



IKER  
GAZTE  
NAZIOARTEKO  
IKERKETA EUSKARAZ

## II. IKERGAZTE

NAZIOARTEKO IKERKETA EUSKARAZ

2017ko maiatzaren 10, 11 eta 12  
Iruñea, Euskal Herria

ANTOLATZAILEA:  
Udako Euskal Unibertsitatea (UEU)

### GIZA ZIENTZIAK ETA ARTEA

**Euskarazko ezeztapenaren  
tratamendu automatikorako  
azterketa**

*Bego Altuna, María Jesús Aranzabe  
eta Arantza Díaz de Ilarraza*

127-134 or.

<https://dx.doi.org/10.26876/ikergazte.ii.01.17>

ANTOLATZAILEA:



ELKARLANEAN:



LAGUNTZAILEAK:



# Euskarazko ezeztapenaren tratamendu automatikorako azterketa

Altuna, Begoña eta Aranzabe, María Jesús eta Díaz de Ilarraza, Arantza<sup>1</sup>

Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU). begona.altuna@ehu.eus

## Laburpena

Hizkuntzaren Prozesamenduak testuetako informazioaren erauzketa eta analisisa du helburu. Lan honetan euskarazko ezeztapenaren analisisa eta etiketatzea aurkezten ditugu, denbora-informazioaren prozesamenduan osagarri gisa, ezeztapenak gertaeren faktualitatearen ebazpena eta gertaera horietako entitateen parte-hartzea baldintzatzen baititu. Gaiaren inguruko lanak aztertu ditugu, euskaraz ezeztapena nola gauzatzen den aztertu dugu eta euskarazko ezeztapen-informazioa kodetzeko eskema moldatu dugu. Kodetzearen inguruan hartutako erabakiak ebaluatzeko eta ikasketa-automatikoan baliatuko dugun corpusa sortzeko, eskuzko etiketatze-esperimentu bat egin dugu. Esperimentuaren emaitzetan oinarrituta, errore-analisi bat egin dugu bai atazaren zailtasuna ebaluatzeko, bai etiketatze-gidalerroak hobetzeko.

**Hitz gakoak:** Hizkuntzaren Prozesamendua, ezeztapena, denbora-informazioa, etiketatzea

## Abstract

*Natural language processing aims to the information extraction and analysis. In this paper, we present the analysis and annotation of negation in Basque, as a complement to the temporal information processing, since negation affects event factuality resolution and event role definition. We have analysed the background and we have studied how negation is expressed in Basque and we have normalised that information. In order to evaluate the normalisation scheme and to create a corpus for machine learning, we have conducted a manual annotation experiment. Based on the experiment results, we have analysed the errors to measure the difficulty of the task and to improve the annotation guidelines.*

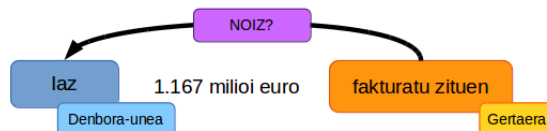
**Keywords:** Natural language processing, negation, temporal information, annotation

## 1 Sarrera

Interneten etengabe sortzen dira testu kantitate handiak. Testu horiek, ordea, era masibo eta automatikoan “ulertu” behar dira gure egunerokotasunean erabilgarri izateko. Hain zuzen ere, Hizkuntzaren Prozesamenduan (HP) testuen ulermen automatikoa bilatzen da testuetako informazioaren erauzketa eta analisisa gauzatuta besteak beste, informazio hori galdera-erantzun sistemetan edo itzulpen automatikoa bezalako atazetan erabili ahal izateko.

Gure ikerketa-arloa denbora-informazioaren prozesamendua da. Arlo horretan gertaerak (gertatzen diren ekintzak eta egoerak), denbora-une eta tarteak, eta horien artean sortzen diren erlazioak aztertzen eta prozesatzen dira. Esaterako 1. irudiko “Iaz 1.167 milioi euro fakturatu zituen” esaldian “fakturatu zituen” gertaera eta “iaz” denbora-unea ditugu. Horien artean denbora-erlazio bat (*noiz?*) sortzen da gertaera data horretan gertatu zela adierazten diguna.

