "nahasi" hitzei ere balentzia esleitu die³. Baina hitz horien balentziak ez dira guztiz zuzenak, testuingurua tarteko. Adibidez, "arazo" hitzaren kasuan, eskuinean ezeztapen-marka du eta ezkerrean polaritate negatiboko osagaia. Testuinguru horren ondorioz, "arazo" hitzaren balentziaren zeinuak aldaketa jasan beharko luke. Bestalde, "gustatu" aditzak $[+0, 5]$ balentzia du baina eskuinean "gehi"gi" adberbioa du. Adberbio horrek aditzean eragiten du, haren balentzia indartuz. Azkenik, "nahasi" adjektiboaren eskuinean, "samar" adberbioa adjektiboa intentsifikatzen ageri da. Gainera, adberbioak bustidura adierazkorra du: "xamar", adberbioa beraren adierazkortasuna indartuz. Beraz, "arazo" hitzak $[-3]$ baino balio negatiboagoa eduki beharko luke, bustidura adierazkorra duen adberbio intentsifikatzailearengatik.

Hiru kasu horietan, hitzen sentimendu balentzian eragiten dutenei **testuinguruzko balentzia aldatzaileak** (*contextual valence shifters*, ingelesez) deitzen zaie. Lan honen helburua, euskarazko sentimendu aldatzaileak identifikatzea da eta gramatika-kategoria ezberdinetakoak identifikatu nahi ditugu: fonologiatik eta morfologiatik hasi eta sintaxia eta diskurtsoraino. Balentzia aldatzaile horiek identifikatzen badira, lexikoian oinarritutako sentimendu analisirako tresnak eraginkorragoak izan daitezke, Morsy eta Rafearen (2012) lanean bezala.

2. Arloko egoera eta ikerketaren helburuak

Testuinguruzko balentzia aldatzaileak identifikatzea sentimendu analisiaren atzetako bat da eta bertako arlo askotan da erabilgarria. Sentimendu analisisian, hiru arlo nagusi bereizi izan dira: *i*) dokumentu mailako sentimendu sailkapena, hots, testu batek balorazio positiboa den negatiboa duen zehaztea (Pang *et al.*, 2002; Turney, 2002), *ii*) esaldi mailan, subjektibitate sailkapena, hau da, esaldi batek informazio objektiboa (informazio faktuala) edo informazio subjektiboa duen zehaztea (Wiebe *et al.*, 1999) eta *iii*) aspektu eta entitate maila non iritzi positibo edo negatiboaren jomuga⁴ identifikatu nahi den (Hu eta Liu, 2004).

Arlo horiek burutzeko, ezinbestekoa da baliabideak sortzea. Testuinguruzko balentzia aldatzaileak identifikatzeko ere baliabide batzuen beharra dago: *i*) corpusarena, bertan identikatuko baitira testuinguruzko balentzia aldatzaileak eta *ii*) sentimendu lexikoiarena, testuko hitzei sentimendu balentzia esleitzeko baliatuko dena eta balentzia aldatzaileek hitz horien balentzia eragiten duten neurtzen laguntzen duena.

- Sentimendu analisirako corpusak hainbat hizkuntzatan daude: ingelesez (Taboada, 2008), arabieraz (Abdul-Mageed eta Diab, 2012) eta gaztelaniaz (Plaza-del Arco *et al.*, 2016), besteak beste.
- Sentimendu lexikoiak sortzeko bi modu daude. Batetik, hiztegiak oinarrituz sor daitezke. Hiztegiek hitzen sinonimoak eta antonimoak dituzte eta hori lexikoiaren sorkuntzan baliagarria da. Lan gehienak WordNet datu-basean (Miller *et al.*, 1990) oinarritzen dira. Modu honetan lexikoiak erraz eta azkar sor daitezke eta sor daitezkeen akatsak eskuz konpondu. Baina domeinuari oso loturik daude eta hori desabantaila da. Corpusean oinarritutako lexikoiak bi helburu dituzte: *i*) balentziadun hitz multzo bat edukita, domeinu bati loturiko corpuseko sentimendu hitzak eta haien balentzia jakitea (Hatzivassiloglou eta McKeown, 1997) eta *ii*) erabilera orokorreko lexikoi bat domeinu zehatz batera egokitzea (Wiebe eta Mihalcea, 2006).

Bukatzeko, gure ikergaia den testuinguruzko sentimendu aldatzaileen inguruan ez dago hainbeste aurrekari guk dakigula. Polanyi eta Zaenen (2006) da erreferentziatzeko lana. Bertan, gramatika maila ezberdinetako balentzia aldatzaileak deskribatzen ditu ingeles hizkuntzan. Esaterako, esaldi mailan, ezeztapena eta intentsifikatzaileak, modalak eta ironia balentzia aldatzaileak direla dio. Diskurtso mailan, berriz, lokailuak, diskurtso-egitura eta zeharkako estiloa dira balentzia aldatzaileetako batzuk. Erreferentziatzeko lan horretako metodologiari jarraituz, errumanierako balentzia aldatzaileen inguruko lana egin da (Colhon *et al.*, 2016).

