



IKER
GAZTE
NAZIOARTEKO
IKERKETA EUSKARAZ

IV. IKERGAZTE

NAZIOARTEKO IKERKETA EUSKARAZ

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

ANTOLATZAILEA:
Udako Euskal Unibertsitatea (UEU)

INGENIARITZA ETA ARKITEKTURA

**Twitterreko Euskal
Komunitatearen Eduki Azterketa
Pandemia Garaian**

*Joseba Fernandez de Landa,
Iker Garcia, Ander Salaberria
eta Jon Ander Campos*

137-144 or.
<https://dx.doi.org/10.26876/ikergazte.iv.03.17>

ANTOLATZAILEA:



LAGUNTZAILEAK:



Twitterreko Euskal Komunitatearen Eduki Azterketa Pandemia Garaian

Fernandez de Landa, J., García, I., Salaberria, A., Campos, J.A.

HiTZ Zentroa - Ixa, Euskal Herriko Unibertsitatea (UPV/EHU). 20018 Donostia, Euskal Herria.

joseba.fernandezdelanda@ehu.eus

Laburpena

Hizkuntzaren Prozesamenduak eskaintzen dituen teknika ez-gainbegiratuak erabiliz Twitterreko euskal komunitatean COVID-19aren pandemiak izan duen eragina aztertzea da lan honen asmoa. Azterketa hau aurrera eramateko sare sozial horretako erabiltzaileen euskarazko txioak masiboki bildu eta denboraren arabera ordenatu dira. Pandemiaren eragina neurtzeko, denboraren arabera edukia nola aldatu den aztertu da, horretarako testuetan azaltzen diren hitz zein emojiak aldaketa kuantitatibo zein kualitatiboak baliatu dira. Azterketa kuantitatiboan, terminoek garai desberdinetan izan duten maiztasunaren aldaketari erreparatu zaio, maiztasunen erregresio lineala erabiliz. Azterketa kualitatiboan, hitzen bektore trinkoak baliatu dira, pandemiaren garai desberdinetan hitz eta emoji adierazgarrienek esanahian izan duten bilakaera aztertzeko.

Hitz gakoak: Twitter, euskara, Covid-19

Abstract

The aim of this work is to study the impact of the COVID-19 pandemic on the Basque Twitter community using unsupervised techniques based on Natural Language Processing. In order to carry out this study, large quantities of tweets were gathered and sorted by time from Basque Twitter users. To analyze the impact of the pandemic, the variability of the content over time has been studied, through quantitative and qualitative changes in the words and emojis that appear in the texts. In the quantitative analysis, the shift at the frequency of the terms was calculated using linear regression over frequencies. In the qualitative analysis, Word Embeddings were used to study the changes in the meaning of the most significant words and emojis at different times during the pandemic.

Keywords: Twitter, Basque language, Covid-19

1. Sarrera eta motibazioa

COVID-19aren ondorioz eman den osasun krisialdi globalak eragin handia izan du gizartearen arlo guztietan. Eragin hau aztertzeko metodologia desberdinak aurkeztu dira pandemia garaian, birusaren kutsakortasun mailaren analisi estatistikoa izan delarik usuena. Azterketa guzti hauek birusaren joera ulertzeko oso baliagarriak diren arren, ez dira eraginkorrak COVID-19ak suposatu duen aldaketa soziala ulertzeko.

Azken urteotan sare sozialak gizartearen ispilu bilakatu dira hauen erabilera asko handitu delarik birusari aurre egiteko proposatu diren osasun neurrien ondorioz (Chakraborty *et al.*, 2020). Honetaz gain, datu masiboak prozesatzeko gaitasuna inoiz baina handiago da gaur egun. Orain dela urte gutxi batzuk egingaitzak ziren gauza asko egingarri bilakatu dira *hardware* atalean eman diren aurrerapen teknologikoen ondorioz (Micikevicius *et al.*, 2018). Honekin batera, sare neuronaletan oinarritutako teknikei esker *word embedding* bezala ezagutzen diren hitzen errepresentazio aberatsak lortzeko gaitasuna asko garatu da (Mikolov *et al.*, 2013; Devlin *et al.*, 2019). Hortaz, *Twitter* bezalako sare sozial batean ematen diren elkarrizketa publikoak monitorizatzek gizarteak gertaera desberdinen aurrean dituen iritziak neurtzeko aukera ezin hobea eskaintzen du. Honetaz gain, aldaketa ugari garaitan sare sozialen esangura oraindik gehiago nabarmentzen da eta oso erabilgarriak direla erakutsi dute aurretik eman diren beste fenomeno eta ekintza batzuk izan dituzten eragin sozialak aztertzeko (Buntain *et al.*, 2016; Wang eta Zhuang, 2017).