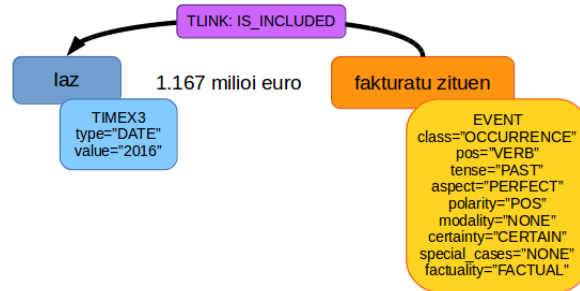
1 Irudia: “Iaz 1.167 milioi euro fakturatu zituen” esaldiko denbora-informazioa



Denbora-informazioa prozesatu ahal izateko, makinek ulertzeko moduan eman behar da. Horretarako, denbora-informazioa eskema baten bidez normalizatu behar da eta informazio hori jasotzen duten

testu-corpusak sortu behar dira. Euskarazko denbora-informazioa kodetzeko, EusTimeML etiketatze-lengoaia (Altuna *et al.*, 2016) sortu dugu TimeML lengoiaian (Pustejovsky *et al.*, 2003) oinarrituta eta denbora-informazioa duen corpusa osatzen ari gara. 2. irudian ikus dezakegu EusTimeMLren bidez kodetuta 1. irudiko esaldiko denbora-informazioa. Hala, denbora-adierazpena TIMEX3 etiketarekin markatu dugu, gertaera EVENTekin eta horien arteko denbora-erlazioa TLINKekin. Bakoitzaren ezaugarriak ere modu normalizatuan adierazi ditugu atributuen bidez. Informazio hori baliatuta, ikasketa automatikoko algoritmoak garatzen dira testu berrietan informazioa detektatu ahal izateko.

2 Irudia: EusTimeML bidez etiketatutako “Iaz 1.167 milioi eruo fakturatu zituen” esaldia



Denbora-informazioa era normalizatuan egoteak ere denbora-lerroen sorrera ahalbidetzen du. Ondoko (1) adibideko esaldietako gertaerak (beltzez) eta denbora-adierazpenak (kurtsibaz) 3. irudian denbora-lerro batean kokatu ditugu kronologikoki antolatuta.

- (1) Dokumentuaren sorrera data: 2013-10-14

Kooperatibak 60,4 milioiko **galerak** izan ditu *2013ko lehen seihilekoan*.

Ez du **jaso maiatzean onartutako** funtsa.

**Gaur bilera** egingo du Mondragoneko kontseiluak.

*Iaz* 1.167 milioi euro **fakturatu** zituen.

3 Irudia: (1) adibidearen denbora lerroa



Denbora-informazioaren prozesamenduan garrantzi handikoa da ezeztapenaren azterketa. Alde batetik, ezinbestekoa da gertaera bat baieztatu edo ezeztatu ematen den jakitea gertaera hori gertatu edo gertatuko den (gertaeraren faktualitate balioa) jakiteko. Bestetik, perpauseko zein zati dagoen zuzenean ukatuta jakiteak gertaera batean parte hartzen duten elementuak erabakitzen lagun dezake. Informazio hori oso erabilgarria izango da denbora-lerroen (Laparra *et al.*, 2015) sorrera automatikoan. Adibidez, (1) adibideko “jaso” gertaerak faktualitate balio negatiboa du eta ez dela gertatu adierazten da. Horregatik ez da 3. irudiko denbora-lerroan agertzen. Bestalde, entitate zehatz batek gertaera batean parte-hartzen ez badu, (2) adibideko<sup>1</sup> “Bob Corker senatari errepublikarra”k bezala, gertaera hori ez da entitate horren denbora-lerroan agertuko.

- (2) [Bob Corker senatari errepublikarra] ere **ez** zegoen pozik.

Lan honetan, euskarazko ezeztapena aztertu dugu tratamendu automatikoari begira. Horretarako, Morante *et al.*-en (2011) eta Blanco and Moldovanen (2011) lanetan oinarritu gara, baita italierarako

<sup>1</sup>Ezeztapena adierazteko ondoko kodea erabiliko dugu: ezeztapen-marka beltzez, fokoa kortxete artean eta irismena azpimarratuta.

egindako lanean (Altuna *et al.*, 2017) ere. Sarrera honen ondoren, 2. atalean artearen egoera eta gure ikerketaren helburuak azalduko ditugu. 3. atalean gure lanaren berri emango dugu: euskarazko ezeztapenaren azterketa 3.1 azpiatalean, ezeztapenaren erabilera denboraren prozesamenduan 3.2 azpiatalean, eta eskuzko etiketatze-saiakera 3.3 azpiatalean. Ondoren, 4. atalean ikerketaren ondorioak adieraziko ditugu eta, amaitzeko, 5. atalean etorkizunerako ildoak deskribatuko ditugu.