³SO-CAL tresnaren euskarazko bertsioan dagoen *Sentitegi* sentimendu lexikoiaren estaldura tarteko, tresnak sentimendua duten beste hitzei ez die balentzia esleitu.

⁴Iritziaren jomugak iritziaren mintzagaiari egiten dio erreferentzia. Hots, "liburua ederra da" esaldian, iritziaren jomuga "liburua" da.

Gure helburua dokumentu mailako sentimendu sailkapena ardatz hartuta eta corpora eta sentimendu lexikoa baliabide izanda, euskarako testuinguruzko balentzia aldatzaileak identifikatzea da.

3. Ikerketaren muina

3.1. Ikerketarako baliabideen sorkuntza

Euskarazko balentzia aldatzaileak identifikatzeko, bi baliabide sortu eta garatu ditugu: Euskarazko Iritzi Corpora (Alkorta *et al.*, 2016) eta euskarazko sentimendu lexikoa, *Sentitegi*, (Alkorta *et al.*, 2018).

Euskarazko Iritzi Corpora sortzeko erreferentziatzat *SFU Review Corpus* (Taboada, 2008) corpora erabili dugu. Corpus hark dituen antzeko ezaugarriak dituen corpora eratu nahi izan dugu. Horrela, 240 iritzi-testuko corpus bat sortu dugu eta iritziak sei gaien ingurukoak dira: eguraldia, literatura, musika, zinema, politika eta kirola. Gai bakoitzeko 40 iritzi-testu bildu ditugu, horietako erdiak balorazio positibokoak eta beste hainbeste balorazio negatibokoak. Aipatu behar da, literaturan, balorazio negatiboko iritziak biltzea zailagoa izan dela, Egañak (2013) bere lanean esplikatzen duenez, azken 30 urteetako euskal literatura kritiken % 84k balorazio positiboa baitute.

Sortu behar izan dugun bigarren baliabidea *Sentitegi* izeneko euskarazko sentimendu lexikoa izan da. Lexikoi bat sortzea lan konplexua da eta denbora mugatua eta euskararen baliabide faltarengatik, testuen balorazioa positiboa edo negatiboa esleitzeko SO-CAL tresnaren ingelesezko eta gaztelaniazko sentimendu lexikoiak (Brooke *et al.*, 2009; Taboada *et al.*, 2011) euskaratu eta eskuz zuzendu ditugu Elhuyar (Elhuyar, 2013) eta Zehazki (Sarasola, 2005) hiztegi elebidunak erabiliz. Bi sentimendu lexikoi horiek itzultzen bukatzean, 1.237 hitzeko lexikoi bat lortu dugu. Hitz horiek lau gramatika-kategorietakoak dira: izenak, adjektiboak, aditzak eta adberbioak.

1. taula. *Sentitegi* euskarazko sentimendu lexikoiko zenbait adibide.

Hitza lexikoian	Sentimendu balentzia
bikain	[+5]
on	[+2]
eskas	[-1]
txar	[-3]

1 Taulan, *Sentitegi* euskarazko sentimendu lexikoiko lau adibide daude. Adjektiboak dira eta eskalan daude. Eskala hori [-5] eta [+5] artekoa da eta eskala horretan “bikain” hitzak [+5]eko balentzia du eta “txar” hitzak, berriz, [-3]ko balentzia. Azkenik, “on” eta “eskas” hitzen balentziak tartean daude.

3.2 Metodologia mailaz maila

Azpiatal honetan, aurreko bi baliabideak erabiliz gramatika maila bakoitzean balentzia aldatzaileak identifikatzeko erabili dugun metodologia esplikatuko dugu. Halaber, kasu batzuetan, metodologian hurbilpen bat aukeratu edota baliabide jakin bat erabili behar izan da, eta horiek ere aipatuko ditugu.

Balentzia aldatzaile morfologiko eta fonologikoak identifikatzeko, lehenik eta behin, iturri bibliografikoetan euskarazko hizki eta bustidura adierazkorri buruzko informazioa bildu dugu. Erabili ditugun iturrien artean Mugika (1982), Euskara Institutua (2010) eta Oñederra (1990) daude. Bildutako informazioarekin taula bat osatu dugu, hizki edo bustidura adierazkorren informazioarekin, 2. Taulan ikusten den moduan.

2. taula. Zenbait hizkiren eta bustidura adierazkorren taula, bakoitzaren ezaugarriekin.