Artikulu honetan, *Twitter* sare sozialak eskaintzen duen informazioaren bitartez euskal komunitateak COVID-19aren pandemiaren aurrean izan duen erantzuna aztertu nahi dugu, honela pandemiak euskal gizartearen izan duen

eragina hobeto ulertu ahal izateko. Azterketa hau aurrera eramateko, 2019ko irailetik 2021eko otsaila arte, Twitterreko euskal komunitateak egindako txioak jaso eta aztertu ditugu Hizkuntzaren Prozesamenduko teknika desberdinak erabiliz. Pandemiak garai desberdin ugari izan dituenek: konfinamendua, alarma egoera, normalitate berria... jasotako txioak multzo desberdinetan banatu ditugu gertaera desberdinek izan duten eragina zehatzago aztertu ahal izateko. Analisi honetan, termino adierazgarriek pandemian zehar izan duten aldakortasuna aztertzen dugu hauen testuinguruarekin batera.

2. Arloko egoera eta ikerketaren helburuak

COVID-19aren pandemia hasi zenetik artikulu asko argitaratu dira *Twitter* sare sozialeko jardura monitorizatzen dutenak. Orokorrean monitorizazio hau komunitate edo herrialde zehatzetan finkatu da (Ordun *et al.*, 2020; Lyu *et al.*, 2020), nahiz eta gaiari buruzko desinformazioa modu globalean aztertu dutenak ere egon diren (Chen *et al.*, 2020; Shahi *et al.*, 2021). Sare sozialetan dagoen testu ez egituratua prozesatzeko Hizkuntzaren Prozesamenduko teknika desberdinak erabiltzen dira. Ordun *et al.* (2020) lanean teknika ez gainbegiratuak erabiltzeko proposamena egiten dute. COVIDaren inguruan tratatu diren gai desberdinak nabarmentzeko *Topic-Modeling* teknika erabiltzen dute LDA algoritmoa aplikatuz (Blei *et al.*, 2003). Ondoren antzeman dituzten gai hauek modu bisualean erreprezentatzeko UMAP (McInnes *et al.*, 2018) dimentsio txikiagotze teknika erabiltzen dute. Horrez gain, euskarazko Twitter komunitatearen gaineko eduki analisiak burutu dira ere, *Topic-Modeling* bitarteko eduki analisia zein interakzioen analisia eginez (Fernandez de Landa *et al.*, 2019).

Badaude beste lan batzuk teknika gainbegiratuak erabiltzen dituztenak sare sozialetako edukia aztertzeko (Chen *et al.*, 2020; Shahi *et al.*, 2021; Müller *et al.*, 2020). Sailkapen algoritmo gainbegiratuak entrenatu ahal izateko aurretiko eskuzko lana beharrezkoa da, hau da, pertsona aditu batek adibide desberdinak etiketatu beharko ditu, ondoren ikasketa automatikoko algoritmoak hauen gainean aplikatu ahal izateko. Algoritmo ugari daude modu gainbegiratuan lan egiten dutenak baina azkenaldian erabilienak aurretik entrenatuak izan diren hizkuntza eredu erraldioen berfintzean oinarritzen dira. Hau da hain zuzen ere Müller *et al.* (2020) lanean proposatzen dutena, bertan, COVIDaren inguruko txioetara BERT (Devlin *et al.*, 2019) hizkuntza eredu egokitzen dute, 5 sailkapen ataza desberdinetan hobekuntza nabarmena erakutsiz.

Hortaz, lan honen helburua, Hizkuntzaren Prozesamenduak eskaintzen dituen teknika ez-gainbegiratuak erabiliz Twitterreko euskal komunitatean COVID-19aren pandemiak izan duen eragina aztertzea da. Azterketa hau aurrera eramateko sare sozial horretako erabiltzaileen euskarazko txioak masiboki bildu eta denboraren arabera ordenatu dira. Pandemiaren eragina aztertzeko, denboraren arabera edukia nola aldatu den aztertu da, horretarako testuetan azaltzen diren hitz eta emoji aldaketa kuantitatibo zein kualitatiboak baliatuz. Azterketa kuantitatiboan, aldakortasunaren erakusletzat terminoek garai desberdinetan izan duten maiztasunaren aldaketa erabili da, maiztasunen erregresio linealaren malden balioak erabiliz. Azterketa kualitatiboan, hitzen esanahia hauen testuinguruak zehazten duela premisa bezala hartuz, pandemiaren garai desberdinetan hitz eta emoji adierazgarrienek esanahian izan duten bilakaera aztertu da hitz bektore trinkoak erabiliz.

3. Ikerketaren muina

Twitter sare sozialean aritzen den euskal komunitatea ikertzeko, bertatik datuak erauzi dira. Sare sozial honek euskarazko komunitate aktiboa dauka eta bertatik testuan oinarritutako datu kantitate erraldoiak lortu daitezke. Era honetan datuen bilketa eta analisia nola egin den azalduko da. Lehenengo zatian datuen bilketaren nondik norakoak azalduko dira, bilketa denboraren arabera nola antolatu den eta zenbat txio lortu diren. Datuak lortuta, hauen gaineko analisia nola egin den azalduko da, informaziotik jakintza nola atera dugun azalduz. Horretarako bi teknika erabili dira, bata kuantitatiboa edo hitzen maiztasunean oinarritua eta, bestea, kualitatiboa edo hitzen esanahian oinarritua.