## 2 Artearen egoera eta ikerketaren helburuak

Arestian aipatu bezala, ezeztapenaren prozesamenduak gertaera bat gertatu den edo entitate batek presentzia duen jakitea bezalako erabilerak ditu HPan. Ezeztapenaren azterketarako lehen saiakerak biomedikuntzaren arloan egin ziren, arlo horretan entitate edo gertaera baten presentzia edo absentzia informazio garrantzitsua delako. GENIA corpusa (Kim *et al.*, 2008) izan zen lehena eta BioInfer (Pyysalo *et al.*, 2007) eta BioScope corpusek (Vincze *et al.*, 2008) ildo berari jarraitu zioten. Azken hori izan zen ezeztapena eta bere irismena espreski etiketatuta zituen lehena eta gerora ere eredutzat hartu zena.

BioScope corpusaren etiketatze-gidalerroetan oinarrituta, hainbat etiketatze-saiakera eta corpus osatu dira. Morante *et al.*-ek (2011) BioScoperen gidalerroak moldatu eta ezeztapena eta irismenaz gain, gertaerak ere etiketatu zituzten *ConanDoyle-neg* Sherlock Holmesen istorioen corpusean (Morante eta Daelemans, 2012). BioScoperen beste moldaketa batetik sortu zen SFU Review Corpus<sup>2</sup> iritzi-corpusaren ezeztapenaren etiketatzea (Konstantinova *et al.*, 2012). Ikasketa automatikoan, Councillek (2010) BioScope ezeztapena automatikoki etiketatzeko tresna bat entrenatzeko erabili zuen.

PropBank-en (Palmer *et al.*, 2005) ere ezeztapena landu da. Alde batetik, PropBank-en etiketa-multzoak (Bonial *et al.*, 2010), ezeztapena adierazten zuten modifikatzaileentzat (*not*, *n't*, etab.) NEG etiketa zuen. Blanco eta Moldovanek (2011), PropBank oinarritzat hartuta, ezeztapena eta bere fokoa etiketatu zituzten. Lan horretatik sortutako corpusa eta *ConanDoyle-neg* corpusa erabili ziren \*SEM 2012ko *Resolving the Scope and Focus of Negation* ataza partekatuan<sup>3</sup>.

Lan gehienak ingeleserako egin badira ere, beste hizkuntza batzuetan ere ezeztapena etiketatzeko lanak egin dira. Gaztelaniaz, Jiménez *et al.*-ek (to appear) SFU Review Corpusaren gaztelaniazko bertsioaren gaineko etiketatzea egin dute SemEval 2015eko ABSA gidalerroetan<sup>4</sup> oinarrituta, eta medikuntzaren arloan Stricker *et al.*-ek (2015) medikuntza-aurkikuntzen (behaketak, gaixotasunak, etab.) baieztapena edota ezeztapena etiketatu dituzte. Italiararako, biomedikuntzan entitateak (gaixotasunak, drogak, etab.) eta horiek agertzen diren testuinguru negatiboak adierazi dira (Attardi *et al.*, 2015) eta denbora-informazioaren prozesamenduan ezeztapenaren, bere irismenaren eta fokoaren etiketatzea egin da faktualitatearen ebazpena hobetzeko (Altuna *et al.*, 2017).

Euskararen tratamendu automatikoan ezeztapena etiketatu da Euskaren Prozesamendurako Erreferentzia Corpusean (EPEC)<sup>5</sup> bai maila sintaktikoan (Aranzabe, 2008), bai maila semantikoan (Estarrona, 2014).

Lan honetan ere euskarazko ezeztapena aztertu dugu gertaeren prozesamenduaren osagarri gisa. Gure hipotesiaren arabera, ezeztapenaren identifikazioak gertaeren faktualitate balioa erabakitzen lagun dezake, baita denbora-lerroen sorreran ere. Horretarako, euskarazko ezeztapena aztertu dugu eta berori etiketatzeko gidalerroak moldatu ditugu. Azkenean, ezeztapena eskuz etiketatu dugu corpus batean, etorkizunean ezeztapenaren eta faktualitatearen arteko erlazioa aztertu ahal izateko.

## 3 Ikerketaren muinak

Atal honetan euskarazko ezeztapenaren azterketaren eta eskuzko etiketatzearen berri emango dugu. Ezeztapena eta bere ezaugarriak aztertu ditugu, horiek kodetzeko etiketatze-eskema bat definitu dugu eta ondoren, etiketatze-eskema horri jarraituta, bi etiketatzailek corpus bat eskuz etiketatu dute. Azkenik, eskuzko etiketatzearen emaitzak eta etiketatzaileen arteko desadostasunak aztertu ditugu.