Hizkia	Izena	Adjektiboa	Aditza	Sailkapen semantikoa	Adibidea
-zale	X			Zaletasuna	Ardozale
ez-	X	X	X	Ezekotasuna	Ezberdin
-z- >-x-				Hurbiltasuna/txikitasuna	Gazte >gaxte, zahar >xahar

Ondoren, corpuseko hitzak banan-banan aztertzen joan eta bertan topatutako hizkiekin eta bustidura adierazkorrekin, 3. Taula eratu dugu. Taula osatzeko, Aldezabal *et al.*en (2006) lana erabili dugu, bertan, hizki bakoitzari loturiko informazioa agertzen baita, tartean baita hizkiek hitzen balentzia nola eragiten duten ere. Esaterako, 3. Taulan, bigarren adibidean, “-txo” hizkiak hitzaren balentzia ahultzen du. Aldiz, hirugarren adibidean, bustidura adierazkorrek hitzaren balentzia indartzen du. Lehen adibidean, ez dago ez bat ezta bestea ere.

Sintaxiak oso eremu zabala hartzen du eta ezeztapena, zehazki **ezeztapen-markak sentimendu balentzia aldatzaile** diren, aztertu dugu. Horretarako, lehenik eta behin egin dugun urratsa euskarazko ezeztapen-marken zerrenda osatzea izan da. Ezeztapen-marken zerrenda osatu ahal izateko iturri bibliografikoetara jo dugu. Altuna

3. taula. Corpusean agertutako hizkien eta bustidura adierazkorren azterketa.

Adibidea	Hitza	Hizkia	+	-	0	Sail. semantikoa	Iz.	Adj.	Adi.
1	istorio								
2	erasotxo	-txo		X		Txikitasuna	X		
3	pattal	-t- → -tt-	X			Hurbiltasuna			

et al. (2017) lanari jarraituz, adberbioek (“ez”, “salbu”), “ezin izan” aditzak, postposizioek (“-rik gabe”), menderagailuek (“ezik”, “ezean”) eta juntagailuek (“ezta”) adierazten duten ezeztapena aztertu dugu eta ez, izenetan (“ukazioa”) eta aditzetan (“ukatu”) agertzen dena. Horiek, gainera, gure hurbilpenetik sentimendu lexikoian leudeke. Ezeztapen-marken zerrenda osatu ondoren, ezeztapen-marka horiek dituzten esaldiak Euskarazko Iritzi Corpuseko 96 iritzi-testuetatik atera ditugu. Guztira 359 ezeztapen-marka bildu ditugu. Horietatik 296 “ez” ezeztapen-marka dira, “gabe” 47tan agertu da, “ezin” 38 aldiz, eta bestelako ezeztapen-markak maiztasun gutxiagoz agertu dira: “salbu” bi aldiz eta “izan ezik”, “ezean” eta “ezta” behin.

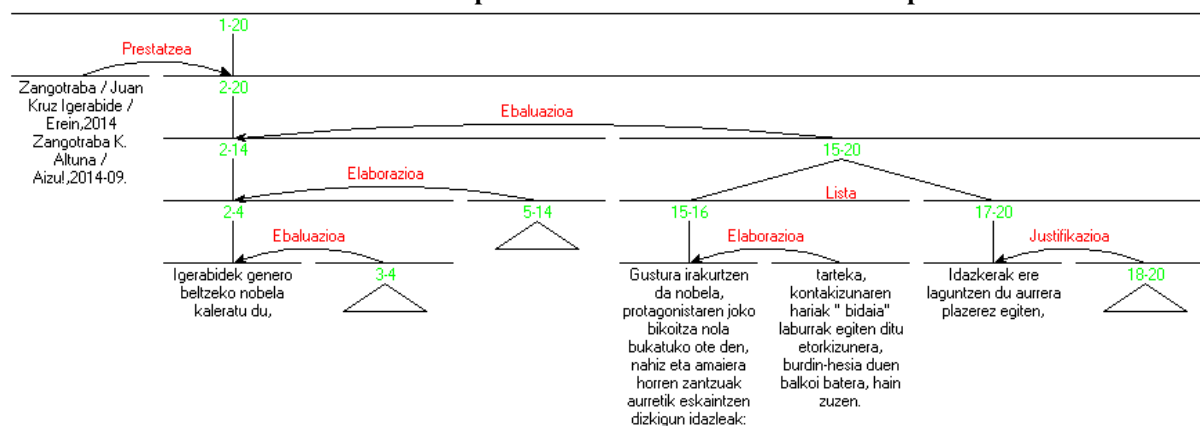
Bigarren urratsean, esaldi horietako hitzei balentzia esleitu diegu ikerketarako baliabideen sorkuntzan aipaturiko *Sentitegi* sentimendu lexikoia erabiliz. 4. Taulan, esaterako, sentimendu lexikoia lehen errenkadako “hobeki” hitzari [+2] balentzia esleitu dio. Bigarren errenkadako “irabazi” hitzari ere balentzia bera ([+2]) esleitu dio.

4. taula. Esaldietako hitzei balentzia jartzea eta ezeztapen-marken eragina aztertzea.