3.1. Datuen erauzketa

Gizartea aztertu eta bertan ematen diren dinamika latenteak identifikatzeko asmoarekin datu iturri oparotzat jotzen dugu Twitter sare soziala. Euskal komunitatea bezalako komunitate txikien azterketarako ere datu kopuru masiboak eskaintzen baitugu sare sozial honek. Gizartearen isla zehatza izateko definitzen den lagin orok bere errore tartea daukan moduan, datuak biltzeko modu honek ere mugak edukiko ditu, datu bilketa Twitterreko erabiltzaileak diren pertsonetara mugatzen baita. Hala ere, bertan identifikatzen diren erabiltzaile guztietatik datuak denbora errealean masiboki erauzteko aukera ematen du, datu kantitate erraldoiak batuz. Era honetan, 10,000 euskal erabiltzaile

baino gehiago identifikatu dira, hauetatik 4M txio pertsonal eta 4M birtxio lortuz. Datuak 2021. urteko otsailean lortu ziren Twitterreko API publikoa baliatuz.

Bildutako datuak 5 garai ezberdinen arabera zatitu dira, denboraren arabera edukia azterketa bat ahalbidetzeko asmoz. 1. taulan ikusi daitekeen moduan zatiketa bakoitza momentu deigarri batzuekin identifikatu da, garai tematiko horien edukia aztertzea helburu izanik. Gainera, momentu tematiko bakoitzaren hasiera eta bukaera datak zein txio eta birtxio kopuruak ikusi daitezke. Lehenik eta behin, 0. momentu bat finkatu da, pandemiaren aurreko garaia datuak edukitzera bidean, 2020 baino lehenagoko datuak hartuz. 1. momentuan 2020 urtea hasi eta konfinamendua arteko garaiak hartuko du. 2. momentua etxean konfinatuta egoteko aginduak iraun bitarteko garaiak hartuko du, momentu zentral hau garrantzitsuen kontsideratuko da. 3. momentuak konfinamendua amaitu eta bigarren alarma egoera arteko garaiak hartuko du, normalitate berri bezala definitu da momentu hau. Azkenik, 4. momentuak bigarren alarma egoeratik datuen erauzketa arteko momentua hartu du, 2021eko Otsailean.

Momentua	Hasiera data	Bukaera data	Txio pertsonalak	Birtxioak (RT)
0. 2020 aurretik	2019/09/01	2019/12/31	224.169	275.042
1. Konfinamendu aurretik	2020/01/01	2020/03/14	155.302	196.500
2. Konfinamendua	2020/03/15	2020/06/21	296.627	349.368
3. Normalitate berria	2020/06/22	2020/10/24	343.372	362.279
4. Bigarren alarma egoera	2020/10/25	2021/01/31	415.388	347.533

1. Taula: Erauzitako euskarazko txioen antolaketa denboraren arabera.

3.2. Datuen analisia

Datuen analisirako txio pertsonal eta birtxioak kontutan hartzea erabaki da, erabiltzaile bakoitzak denbora lerroan publiko egiten duen edukia parte baitira bi elementu hauek. Era honetan, norberaren txioen zein konpartitzen diren txioen (birtxioak) testuetan oinarritu da ikerketa hau. Testuaren osagarri nagusia diren hitzez gain emojiak ere kontutan hartu dira, bestelako hitzen antzera esanahia duten karaktere edota karaktere multzoak baitira. Geroz eta ohikoagoak bilakatzen ari diren emoji hauek ez dira hiztegiko definizio zehatza duten hitzak, baina agertzen diren testuinguruari esker beren esanahia ezagutu daiteke. Horregatik, Twitterreko euskal komunitatearen baitan pandemiak eragindako eraginak aztertzeko terminoei so egin zaie, hauen erabileran zentratuz. Denboraren arabera terminoen erabilera nola aldatu den erakutsiko da, aldaketa hauetan pandemiaren eraginera aztertuz aldi berean.

Erabiltzaileen adierazpenetan gertatutako aldaketak neurtzeko, terminoen erabilera maiztasun zein esanahian izan diren aldaketak kontutan hartu dira. Fokua bi aspektu horietan jarritz, analisi kuantitatibo eta kualitatiboa jorratu dira, erabilera kantitateak eta erabilera moduak aztertzeko teknika bezala aurkeztuz. Alde batetik, terminoen maiztasunak momentuen arabera nola aldatu diren aztertu da, pandemia garaian aurrera joan ahala garrantzi gehien hartu duten terminoak zein garrantzi gehien galdu dutenak jasoz. Beste alde batetik, termino zehatzek denboraren arabera izan dituzten aldaketa semantikoak ere aztertu dira, pandemiak esanahietan eduki duen eragina erakutsiz.