<sup>2</sup>[https://www.sfu.ca/~mtaboada/research/SFU\\_Review\\_Corpus.html](https://www.sfu.ca/~mtaboada/research/SFU_Review_Corpus.html)

<sup>3</sup><http://www.clips.ua.ac.be/sem2012-st-neg/>

<sup>4</sup><http://alt.qcri.org/semeval2016/task5/>

<sup>5</sup><http://ixa.si.ehu.es/node/4477>

### 3.1 Ezeztapena euskaraz

Ezeztapena bere eraginpean dagoen proposizioaren egia-balioa aldatzen duen hizkuntzako fenomeno da (Martí *et al.*, 2016). Ezeztapena sintaktikoki gauzatzen da ezeztapen-marka esango diegun elementu sintaktiko eta lexikoen bidez. Markez gain, ezeztapen-marka horien eragin-esparruak, hau da, ukatutako proposizioaren adierazpen sintaktikoak, eta zuzenean ezeztatutako elementuak ere ezeztapen-informazioa osatzen dute.

Ezeztapenaren elementuak kodetzeko etiketatze-eskema bat definitu dugu. Etiketatze-eskema definitzeko, bi lanetan oinarritu gara nagusiki: Morante eta Daelemansenean (2012), ezeztapen-markak eta irismena etiketatzeko eta Blanco eta Moldovanenean (2011) fokoa etiketatzeko. Bi lan horiek batzen dituen italierarako lana (Altuna *et al.*, 2017) ere kontuan hartu dugu, denbora-informazioaren prozesamenduari begira egin baita. Hurrengo azpiataletan ezeztapenean parte-hartzen duten elementuak deskribatuko ditugu, euskararako sortu ditugun etiketatze-gidalerroetako sailkapenari jarraituz.

#### 3.1.1 Ezeztapen-markak

Gure lanean adberbioek (*ez* (3), *salbu*), *ezin* aditzak (4), postposizioek (*-rik gabe*), menderagailuek (*ezik*, *ezean*) (5) eta juntagailuek (*eza*) adierazten duten ezeztapena landu dugu. Aditz batzuek (*ekidin*, *ukatu*, etab.) eta izen batzuek (*debekua*, *ukazioa*, etab.) ere proposizio baten ezeztapena adieraz dezakete, baina ez ditugu lehen urrats honetan aztertu.

(3) Garraio-ministroak hegazkin europarrak **ez** erosteko eskatu die aire-lineei.

(4) Hori gertatzen bada, Fagor Etxetresnek **ezingo** lukete aurrera egin.

(5) Finantzazio hori lortu **ezean**, kooperatiba Konkurtso Legearen baitan sartu beharko litzateke.

Ezeztapen-marka horiei NEG\_CUE etiketa esleitu diegu. Nabaria da, gainera, ezeztapen-marka batzuek (*ezin* (4) eta *eza*, adibidez) ezeztapenaz gain informazio gehiago dutela; *ezinen* kasuan ezeztapena eta ahala adierazten dira. Erabiltzen ditugun tresnak direla eta, token<sup>6</sup> mailan lan egiten dugunez, ordea, token osoa etiketatu dugu ezeztapen-marka bezala.

#### 3.1.2 Ezeztapenaren irismena

Ezeztapenak proposizio baten egia-balioa aldatzen duela aipatu dugunez, ezeztapen-markek proposizio hori adierazten duten testu-zatiari eragingo diote. Testu-zati horri deritzogu ezeztapenaren irismena. Esaterako, (6) adibidean “hegazkin europarrak erosteko” da ezeztapenaren irismena.

(6) Garraio-ministroak hegazkin europarrak **ez** erosteko eskatu die aire-lineei.

Ezeztapen-marka, orokorrean, irismenetik kanpo geratuko da, ez baitu bere polaritatea aldatzen, baina *ezin* aditzaren kasuan (7), esaterako, barruan hartzea erabaki dugu ezeztapenaz gain informazio gehiago duelako.

(7) Oso larri dago, 800 milioiko zorrari buelta eman **ezinik**.

Ezeztapenaren irismenari SCOPE etiketa esleitu diogu eta irismenaren barruko hitz guztiak hartuko ditu.

#### 3.1.3 Ezeztapenaren fokoa

Ezeztapenaren fokoa ezeztapenaren irismenean argien edo indar handienez ezeztatutako elementua da (Huddleston eta Pullum, 2002). Hurrenkera edo intonazio bidez antzeman daiteke (Altuna *et al.*, 1985), baina bi faktore horiek ez dira beti nahikoak fokoaren identifikaziorako (Ortíz de Urbina, 2003). Aurkari-tzan agertzen den elementu batek, ordea, fokoa identifikatzeko informazio zehatza eman diezaguke; hala, (8) adibidean fokoa identifikatzeko “baizik eta” aurkaritzazko lokailua baliatu dugu. Kasu horretan, “kooperatiben eredua” eta “gobernatzeko era” kontrajartzen dira eta horregatik erabaki dugu “kooperatiben eredua” fokotzat hartzea.