Testu-kodea	Esaldi-zatia	<i>Sentitegi</i> ko hitza	Hitzaren sentimendu balentzia
MUS20	Pogostkinak <u>ezin</u> hobeki ⁺ atera zituen	<i>hobeki</i>	[+2]
KIR17	<u>Irabazi</u> ezinik ⁻ jarraitzen du Eibarrek	<i>Irabazi</i>	[+2]

Ondoren, corpusaren azterketa egin dugu eta aztertu duguna izan da ea ezeztapen-markek esaldiko hitzen balentzian eragiten duten eta eragiten badute nola eragiten duten. Berriz, 4. Taulara joanez gero, lehen errenkadan, “ezin” ezeztapen-marka ageri da eta bere eskuinean “hobeki” adberbioa dago. Kasu honetan, ezeztapen-markak adberbioaren balentzia ([+2]) indartzen du. Bigarren errenkadan, berriz, ezeztapen-marka berak eskuinetara “irabazi” aditza du baina aurreko kasuan ez bezala, hemen, ezeztapen-markak aditzaren balentzia ([+2]) ahultzen du. Beraz, “ezin” ezeztapen-markak jokamolde ezberdina du ondoko hitzen gramatika-kategoriaren arabera.

Aztertu dugun laugarren gramatika maila diskurtsoa izan da. Egitura Erretorikoen Teoria (Mann eta Thompson, 1988) edo RST (*Rhetorical Structure Theory*, ingelesez) erabili dugu. Lehenik eta behin egin duguna, Euskarazko Iritzi Corpuseko 28 iritzi-testu etiketatzea izan da. Etiketatzerakoan honako urratsak egin dira: i) testua segmentatu eta EDUak, diskurtso unitateak, lortu, ii) testuko unitate zentrala (UZ), hots, testuko EDU garrantzitsuen identifikatu eta azkenik, iii) diskurtso-unitate ezberdinen artean loturak eginez, koherentziazko erlazio erretorikoen bidez, testuaren diskurtso-zuhaitza sortu. Jarraian, honen deskribapen zabalagoa egingo dugu 1 irudia erabiliz.

1. irudia. Euskarazko Iritzi Corpuseko LIB29 iritzi-testua RST hurbilpenaz etiketatuta.

1 irudian, LIB29 iritzi-testua RST jarraituz etiketatu da. Ikusten denez, testua segmentatuta dago eta testu-zati horietako bakoitza EDU bat da. “Igerabidek genero beltzeko nobela kaleratu du” EDU garrantzitsuen, hots, unitate zentrala da eta diskurtso-zuhaitzeko koherentziazko erlazio erretoriko nagusiek horrekin egiten dute lotura.

Erlazio erretorikoak mota ezberdinetakoak dira: nukleo bakarrekoak edo nukleoanitzak. PRESTATZEA, ELA-BORAZIOA, JUSTIFIKAZIOA eta EBALUAZIOA nukleo bakarreko erlazioak dira. Hau da, erlazio erretorikoan EDU bat nukleoa da, EDU garrantzitsuena erlazioan, eta bestea, satelitea, nukleoaren osagarria dena. LISTA erlazio erretorikoa, berriz, nukleo anitzekoa da eta bertan, bi EDUak nukleoak dira eta garrantziari dagokionez maila bereko unitateak dira elkarri lotzean.

Hurrengo urratsean, etiketatutako corpuseko erlazio erretoriko bakoitzaren ezaugarriak zehaztu ditugu, izan ere, ezaugarri horiek konbinatuz balentzia aldatzaileak diren eta badira, zer-nolako balentzia aldatzaileak diren jakitea baita helburua. Adibidetzat 15 eta 16. EDUek osatzen duten ELABORAZIOA erlazio erabiliko dugu.

- 1- Koherentzia erlazioetako EDUen sentimendu orientazioa. Lehen EDUak (15) orientazio semantiko positiboa du eta bigarren EDUak (16), berriz, neutrala.⁵
- 2- Koherentzia erlazioaren orientazio semantikoa. Erlazio erretorikoaren EDU batek orientazio semantiko positiboa du eta besteak neutrala. Erlazio erretoriko osoaren sentimendu orientazioa positiboa da.
- 3- Erlazio erretorikoaren nuklearitatea. 15-16 EDUek osatzen duten erlazioan, 15 nukleo-unitatea (N) da eta 16, berriz, satelite-unitatea (S). Hori geziaren norabideak adierazten du. Beraz, erlazioa NS da eta ez SN.
- 4- Erlazio erretorikoaren eta UZaren arteko distantzia. Distantzia neurtzeko bien artean dauden erlazio kopurua zenbatu dugu. Kasu honetan, bien artean LISTA eta EBALUAZIOA erlazioak daude, beraz, distantzia +2 da.