3.2.1. Metodo kuantitatiboa: terminoen maiztasunaren aldaketa denboran zehar

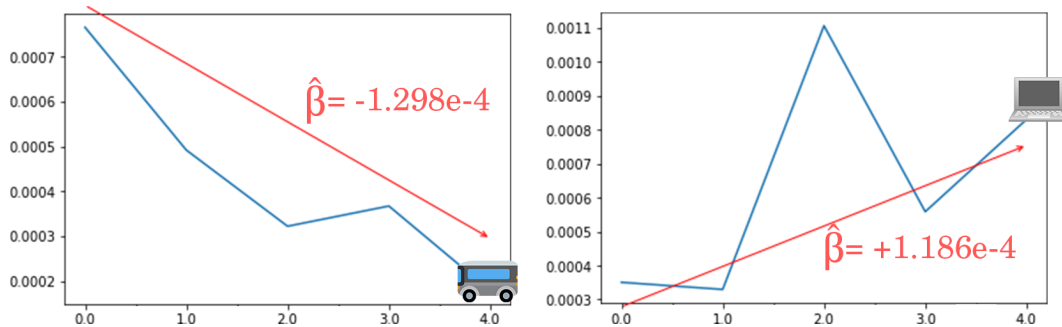
Terminoaren erabilera maiztasunaren aldakortasunean oinarritutako metodo kuantitatibo honekin, Twitterreko euskal komunitateak adierazteko erabili dituen terminoen azterketa egin nahi da: pandemiaren eraginez zein termino hasi den gehiago erabiltzen eta zein termino gutxiago. Horrela, denboran zehar terminoen erabilera maiztasuna nola aldatu den ezagutzeko asmoarekin, termino guztien maiztasunen erregresio linealeak kalkulatu dira. Horrekin batera, erregresio linealeko balio horiek ordenatu eta balio altuenak zein baxuenak hartuko dira, erabileran igoera eta jaitsiera identifikatzeko modu bezala.

Lehen pausu modura, termino guztiak lematizatu dira IXA pipes (Agerri *et al.*, 2014) erabilita, euskararen aberastasun morfologikoa sinplifikatzeko asmoarekin. Terminoen lematizazioa burutu ostean, 1. ekuazioan ikusi daitekeen bezala, termino bakoitzaren maiztasuna kalkulatzeko da zehaztu dugun pandemiaren momentu bakoitzerako. Termino bakoitzeko 5 maiztasun ezberdin edukiko genituzke, maiztasun balio bat momentu bakoitzeko. Terminoen joera kalkulatzeko, 2. ekuazioko erregresio linealeko sistema ebatzi da, termino bakoitzaren momentu ezberdinetako maiztasunen balioetan oinarrituz. Ekuazio honetan, $x_0 \dots x_N$ balioek denbora errepresentatzen dute, $y_0 \dots y_N$ balioek tarte bakoitzean termino bakoitzak duen maiztasuna eta N tarte kopuru totala da. Horrela, termino bakoitzaren erregresio lineal maldaren balioa ($\hat{\beta}$) izango da termino horren joeraren erakuslea.

$$y = \frac{\text{Zenbat txioetan azaltzen da terminoa}}{\text{Txio kopuru totala}} \quad (1)$$

$$\hat{\beta} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^N (x_i - \bar{x})^2} \quad (2)$$

Malda edo joeran ($\hat{\beta}$) balio positiboak erabileran gora egin duela adieraziko du, balio negatiboak erabileran behera egin duela adieraziko duen bitartean. 1. irudian adibide gisa ikusi daitezke ordenagailu (💻) eta autobusaren (🚌) emojiaren joera. Momentu bakoitzean (ardatz horizontala) duten maiztasun balioen (ardatz bertikala) erregresio linealaren maldaren balioari esker ($\hat{\beta}$), ordenagailua (💻) emojiaren erabilera handitu eta autobusaren (🚌) emojiaren erabilera txikitu dela erakutsi daiteke.



1. Irudia: Ordenagailu (💻) eta autobusaren (🚌) emojiaren joeraren kalkulua

Hortaz, corpusean azaldu diren termino guztiak beren joeraren arabera ordenatu dira ranking batean. Ranking honetatik, goranzko joera altuena duten 10 terminoak, zein, beranzko joera altuena duten 10 terminoak hartu dira, gehien hazi eta murriztu diren terminoak lortuz hurrenez hurren. Terminoen joeren aldakortasunaren emaitzak 2. taulan ikusi daitezke, gehien hasi direnak eta gehien murriztu direnak jasoz.

Terminoa	Joera	Terminoa	Joera
covid	7,31	aurkezpen	-4,60
neurri	6,82	greba	-4,43
osasun	6,17	feminista	-4,42
pandemia	6,13	jardunaldi	-4,23
txerto	5,02	altsasu	-4,14
positibo	3,77	antolatu	-3,83
online	3,44	pentsio	-3,80
kasu	3,20	hitzaldi	-3,48
zaindu	3,07	emakume	-3,22
konfinamendu	2,80	elkartasun	-3,20
birus	2,79	katalunia	-3,16
krisi	2,78	areto	-3,11
ostalaritza	2,75	aurkeztu	-3,11
#covid19	2,70	egitarau	-3,09
maskara	2,60	manifestazio	-2,79

(a) Aldakortasun positibo handiena.

(b) Aldakortasun negatibo handiena.