(8) Nabarmendu du **ez** duela [kooperatiben eredua] zalantzan jartzen, baizik eta “gobernatzeko era” bakarrik.

<sup>6</sup>Hutsunetik hutsunerako hizki segida.

Fokoari FOCUS etiketa esleitu diogu eta beti egongo da ezeztapenaren irismenaren barruan.

### 3.2 Ezeztapena denbora-informazioaren prozesamenduan

Ezeztapena aztertzearen helburu nagusia horrek gertaeretan duen eragina aztertzea da. Gure hipotesiaren arabera, gertaera bat ezeztapenaren irismenaren barruan badago, bere polaritatea alda daiteke. (9) adibidean baiezka ematen den (polaritate positiboa duen) eta faktualitate positiboa (gertatu dela edo gertatuko delako balioa) duen “erosteko” gertaera (10) adibidean ezeztapenaren irismenaren barruan agertzen da eta ondorioz, bere polaritatea eta faktualitatea balioa aldatu dira.

(9) Garraio-ministroak hegazkin europarrak [erosteko] eskatu die aire-lineei.

(10) Garraio-ministroak hegazkin europarrak **ez** [erosteko] eskatu die aire-lineei.

Ezeztapenaren irismenaren baitan dauden gertaera guztien faktualitate-balioa ez da aldatzen, ordea. (11) adibideko “bilera horretatik” gertaerak bere faktualitate-balio positiboa mantentzen du, hau da, gertatu da. Ondorioz, ezeztapenaren irismenaz gain, faktualitatea ebazteko zein beste ezaugarri aztertu ere erabaki beharko dugu etorkizunean.

(11) Sindikatuak **ez** zuten [ezer onik] espero atzoko *bilera horretatik*.

Bestalde, denbora-informazioaren prozesamenduan informazio hori erabiliko duten tresnak sortzea dugu helburu. Automatikoki sortu nahi dugun gauzetako bat denbora-lerroak dira. Denbora-lerroetan gertaerak kronologikoki antolatzen dira gertaeren eta gertatu diren denbora-uneen arteko erlazioak baliatuz. Entitate baten (pertsona, erakundea, etab.) gertaerak denbora-lerroan kokatu nahi izanez gero, entitateak zein gertaeratan parte hartu duen jakin behar dugu. Hala, entitate horiek espreski ezeztatuta agertzen diren, bestela esan, ezeztapenaren fokoa diren jakitea oso interesgarria da, horiek ez baitute gertaeran parte hartuko.

### 3.3 Esperimentazioa

Ezeztapenaren inguruan hartutako erabakiak ebaluatzeko, corpus bat eskuz etiketatu dute bi etiketatzailek ezeztapena etiketatzeko gidalerroei jarraituz. Zehazki, 20 dokumentuko albiste-corpus bat etiketatu dute Content Annotation Tool (CAT) (Bartalesi Lenzi *et al.*, 2012) tresna erabilita.

#### 3.3.1 Corpora

Etiketaturako 20 dokumentu horiek EusMEANTIME corpuseko (MEANTIME corpusaren (Minard *et al.*, 2016) euskarazko bertsioa) lagin bat da. Lagina albiste ekonomiko osatuta dago, 5.247 hitz ditu eta 33 ezeztapen etiketatu dituzte. Corpus hori EusTimeML bidez ere etiketatuta dago eta espero dugu etorkizunean bi etiketatzek uztartu ahal izatea.

1 Taula: Esperimentuan erabilitako corpusaren datuak

|                         |      |
|-------------------------|------|
| Dokumentu kopurua       | 20   |
| Token kopurua           | 5247 |
| Ezeztapen kopurua (ggb) | 33   |

#### 3.3.2 Emaitzak eta etiketatzearen analisia

Etiketatzailen arteko adostasuna Diceren koefizientea (Dice, 1945) erabiliz neurtu da. 2. taulan agertzen den bezala, etiketa bakoitzeko lau balio lortu ditugu adostasunaren ebaluaziorako: etiketen hedapenaren bat etortzea batez beste dokumentu bakoitzeko, micro-average (markable); tokenen gainjarketa batez beste dokumentu bakoitzeko, micro-average (token); etiketen hedapenaren bat etortzea batez beste corpusean, macro-average (markable); eta tokenen gainjarketa batez beste corpusean, macro-average (token).