Guk hiru aspektu landu ditugu: i) erlazioetan, **nukleoak, sateliteak, lehen EDUak edo azken EDUak duen orientazio semantikoan adostasun handiagoa** erlazio osoarekiko; ii) ea erlazioek nukleoarekiko duten distantziaren⁶ arabera, **erlazioen eta testu osoaren orientazio semantikoen artean nolako adostasuna** dagoen eta iii) **erlazio erretoriko motak zer eragin duen haren orientazio semantikoan**.

3.3. Lanaren emaitzak mailaz maila

Lan honen emaitzak dira: i) balentzia aldatzaileak gramatika maila guztietan daudela eta ii) balentzia aldatzaileek hitz, sintagma edo esaldiaren sentimendu balentzia indartu edo ahuldu dezaketela edo ez dutela eraginik.

2. irudia. Lan honetan identifikatutako gramatika maila ezberdinetako balentzia aldatzaileak.

	Fonologia	Morfologia	Sintaxia	Diskurtsoa
Balentzia aldatzaile	Indartu	<s>/<z> -> <x> <t> -> <tt> <z> -> <tx>	-garri -min -tsu -tza -en -tzar -zale -nahi -gura	ezin · UZrekiko distantzia · nukleoa · erlazioetako azken zatia · KONTZETSIOAK nukleoa
	Ahuldu	-gabe des- -egi -keria ez-/ez ezin-/ezin -xe/-txe -txo -gaitz -ska/-xka -txa a-	ezin gabe izan ezik salbu ezta ez	· EBALUAZIOAK nukleoa
	Eraginik ez	-ezia -tasun	· edo/edota/ala + ez · ez ezik · egitura · lexikalizatuak	

2 irudian, **balentzia aldatzaile fonologikoen** zerrenda ageri da lehen zutabearen. Corpusean, guztira, hiru bustidura adierazkor mota daude eta horiek balentzia indartzen dute. Esaterako, “lagun” “lagun” ([+2]) baino indartsuagoa da. Bustidura adierazkorak “lagun” hitzaren balentzia positiboa are positiboago bihurtzen du.

Bestalde **balentzia aldatzaile morfologikoetan**, 2. irudiko bigarren zutabearen, ageri diren aurrizki gehienak ezeztapenarekin lotuta daude eta horiek hitzaren balentzia ahultzen dute. Esaterako, “orekatu” ([+1]) hitzari “des-” aurrizkia⁷ itsasita (“desorekatu”), haren balentzia ahuldu eta negatibo ([−3]) bilakatuz. Atzizkietan ugaritasuna

⁵Erlazio erretoriko batean, EDU batek orientazio semantiko positiboa duenean eta besteak neutrala, bien batura beti ez da orientazio semantiko positiboa. Baina kasu horretan horrela da.

⁶Erlazio erretorikoaren eta unitate zentralaren arteko distantzia neurtzeko bien artean dauden erlazio erretoriko kopurua hartuko da aintzat.

⁷Taboada *et al.* (2011) lanak ezeztapen-markei [−4] balentzia esleitzen die eta guk irizpide hori aintzat hartuz eta morfologiara ekarri, ezeztapenari lotutako hizkiei [−4] balentzia esleitu diegu.

eta aniztasuna dago; balentzia indartu edo ahuldu dezaketelako. Esaterako, corpusean ageri den “krisi” ([−3]) hitza aztertzen badugu, ikusiko dugu “-tzar” atzizki handigarriak (“krisitzar” hitza osatuz) bere balentzia indartu eta are negatiboago bilakatzen duela. Horren kontrara, “-txo” atzizki txikigarriak “krisi” hitzaren balentzia ahultzen du (“krisitxo”), bere balentzia [−3] baino baxuago bilakatuz, nahiz eta kasu honetan, balentziaren zeinuan (−/+) ez den aldaketarik gertatzen; negatiboa izaten jarraitzen baitu. 2. irudiko bigarren zutabeko azken errenkadan, sentimendu balentzian eraginik ez duten bi hizki (“-ezia” eta “-tasun”) ikus ditzakegu. Hizki horiek itsasita dauden hitzek edozein zeinutako sentimendu balentzia eduki dezakete, baina aipatu ditugun hizkiak hitz horiei itsasteak ez du haien sentimendu balentzian inolako eraginik. Adibide pare bat jartzearen, ‘pobre’ hitzak [−2] balentzia du eta “pobrezia” hitzak ere balentzia bera du, ez delako aldaketa semantikorik egon. Modu berean, “poz” adjektiboak [+2] balentzia du eta “poztasuna” hitzak bere balentzia bera du. Bi kasu hauetan, hizkiek ez dute bere baitan ebaluaziorik, tasuna edo ezaugarria adierazten dute.