2. Taula: Hitzen erabileraren aldakortasuna denboraren arabera.

Erabileran gora egin duten terminoak 2a taulan ikusi daitezke, hauetako batzuk pandemiarekin zuzenean erlazionatuta egonik (covid, neurri, osasun, pandemia, txerto, positibo, kasu, zaindu, birus, #covid19) edo honen albo-eraginekin (online, konfinamendu, krisi, ostalaritza, maskara). Azpimarratu beharra dago, hitzen maiztasunaren aldakortasunean joera positibo handiena izan duten termino gehienak zuzeneko erlazioa dutela pandemiaren

gaiekin. Horrek argi erakusten du pandemiaren eragina handia izan dela euskarazko Twitter komunitatean. Gainera, osasunarekin erlazioatutako terminoez gain, albo eraginekin erlazioatutako terminoak dauzkagu: online, konfinamendu, krisi, ostalaritza, maskara. Albo eragin hauek gizartearen iruditeri kolektiboan utzitako arrastoak ikusi daitezke hitz hauetan: online moduen areagotzea komunikazioan, ostalaritzan murrizketa modu eta ordutegietan, maskararen erabilera egunerokotasunean... Krisi hitzaren maiztasunaren gorakada ere adierazgarritzat jo daiteke, egoera bera definitzeko baliatu baitute hitza.

Bestalde, erabilera maiztasunean jaitsiera nabarmenena izan duten terminoak 2b taulan ikusi daitezke, honakoan terminoek gai politikoekin (greba, feminista, Altsasu, pentsio, emakume, elkartasun, Katalunia, manifestazio) zein ekimen kolektiboekin (aurkezpen, jardunaldi, antolatuta, hitzaldi) erlazioatuta daudelarik. Hortaz, Twitter sare sozialean ohikoak ziren gai politikoaren beherakada nabarmena izan dela baieztatu daiteke. Feminismoa (feminista, emakume), ekonomia (greba, pentsio) eta bestelako gaiek (Katalunia, Altsasu) euskal komunitatearen baitan garrantzia galdu dutela ikusi daiteke, fokoa tokiz aldatu dela erakutsiz. Era berean, ekintza politikoarekin edo aldarrikapenarekin zerikusia daukaten praktikek ere zeresana galdu dutela ematen du. Teoria zein praktika politikoaren garrantzi galera nabarmena erakusten du honek, batez ere aldarrikapen politikoarekin guztiz loturiko sare sozial batean.

Laburbilduz, deigarria da plano ideologikoan indartsuak diren zenbait kontzeptuen erabilera txikitu dela, osasuna bezalako kontzeptuek zentraltasuna lortu duten bitartean. Era berean, maiztasuna handitu duten zenbait hitz egunerokotasunean zehar ematen diren praktikekin erlazioatuta daude, abstrakzioetik errealitate hurbilera igaroz. Are gehiago, eguneroko bizitzaren arauketarekin erlazioatuta dauden hainbat terminoen maiztasuna areagotu da: maskarila eramateko beharra, tabernetan egoteko baimen edo harremantzeko moduak, distantziara komunikatzeko era edota etxean itxita egotearen agindua. Esan daiteke biopolitikarekin (Foucault, 2009) edo gizakiaren egunerokotasuneko ekintzen gaineko arautzearekin erlazioatutako gaietara igaro dela fokua. Hots, gai orokorretan zentratetik, egunerokotasunean eragina daukaten ekintzetan gehiago zentratzera igaro da euskarazko Twitterreko komunitatea.

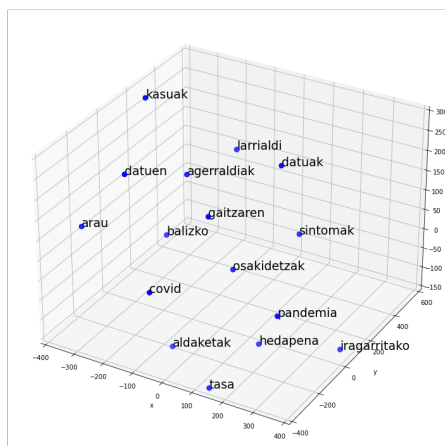
3.2.2. Metodo kualitatiboa: terminoen esanahiaren aldakortasuna denboran zehar

Azterketa kualitatiboaren asmoa, terminoen esanahiaren aldaketa denboran zehar nola eman den aztertzea da. Hitzek esanahia aldatzen dute gizartearen beharren arabera, hizkuntza egoera zehatzetara moldatuz. Pandemiaren garaian aldaketak, egon badira, nola eman diren ezagutzeko, hitzen testuinguruan oinarritutako mapaketa bektorialak erabiliko dira. Testuinguruan oinarritutako mapaketa hauek, terminoen esanahi semantikoa mantentzeko kapazak dira, hitz baten ezaugarriak beren ingurukoek definitzen dituelako (Firth, 1957). Honi esker, emojiak daukaten esanahi atxikia ezagutzeko parada edukiko da ere, hauek ere testuingurua baitaukate. Asmoa, momentu bakoitzeko eredu bat sortzea da, terminoen ezaugarriak denboraren arabera moldatu diren edo ez ikusi eta aztertzeko.