Emaizetatik ondoriozta daiteke ezeztapena adierazten duen elementua identifikatzea ez dela lan zaila izan, ia bat etortze osoa (0.969 eta 0.982ko balioak) lortu baitugu. Desberdintasun bakarrak “proiektua onartu **ez** zelako” eta “jaso **gabe**” etiketatzaila bakarrak etiketatu dituela dira eta, ezeztapen argiak izaki uste dugu etiketatzaila batek ahaztu dituela.

2 Taula: Ezeztapen-marken etiketatzean etiketatzailen arteko adostasuna

| Etiketa | Micro-average<br>(Markable) | Micro-average<br>(Token) | Macro-average<br>(Markable) | Macro-average<br>(Token) |
|---------|-----------------------------|--------------------------|-----------------------------|--------------------------|
| NEG.CUE | 0.969                       | 0.969                    | 0.982                       | 0.982                    |
| SCOPE   | 0.562                       | 0.806                    | 0.55                        | 0.858                    |
| FOCUS   | 0.381                       | 0.55                     | 0.373                       | 0.549                    |

Irismenerari eta fokoari dagokionez, ordea, emaitzak ez dira hain onak izan, nabarmentzekoa bada ere irismenen gainjartzea 0.8tik gorakoa izan dela. Emaizta horietatik ondoriozta dezakegu irismena identifikatzea ere ez dela ataza zaila, desberdintasunak aurki badaitezke ere. Fokoaren kasuan, ordea, emaitzak baxuak izan dira. Fokoaren identifikazioan ez da nahikoa sintaxia, subjektibotasunak intonazioak rol garrantzitsua jokatzen baitute. Etiketatzaile bakoitzak hautatutako fokoa zein den agertzen da (12) eta (13) adibideetan. (12) adibidean, informaziorik zehatzena hartu da foko gisa eta (13) adibidean normalean perpausaren fokoak hartzen duen tokia erabili da foko aukeratzeko.

(12) Salmenten beherakada ez da Estatu Batuetako autogileetara [soilik] mugatu.

(13) [Salmenten beherakada] ez da Estatu Batuetako autogileetara soilik mugatu.

## 4 Ondorioak

Lan honetan euskarazko ezeztapena analizatu dugu eta ezeztapenaren eskuzko etiketatzea egin dugu HPan ezeztapenaren inguruan egindako lanetan oinarrituta. Euskarazko HPan ezeztapena lantzean lehen aldiz, ezeztapen-markez gain, ezeztapenaren irismena eta fokoak ere definitu ditugu.

Ezeztapen-informazioa etiketatzeko gidalerroak osatu ditugu eta eskuzko etiketatzea egin dugu horiei jarraituz. Ezeztapen-marken identifikazioan ia erabateko adostasuna lortu dugu, ataza erraza delako seinale. Irismenaren eta fokoaren emaitzak, ordea, ez dira hain onak izan. Irismenean tokenen gainjartzea nahiko altua izan da (0,8tik gorakoa) eta gidalerroak berraztertuta, adostasuna hobetzea espero dugu. Fokoaren etiketatzea askoz ere ataza zailagoa da, etiketatzailearen subjektibotasuna eta intonazioaren faltagatik; ondorioz, nola hobetu aztertu beharko dugu.

Etiketatzeko gainbegiratu ostean, ezeztapena etiketatuta duen corpusa izango dugu. Tamainaz txikia da eta 33 ezeztapen baino ez dira aurkitu. Etorkizunean corpusa zabaldu ostean, ikasketa automatikorako erabili ahal izango da, ikasketa automatikoko tresnek corpus etiketatuetatik ikasten baitute.

Corpus etiketatua izateak gertaeren faktualitatearen analisia hobetzen ere lagunduko digu. Ezeztapenaren irismenaren barruan dauden gertaeren faktualitatea aztertuko dugu ezeztapenaren eragina neurtzeko. Hala, denbora-informazioaren azterketa eta prozesamendua hobetu ahal izango dugu. Etorkizunean, entitateetan oinarritutako denbora-lerroak sortzean, bai faktualitatearen bai ezeztapenaren fokoaren informazioa baliagarria izango zaigu denbora-lerroan agertuko diren gertaerak aukeratzeko.

## 5 Etorkizunerako planteatzen den norabidea

Helburuen atalean esan bezala, ezeztapena gertaeren prozesamenduan osagarri legez erabiltzeko aztertu dugu. Etorkizunean hiru erabilera eman nahi dizkiogu ezeztapenaren azterketa eta etiketatze honi: i) gertaeren faktualitatearen eskuzko etiketatzean ezeztapena eta bere irismena kontuan hartuko ditugu; ii) ezeztapenaren etiketatzea, ikasketa automatikoan baliatuko dugu. Horretarako, ezeztapena etiketatua duen corpusa zabaltzen jarraituko dugu eta etiketatutako informazioa ezeztapena prozesatzeko ikasketa automatikoko tresnetan baliatuko dugu. Azkenik, iii) faktualitatea automatikoki ebazteko erabiliko dugu ezeztapenaren inguruko informazioa.