Sintaxi mailan, 2. irudiko hirugarren zutabean, **ezeztapen-markek sentimendu balentzian nola eragiten duten** agertzen da. Hiru balentzia aldaketa mota egon daitezke: *i*) indartzea, *ii*) ahultzea edota *iii*) aldaketarik ez. “Ezin” da hitz-multzoen, sintagma edo esaldien, sentimendu balentzia indartzen duen bakarra. (2). adibidean, “ezin” ezeztapen-marka adjektibo baten ezkerrean agertzen da eta horren ondorioa da adjektiboaren sentimendu balentzia indartzea. Horrenbestez, adjektibo horren balentzia [+4]tik gora kokatzen da. Adjektiboa edota adberbioa “ezin” ezeztapen-markaren eskuinetara agertzean, hauen balentzia indartze bat gertatzen da.

- (2) Dena nahasten da maisulan ezin ederragoa_[(+4)] osatzeko. (MUS21).⁸

Hala ere, gutxitan gertatzen den fenomeno da, Euskarazko Iritzi Corpusean % 1,96an (359tik 7tan), hain zuzen ere. Kasurik ohikoena ezeztapen-markek balentzia ahultzea da. Corpusean, kasuen % 89,98an (359tik 323tan) sentimendu balentzien ahultzea gertatzen da eta ezeztapen-marka mota asko dira hori egiten dutenak.

- (3) Irabazi_[(+2)] ezinik jarraitzen du Eibarrek, baina oso puntu ona eskuratu du Getaferen zelaian. (KIR17)

(3) adibidean, “ezin” ezeztapen-marka ageri da baina adibide horretan, aditzaren eskuinera ageri da eta horren ondorioa, aditz horren balentziaren ahultzea da. Ikusten denez, “ezin” ezeztapen-markak jokamolde bikoitza du (ikus (2) eta (3) adibideak) eta bere inguruan agertzen den hitzaren gramatika-kategoriak zerikusi handia du jokamolde bat edo bestearekin. Balentzia ahultzea gertatzen den kasuetan, gogoan izan behar da, ahultzeak hitz-multzo edo esaldien balentzian zeinu aldaketa ekar dezakeela. Hau da, hitz-multzo edo esaldi batek balentzia [±3] baino baxuagoa bada, ezeztapen-marken ondorioz ([±4] sentimendu balentziaduna), haien balentzian zeinuaren aldaketa gertatuko da. Aldiz, haien balentzia [±4] edo altuagoa bada, nahiz eta balentzian ahultze bat egon, balentzian ez da zeinu aldaketarik gertatuko.

Azkenik, Euskarazko Iritzi Corpusean, kasuen % 8,08an (27 kasutan 359etatik), ezeztapen-markek ez dute hitz-multzo edo esaldiaren sentimendu balentzian eragin.

- (4) Ikuspuntu politikotik_[(−1)] ez ezik, ekonomikotik_[(+3)] ere Greziak esperantza ekarri du Europako hegoaldeko beste herrietara, tartean Euskal Herrira. (POL08)
- (5) Sei puntu baino ez dituela, hamaseigarren postuan da Real sailkapenean. (KIR27)

(4) adibidean, “ez ezik” ezeztapen-markak kontraste balioa du eta “ikuspuntu politikotik”_[(−1)] ezeztatzen badu ere, ezeztapen horren helburua, azken batean, esaldiaren hurrengo zatian informazio gehikuntza egitea da (Silvennoinen, 2017). Horrelako egituretan, “ez ezik” ezeztapen-markak bere aurretik dagoen elementua (“Ikuspuntu politikotik”_[(−1)]) adibidearen kasuan) sintaktikoki ezeztatzen du. Baina sentimendu balentziaren ikuspegitik, ez da balentziarik ezeztatu eta ondorioz, ahuldu; azken batean, informazioa gehitzeko funtzioa bete duelako. Azkenik, (5) adibidean, egitura lexikalizatu baten adibidea ageri da. “Baino ez” egitura “ez” ezeztapen-markaz osatuta dago baina “baino” juntagailuarekin batera agertzen da eta biek unitate bakarra osatzen dutenez, “ez” markak ez du eraginik. Hau da, “baino ez” egitura “bakarrik” hitzaren parekidea eta esanahi berekoa da. Horrenbestez, egitura lexikalizatu bat dela eta ezeztapen-markak bere irismena galdu duela esan daiteke. Horrek “baino ez” egiturak bere sentimendu balentzia jakina duela esan nahi du.

Balentzia aldatzaileen bila, aztertzen ari garen hirugarren gramatika maila diskurtsoarena da. Diskurtsoan, batez ere, **koherentziatzko erlazio erretorikoetan**, bere zatietako bakoitzaren, nukleoaren edo satelitearen, edo lehen EDUaren edo bigarrenaren, orientazio semantikoak erlazio erretoriko osoaren orientazioarekin zenbateko

⁸Adibideetan tipografiak honako esanahia du: azpimarrak ezeztapen-markaren irismena adierazten du, **beltzez** dauden hitzak ezeztapen-markak dira eta parentesi artean hitzen sentimendu balentzia agertzen da. Gainera, (2) adibidearen kasuan badago beste berezitasun bat. “Ezin” ezeztapen-marka irismenaren partetzat hartu da ezeztapen-marka horrek bere baitan bi esanahi dituelako: *i*) ezeztapena eta *ii*) posibilitatea.

adostasuna duen neurtu dugu. Cohen kappa neurriak erakusten dutenez, nukleoak satelitearekin baino adostasun handiagoa erakusten du (0,73 vs 0,65). Bestalde, erlazioak azken EDUak lehenak baino adostasun handiagoa erakusten du (0,78 vs 0,58). Beraz, nukleoak eta erlazioak azken EDUak erlazio osoaren orientazio semantikoarekiko adostasuna indartzeko joera erakusten dute eta balentzia adostasuna indartzen duten aldatzaileak dira.