Hitzen zein emojiaren esanahia ezagutzeko, word2vec (Mikolov *et al.*, 2013) teknika erabili da, hitz eta emojiaren erlazioak erakutsiko dituen mapaketa bektorial bat osatzeko. Teknika honekin, terminoen irudikapen bektorial trinkoak lortzen dira, terminoak espazio bektorial berdinean kokatuz testuinguruaren arabera. Irudikapen bektorialari esker semantikoki antzekoak diren terminoak lortu daitezke, antzeko terminoek antzeko irudikapen bektoriala baitute (Harris, 1954). Era horretan, termino bakoitzaren antzeko bektoreak ezagututa antzekoak diren hitzak ezagutu daitezke, hots, antzeko esanahia duten hitzak lortuko dira. Termino bakoitzaren inguruko hitzek honen esanahia zein den erakutsiko dutenez, hitzen esanahia definitzeko ekingo zaio momentu bakoitzeko. Momentu bakoitzean termino baten antzeko hitzak zeintzuk diren jasoko dira, horrek aldaketak egon diren edo ez aztertzeko aukera emango du.

Hitzen esanahia denboraren arabera nola moldatu den ezagutzeko asmoarekin, 5 momentu ezberdinetarako hitzen bektore irudikapen bana lortu da. Hauetako bakoitzak, momentuaren araberrako ezaugarri semantikokoak baturatu ditu, irudikapen independenteak sortuz. Irudikapen dentso bakoitza sortzeko, 5 token edo hitzeko leihoa erabili da hitzen ezaugarriak 100 dimentsiotan kontzentratuz (window_size=5, 100 dimentsio). Horrela, terminoen irudikapen bektorial trinkoen 5 eredu ezberdin lortu dira, momentu bakoitzeko terminoak dagozkien esparru bektorialean kokatuz testuinguruaren arabera. Teknika honekin lortutako emaitzen adibide bat 2. irudian ikusi daiteke, bertan *Covid* hitzaren eta semantikoki gertuen dauden 16 hitzen irudikapena erakusten da. Horrela, aukeratutako hitzarekin erlazioatutako hitzak lortzen dira, hauek 3. momentuan *Covid* hitzak hartzen duen esanahia definitzen lagunduko dute.

Azterketa kualitatibo hau egiteko, erabilera maiztasunean gorakada nabarmena eduki duten terminoen artean, esanahian aldaketa argiena izan dutenak aukeratu dira: positibo, kasu eta segurtasun. Maskaradun emoji (👤)



2. Irudia: Covid hitzetik gertuen dauden hitzak 3. momentuko bektore espazioan.

ere aukeratua izan da azterketa kualitatiborako, erabilera maiztasun aldaketa handienetakoa duten emojiaren artean kokatzen baita. Era honetan, aukeratutako terminoen gertuko hitzak lortu dira momentu bakoitzerako, horretarako terminoen irudikapen bektorial trinkoak erabiliz. Terminoren aldaketak nola eman diren irudikatzeke, termino bakoitzaren momentu bakoitzeko termino antzekoak 3. taulan ikusi daitezke.

Terminoak	Erlazionatutako hitzak - Denboraren arabera esanahia
positibo	0. teknika, zailtasun, kontzeptu, ikusmen, gertakizun
	1. h5n8, arrazoia, egoiliarri, aktiboko, ontzat
	2. kutsatu, koronabirus, kasu, pcr, infektatu
	3. koronabirus, negatibo, kutsatu, positiboan, pcr
	4. kutsatu, koronabirus, ospitaleratze, biztanleko, atzemandako
kasu	0. afera, galdera, une, kontu, zentzu
	1. oraingo, garai, mota, legegintzaldi, afera
	2. positibo, koronabirus, kasuak, pcr, kutsatu
	3. positibo, koronabirus, proba, kutsatu, test
	4. positibo, kasuak, test, hildako, kutsatu
segurtasun	0. sistemak, hondakinen, murrizteko, administrazio, legearen
	1. prebentzio, kontrol, murrizteko, koronabirusak, kontrolatzeko
	2. prebentzio, distantzia, higie, errespetatu, beharrezko
	3. prebentzio, higie, zorrotz, neurriekin, protokolo
	4. prebentzio, higie, malgutu, ezarritako, mugikortasun
🧑	0. #kutsadura, albistegitan, #nipenani gloria, #bizitzaerdigunera, margaret
	1. isurketa, filtro, argindar, tximinia, laino
	2. saihesteko, bestee kiko, musukoa, maskara, derrigorrezkoa
	3. #maskarabeti, aforo, #euskotrenmetrobilbao, edukiera, hidroalkoholikoa
	4. bidalketa, #htxonline, #getxo, #udalsarea2030, #amasavillabona

3. Taula: Terminoak eta euren esanahien azterketa denboraren arabera.

Positibo terminoa hitz desberdin askorekin erlazionatuta dago (*teknika, zailtasun, kontzeptu, h5n8, arrazoia...*) 0 eta 1 momentuetan. Bestalde, 2, 3 eta 4 momentuen artean *kutsatu* eta *koronabirus* terminoekin erlazio zuzena eta iraunkorra izatera igaro da, pandemiaren eragina agerian utziz. Termino honek erlazio zuzena du pandemiaren gaiekin, gaitzarekin kutsatu diren pertsonak definitzeko baliatua izan baita, pandemia aurreko esanahia guztiz irauliz.