Gainera, lan honetan ez ditugu ezeztapena adierazten duten elementu guztiak aztertu eta beste elementuen (izenak eta aditzak nagusiki) eragina ere aztertu nahi dugu, betiere denbora-informazioaren prozesamenduari begira.

## Erreferentziak

- ALTUNA, BEGOÑA, MARÍA JESÚS ARANZABE, eta ARANTZA DÍAZ DE ILARRAZA. 2016. Euskarazko denbora-egiturak etiketatzeko gidalerroak v2.0. Technical report, Lengoaia eta Sistema Informatikoak Saila, UPV/EHU. UPV/EHU/LSI/TR;01-2016. <https://addi.ehu.es/handle/10810/17305>.
- ALTUNA, BEGOÑA, ANNE-LYSE MINARD, eta MANUELA SPERANZA. 2017. *Proceedings of the Workshop Computational Semantics Beyond Events and Roles*, chapter The Scope and Focus of Negation: A Complete Annotation Framework for Italian, 34–42. Association for Computational Linguistics.
- ALTUNA, PATXI, PELLO SALABURU, PATXI GOENAGA, MARÍA PILAR LASARTE, LINO AKESOLO, MIREN AZKARATE, PIARRES CHARRITON, ANDOLIN EGUSKITZA, JEAN HARITSCHELHAR, ALAN KING, JOSE MARI LARRARTE, JOSE ANTONIO MUJICA, BEÑAT OYHARÇABAL, eta KARMELE ROTAETXE. 1985. *Euskal Gramatika Lehen urratsak (EGLU) I*. Euskaltzaindia, Bilbo: Euskaltzaindiko Gramatika batzordea.
- ARANZABE, MARÍA JESÚS, 2008. *Dependentzia-ereduan oinarritutako baliabide sintaktikoak: zuhaitz-bankua eta gramatika konputazionala*. Donostia: Universidad del País Vasco/Euskal Herriko Unibertsitatea tesia. [http://ixa.si.ehu.eus/sites/default/files/dokumentuak/4118/tesi-txostena\\_eranskinak.pdf](http://ixa.si.ehu.eus/sites/default/files/dokumentuak/4118/tesi-txostena_eranskinak.pdf).
- ATTARDI, GIUSEPPE, VITTORIA COZZA, eta DANIELE SARTIANO. 2015. Annotation and extraction of relations from italian medical records. In *Proceedings of the 6th Italian Information Retrieval Workshop (IIR 2015)*.
- BARTELES LENZI, VALENTINA, GIOVANNI MORETTI, eta RACHELE SPRUGNOLI. 2012. CAT: the CELCT Annotation Tool. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, ed. by Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Jan Odijk, eta Stelios Piperidis, 333–338, Istanbul, Turkey. European Language Resources Association (ELRA).
- BLANCO, EDUARDO, eta DAN MOLDOVAN. 2011. Semantic Representation of Negation Using Focus Detection. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, 581–589. Association for Computational Linguistics.
- BONIAL, CLAIRE, OLGA BABKO-MALAYA, JINHO D. CHOI, JENA HWANG, eta MARTHA PALMER. 2010. Propbank Annotation Guidelines (Version 3.0). Technical report, Center for Computational Language and Education Research, Institute of Cognitive Science, University of Colorado at Boulder.
- COUNCILL, ISAAC G., RYAN McDONALD, eta LEONID VELIKOVICH. 2010. What's great and what's not: learning to classify the scope of negation for improved sentiment analysis. In *Proceedings of the workshop on negation and speculation in natural language processing*, 51–59. Association for Computational Linguistics.
- DICE, LEE RAYMOND. 1945. Measures of the Amount of Ecologic Association between Species. *Ecology* 26.297–302.
- ESTARRONA, AINARA, 2014. *EPEC corpora predikatu-mailan etiketatzeko oinarriak: EPEC-RolSem, BVI eta e-ROLda*. Donostia: Universidad del País Vasco/Euskal Herriko Unibertsitatea tesia. [http://ixa.si.ehu.eus/sites/default/files/dokumentuak/4129/Ainara\\_Estarrona\\_Tesia.pdf](http://ixa.si.ehu.eus/sites/default/files/dokumentuak/4129/Ainara_Estarrona_Tesia.pdf).
- HUDDLESTON, RODNEY D., eta GEOFFREY K. PULLUM. 2002. *The Cambridge Grammar of the English Language*. Cambridge University Press.
- JIMÉNEZ-ZAFRA, SALUD M., MARIONA TAULÉ, M. TERESA MARTÍN-VALDIVIA, L. ALFONSO UREÑA-LÓPEZ, eta M. ANTÒNIA MARTÍ. to appear. SFU ReviewSP-NEG: a Spanish corpus annotated with negation for Sentiment Analysis A Typology of negation patterns. *Language Resources and Evaluation*.
- KIM, JIN-DONG, TOMOKO OHTA, eta JUN'ICHI TSUJII. 2008. Corpus annotation for mining biomedical events from literature. *BMC bioinformatics* 9.1.
- KONSTANTINOVA, NATALIA, SHEILA C.M. DE SOUSA, NOA P. CRUZ, MANUEL J. MAÑA, MAITE TABOADA, eta RUSLAN MITKOV. 2012. A review corpus annotated for negation, speculation and their scope. In *Proceedings of the Eight International Conference on Language Resources and Evaluation*.