Beste ondorietako bat izan da **erlazio erretoriko bat unitate zentraletik gertuago** egon ahala, erlazioa eta tesuaren arteko orientazio semantikoaren adostasuna indartu egiten dela. 1 distantzia bien arteko orientazio semantiko adostasuna % 0,67koa da, 0 distantzian 0,60koa eta 2 distantzian, azkenik, 0,59koa. Hortik aurrera, adostasuna asko jaisten da. Esaterako, 3 distantzian adostasuna % 0,42koa da.

Bukatzeko, ikusi da **erlazio-mota bat edo beste izateak nukleo edo satelitearen orientazio semantikoa erlazio osoan nagusitzea ekar dezakeela**. Esaterako, KONTZETSIOAk beti nukleoa indartzen du (“horiek horrela izanik ere (satelitea, neutrala), emaitza zalantzarik da (nukleoa, negatiboa)”: negatiboa) eta EBALUAZIOAk ahuldu (“Hori egiten du egileak, esaterako, Ordezko atezainaren balada-n, (nukleoa, neutrala), nire ustez liburu-ko ipuin onenetakoa (satelitea, positiboa)”: positiboa). Lehen erlazioaren kasuan, nukleoa eta erlazio osoaren orientazio semantiko adostasuna 0,70raino iristen da eta bigarren kasuan, nukleoa eta erlazio osoaren arteko orientazioaren adostasuna 0,67koa da eta satelitearena 0,82koa.

4. Ondorioak

- i) Euskarak testuinguruzko balentzia aldatzaile fonologikoak ditu. Ikusi dugun moduan, Euskarazko Iritzi Corpusean hiru bustidura adierazkor mota topatu ditugu eta denek balentzia indartzen dute.
- ii) Euskararen badaude balentzia aldatzaile morfologikoak. Aurrikizkiak nahiz atzikizkiak izan daitezke balentzia aldatzaile eta bi zentzutan alda dezakete: indartuz edo ahulduz. “-ezia” eta “-tasun” atzikizkiek, aldiz, balentzian ez dute eraginik.
- iii) Sintaxian, ezeztapen-markak ere sentimendu balentzia aldatzaile dira. Ezeztapenak sentimendu balentzia beti ahuldu egiten duela dirudien arren, azterketak erakutsi digu hori ez dela beti gertatzen. Sentimendu balentzia indartu dezakete edo eraginik ez izan ere, nahiz eta maiztasun gutxiagoz gertatzen diren.
- iv) Diskurtso mailan, erlazio erretorikoetan, nukleoak eta erlazioaren azken zatiak erlazioak EDUaren orientazio semantikoaren eta erlazio erretoriko osoaren orientazio semantikoaren arteko adostasuna indartzen dute. Hala ere, unitate zentralarekiko 0, 1 eta 2 distantzian egoteak erlazio erretorikoaren eta testu osoaren orientazio semantikoaren adostasuna indartzen duela ikusi dugu. Bukatzeko, KONTZETSIOA erlazio-motak nukleoaren orientazio semantikoa indartzen du, bere orientazio semantikoa erlazio osoan gailenduz eta EBALUAZIOAk nukleoaren orientazio semantikoa ahultzen du, satelitearena indartuz eta bera erlazio osoan gailenduz.

5. Etorkizunerako planteatzen den norabidea

Aztertu beharrekoa litzateke intentsifikatzaileak balentzia aldatzaile diren eta badira nolakoak diren. Intentsifikatzaileek ere balentzia indartu (“oso polita”) edo ahuldu (“nahiko polita”) dezaketela dirudi. Aditzen denbora irrealaren balentzia aldatzailea den ere aztertu beharko litzateke. Izan ere, “polita litzateke hori” esaldian, “polita” hitzak +2 balentzia duen arren, balentzia hori ahuldu eta 0an bilakatuko litzateke, denbora irrealaren delako.