Kasu terminoak, 0 eta 1 momentueatan *afera* zein *kontu* bezalako, hitzekin erlazioa dauka, horrez gain ere, denborarekin ere zerikusia duela ikusi da (*garai, une, oraingo*). Ostera, 2, 3 eta 4 momentuetan pandemiarekin

erlazionatutako hitzek hartzen dute tokia (*koronabirus, kasu, kutsatu*). Horrez gain, azpimarratzekoa da *positibo* hitza dela gertuen daukana, kasu positibo bigramaren agerpenaren eraginez ziurrenik. Gainera, *positibo* hitzarekin gertatu bezala, erlazio zuzena izan du pandemiaren gaiekin, honako honetan ere esanahia iraulia izan delarik.

Segurtasun hitza *legea, administrazioa* edo *sistema* bezalako hitzekin batera kokatzen da 0. momentuan, estatuaren gestioarekin erlazionatutako terminoak. Aurrera egin ahala, 1. momentuan kontrolarekin erlazionatutako esanahia izatera igaro da (*kontrola, kontrolatzeko, murrizteko*), beti ere pandemiarekin erlazionatuta (*koronabirus*). Esan beharra dago, 2, 3 eta 4 momentuen artean terminoa oso lotuta egon dela *prebentzio* zein *higiene* terminoekin, norberaren gaineko kontrolarekin erlazionatuta, biopolitikarekin erlazio estua erakutsiz berriz ere.

😊 emojiari erreparatuz, 0 eta 1 momentuetan emojiak ingurumenaren kutsadurarekin erlazionatutako terminoekin batera azaltzen da (*#kutsadura, filtro, tximinia, isurketa, laino...*). Denboran aurrera egin ahala esanahia aldatzen doa 2 eta 3 momentuetan pandemiarekin zerikusi zuzena daukaten terminoen alboan azaltzen da (*aforo, edukiera, hidroalkoholiko...*) eta baita gaitzaren kutsadura ekiditeko musukoa eramateko behararekin (*sahiesteko, derrigorrezkoa, #maskarabeti...*) ere. Horrela, ikusi daiteke emojiaren ezaugarri semantikoak aldatu egin direla pandemian aurrera egin ahala, ingurumenaren kutsadura gaiekin erlazioa edukitzetik pandemiaren kutsatu eta gaiekin erlazioa izatera igarotzen baita.

Terminoek esanahien aldakortasunari begiratu, aukeratutako terminoetan pandemiaren lorratza argia dela erakutsi da. *Positibo, kasu* eta 😊 terminoak egoeraren adierazle ezin hobeak dira, pandemiaren inguruko kontakizunekin zuzenean erlazionatutako terminoak diren heinean, esanahian emandako aldaketa nabaria da. Bestalde, *segurtasun* bezalako terminoari so eginda, osasun egoerarekin baino egunerokotasuneko ekintzen arauketarekin zerikusi gehiago dauka, bizimoduan pandemiak duen eraginarekin erlazio zuzena erakutsiz. Termino hauek maiztasunean gora egiteaz gain, esanahien azterketari esker ikusi da terminoaren adiera ere pandemiari guztiz lotua dagoela, maiztasunaren gorakada pandemiarekin guztiz erlazionatuta dagoela frogatuz. Honekin, gizartearen imajinario kolektiboan pandemiak eragindako aldaketak erakutsi dira. Euskal erabiltzaileen idazteko moduan emandako aldaketa hauek, aldaketa sakonen erakusle izan daitezke, hitzei konnotazio berria emateraino ailegatuz.

4. Ondorioak

Hizkuntzaren Prozesamenduak eskaintzen dituen teknika ez-gainbegiratuak erabiliz Twitterreko euskal komunitatean COVID-19aren pandemiak izan duen eragina aztertzea lortu da lan honetan. Azterketa hau aurrera eramateko sare sozial horretako erabiltzaileen euskarazko txioak masiboki bildu eta denboraren arabera ordenatu dira, hizkuntza gutxituen ingurune batean ere datu kantitate handiak lortuz. Pandemiaren eragina neurtzeko denboraren arabako edukien aldakortasunean zentratzea baliagarria dela frogatu da, testuetan azaltzen diren hitz zein emojiak aldaketa kuantitatibo zein kualitatiboak horretan lagungarri direla erakutsiz. Azterketa kuantitatiboan, terminoek garai desberdinetan izan duten maiztasunaren aldaketa kontutan hartuz, pandemia garaiko termino nabarmenenak lortzeko teknika egokia izan dela ikusi da. Azterketa kualitatiboan, pandemiaren garai desberdinetan hitz eta emoji adierazgarrien esanahiak hitzen sektore trinkoen bidez lortzea ahalbiatu da, gizarte aldaketen nolakotasuna hitzen bidez erakustea lortuz.