6. Erreferentziak

- Abdul-Mageed, Muhammad, eta Mona T Diab. 2012. Awatif: A multi-genre corpus for modern standard arabic subjectivity and sentiment analysis. In *LREC*, 3907–3914. Citeseer.
- Aldezabal, Izaskun, Olatz Ansa, Bertol Arrieta, Xabier Artola, Aitzol Ezeiza, G. Hernández, eta Mikel Lersundi. 2006. Edb1: A general lexical basis for the automatic processing of basque. In *Proceedings of the IRCS Workshop on linguistic databases*. IRCS Workshop on linguistic databases.
- Alkorta, Jon, 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*, 58–62.
- , —, eta —. 2018. Sentitegi: Semi-manually created semantic oriented basque lexicon for sentiment analysis. *Computación y Sistemas* 22.
- 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*.
- Brooke, Julian, Milan Tofiloski, eta Maite Taboada. 2009. Cross-linguistic sentiment analysis: From english to spanish. In *Proceedings of the international conference RANLP-2009*, 50–54.

- Colhon, Mihaela, Mădălina Cerban, Alex Becheru, eta Mirela Teodorescu. 2016. Polarity shifting for romanian sentiment classification. In *INnovations in Intelligent SysTems and Applications (INISTA), 2016 International Symposium on*, 1–6. IEEE.
- Egaña, Ibon, 2013. *Kritikarako hurbilketa literaturaren soziologiatik. Egunkari eta aldizkarietako euskal literatur kritikaren analisia (1975-2005)*. UPV/EHU tesia.
- Elhuyar, Hizkuntza Zerbitzuak, 2013. Elhuyar hiztegia: euskara-gaztelania, castellano-vasco. Usurbil: Elhuyar.
- Euskara Institutua, EHU. 2010. *Sareko Euskal Gramatika*.
- Hatzivassiloglou, Vasileios, eta Kathleen R McKeown. 1997. Predicting the semantic orientation of adjectives. In *Proceedings of the 35th annual meeting of the association for computational linguistics and eighth conference of the european chapter of the association for computational linguistics*, 174–181. ACL.
- Hu, Mingqing, eta Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, 168–177. ACM.
- Mann, William C, eta Sandra A Thompson. 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text-Interdisciplinary Journal for the Study of Discourse* 8.243–281.
- Miller, George A, Richard Beckwith, Christiane Fellbaum, Derek Gross, eta Katherine J Miller. 1990. Introduction to wordnet: An on-line lexical database. *International journal of lexicography* 3.235–244.
- Morsy, Sara A., eta Ahmed Rafea. 2012. Improving document-level sentiment classification using contextual valence shifters. In *Proceedings of the 17th International Conference on Applications of Natural Language Processing and Information Systems, NLDB'12*, 253–258, Berlin, Heidelberg. Springer-Verlag.
- Mugika, Luis María. 1982. Morfología de la composición lexical euskérica. *Fontes linguae vasconum: Studia et documenta* 14.233–272.
- Oñederra, Lourdes. 1990. *Euskal fonologia: palatalizazioa: asimilazioa eta hots sinbolismoa*. Servicio Editorial de la Universidad del País Vasco= Euskal Herriko Unibertsitatea.
- Pang, Bo, Lillian Lee, eta Shivakumar Vaithyanathan. 2002. Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, 79–86. Association for Computational Linguistics.
- Plaza-del Arco, Flor Miriam, M Teresa Martín-Valdivia, Salud María Jiménez, M Dolores Molina-González, eta Eugenio Martínez-Cámara. 2016. Copos: corpus of patient opinions in spanish. application of sentiment analysis techniques. *Procesamiento del Lenguaje Natural* 83–90.
- Polanyi, Livia, eta Annie Zaenen. 2006. Contextual valence shifters. In *Computing attitude and affect in text: Theory and applications*, 1–10. Springer.
- Sarasola, Ibon. 2005. *Zehazki: gaztelania-euskara hiztegia*. Alberdania.
- Silvennoinen, Olli O. 2017. Not only apples but also oranges: Contrastive negation and register. *Studies in Variation, Contacts and Change in English* 19.
- Taboada, Maite, 2008. SFU Review Corpus [Corpus]. Vancouver: Simon Fraser University.
- , Julian Brooke, Milan Tofiloski, Kimberly Voll, eta Manfred Stede. 2011. Lexicon-based methods for sentiment analysis. *Computational linguistics* 37.267–307.
- Turney, Peter D. 2002. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics*, 417–424. Association for Computational Linguistics.
- Wiebe, Janyce, eta Rada Mihalcea. 2006. Word sense and subjectivity. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, 1065–1072. Association for Computational Linguistics.
- Wiebe, Janyce M, Rebecca F Bruce, eta Thomas P O'Hara. 1999. Development and use of a gold-standard data set for subjectivity classifications. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, 246–253. Association for Computational Linguistics.

7. Eskerrak eta oharrak

Lan hau Eusko Jaurlaritzaren PRE_2018_2_0033 bekari, UPV/EHUren UPV/EHU taldea (GIU16/16) proiektuari eta Eusko Jaurlaritzaren BERBAOLA: Hizkuntzaren eta hizketaren teknologia RIS3ko eremuetara aplikatuta (KK-2018/00014) proiektuari esker egin da.