Gizakion bizitzetan pandemiak hartu duen zentraltasunak, gizarte garaikidean inflexio puntu argi bat zehaztu du, iragana eta etorkizuna desberdinduz. Euskarazko Twitter komunitatean ere hala gertatu dela nabarmena da, hitzen maiztasunean emandako aldaketek zein esanahien aldakortasunak hala frogatu baitu. Horrela, gai politiko orokorretan zentratzetik, egunerokotasuneko ekintzen gaineko arautzearekin erlazionatutako gaietara igaro du fokua euskarazko Twitterreko komunitateak. Bestalde, hitzen esanahien aldakortasunak ere pandemiaren eragina islatu du, egoera honek hitzen esanahia aldatzeko adina indar izan duela erakutsiz.

5. Etorkizunerako planteatzeko den norabidea

Lehenik eta behin, datu bilketari esker lortu diren datu kopuru erraldoiekin eduki azterketa desberdinak egin daitezke, hala nola *Topic-Modeling*, interakzioetan oinarritutako azterketak... Horrez gain, lan honetan garatu den metodologiak, bestelako hizkuntza gutxituetan aplikatu daitezke. Datu bilketa zein analisiak hizkuntzarekiko inpendenteak izanda, aplikagarritasun erraza dauka aldaketa askorik egin gabe. Epe luzera begira, Twitterreko komunitatea eta euskal komunitatearen arteko loturak aztertzea izango da helburu garrantzitsua, datu iturri kantitate handien laguntzaz, ikerketa soziala egiteko modu berri eta zehatzagoak proposatzeko.

6. Erreferentziak

- Agerri, Rodrigo, Josu Bermudez, eta German Rigau. 2014. IXA pipeline: Efficient and Ready to Use Multilingual NLP tools. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation*, volume 2014, 3823–3828.
- Blei, David M, Andrew Y Ng, eta Michael I Jordan. 2003. Latent dirichlet allocation. *the Journal of machine Learning research* 3.993–1022.
- Buntain, Cody, Jennifer Golbeck, Brooke Liu, eta Gary LaFree. 2016. Evaluating public response to the Boston Marathon bombing and other acts of terrorism through Twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 10.
- Chakraborty, Tanusree, Anup Kumar, Parijat Upadhyay, eta Yogesh K Dwivedi. 2020. Link between social distancing, cognitive dissonance, and social networking site usage intensity: a country-level study during the COVID-19 outbreak. *Internet Research* .
- Chen, Emily, Kristina Lerman, eta Emilio Ferrara. 2020. Tracking social media discourse about the COVID-19 pandemic: Development of a public coronavirus twitter data set. *JMIR Public Health and Surveillance* 6.e19273.
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee, eta Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–4186.
- Fernandez de Landa, Joseba, Rodrigo Agerri, eta Iñaki Alegria. 2019. Large Scale Linguistic Processing of Tweets to Understand Social Interactions among Speakers of Less Resourced Languages: The Basque Case. *Information* 10.212.
- Firth, John Rupert, 1957. A synopsis of linguistic theory 1930-1955 in studies in linguistic analysis, philological society.
- Foucault, Michel. 2009. *Nacimiento de la biopolítica: curso del Collège de France (1978-1979)*, volume 283. Ediciones Akal.
- Harris, Z. 1954. Distributional structure. *Word*, 10(23): 146-162. .
- Lyu, Hanjia, Wei Wu, Junda Wang, Viet Duong, Xiyang Zhang, eta Jiebo Luo. 2020. Social Media Study of Public Opinions on Potential COVID-19 Vaccines: Informing Dissent, Disparities, and Dissemination. *arXiv preprint arXiv:2012.02165* .
- McInnes, Leland, John Healy, Nathaniel Saul, eta Lukas Großberger. 2018. UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software* 3.861.
- Micikevicius, Paulius, Sharan Narang, Jonah Alben, Gregory Diamos, Erich Elsen, David Garcia, Boris Ginsburg, Michael Houston, Oleksii Kuchaiev, Ganesh Venkatesh, eta others. 2018. Mixed Precision Training. In *International Conference on Learning Representations*.
- Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S Corrado, eta Jeff Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality. In *Advances in Neural Information Processing Systems*, ed. by C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, eta K. Q. Weinberger, volume 26. Curran Associates, Inc.
- Müller, Martin, Marcel Salathé, eta Per E Kummervold. 2020. COVID-Twitter-BERT: A natural language processing model to analyse COVID-19 content on twitter. *arXiv preprint arXiv:2005.07503* .
- Ordun, Catherine, Sanjay Purushotham, eta Edward Raff. 2020. Exploratory analysis of COVID-19 tweets using Topic Modeling, UMAP, and Digraphs. *arXiv preprint arXiv:2005.03082* .
- Shahi, Gautam Kishore, Anne Dirkson, eta Tim A Majchrzak. 2021. An exploratory study of COVID-19 misinformation on Twitter. *Online social networks and media* p. 100104.
- Wang, Bairong, eta Jun Zhuang. 2017. Crisis information distribution on Twitter: a content analysis of tweets during Hurricane Sandy. *Natural hazards* 89.161–181.