- (*LREC'12*), ed. by Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Jan Odijk, eta Stelios Piperidis, 3190–3195, Istanbul, Turkey. European Language Resources Association (ELRA).
- LAPARRA, EGOITZ, ITZIAR ALDABE, eta GERMAN RIGAU. 2015. From TimeLines to StoryLines: A preliminary proposal for evaluating narratives. In *Proceedings of the 1st Workshop on Computing News StoryLines (CNews 2015) at the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2015)*, Beijing, China.
- MARTÍ, M. ANTÓNIA, M. TERESA MARTÍN-VALDIVIA, MARIONA TAULÉ, SALUD MARÍA JIMÉNEZ-ZAFRA, MONTSERRAT NOFRE, eta LAIA MARSÓ. 2016. La negación en español: análisis y tipología de patrones de negación. *Procesamiento del Lenguaje Natural* 57.41–48.
- MINARD, ANNE-LYSE, MANUELA SPERANZA, RUBEN URIZAR, BEGOÑA ALTUNA, MARIEKE VAN ERP, ANNELEEN SCHOEN, eta CHANTAL VAN SON. 2016. MEANTIME, the NewsReader Multilingual Event and Time Corpus. In *Proceedings of LREC 2016*.
- MORANTE, ROSER, eta WALTER DAELEMANS. 2012. ConanDoyle-neg: Annotation of negation in Conan Doyle stories. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, ed. by Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Jan Odijk, eta Stelios Piperidis, Istanbul, Turkey. European Language Resources Association (ELRA).
- , SARAH SCHRAUWEN, eta WALTER DAELEMANS. 2011. Annotation of negation cues and their scope: Guidelines v1. *Computational linguistics and psycholinguistics technical report series, CTRS-003*.
- ORTÍZ DE URBINA, JON. 2003. Ezezko perpausetako fokoak: (galdegaiak eta ukagaiak. In *P. Laffitteren sortzearen mendemugako biltzarra*, ed. by B. Oihartzabal eta J. M. Makazaga (Eds.). Bilbao: Euskaltzaindia.
- PALMER, MARTHA, DANIEL GILDEA, eta PAUL KINGSBURY. 2005. The proposition bank: An annotated corpus of semantic roles. *Computational linguistics* 31.71–106.
- PUSTEJOVSKY, JAMES, JOSÉ M. CASTAÑO, ROBERT INGRIDIA, ROSER SAURÍ, ROBERT J. GAIZAUSKAS, ANDREA SETZER, GRAHAM KATZ, eta DRAGOMIR R. RADEV. 2003. TimeML: Robust Specification of Event and Temporal Expressions in Text. *New directions in question answering* 3.28–34.
- PYYSALO, SAMPO, FILIP GINTER, JUHO HEIMONEN, JARI BJÖRNE, JORMA BOBERG, JOUNI JÄRVINEN, eta TAPIO SALAKOSKI. 2007. BioInfer: a corpus for information extraction in the biomedical domain. *BMC Bioinformatics* 8.50.
- STRICKER, VANESA, IGNACIO IACOBACCI, eta VIVIANA COTIK. 2015. Negated findings detection in radiology reports in spanish: an adaptation of negex to spanish. In *Workshop on Replicability and Reproducibility in Natural Language Processing: adaptive methods, resources and software at IJCAI 2015*.
- VINCZE, VERONIKA, GYÖRGY SZARVAS, RICHÁRD FARKAS, GYÖRGY MÓRA, eta JÁNOS CSIRIK. 2008. The bioScope corpus: biomedical texts annotated for uncertainty, negation and their scopes. *BMC bioinformatics* 9.1.

## Eskerrak eta oharrak

Lan hau eusko Jaurlaritzako PRE\_2016\_2\_294 bekari esker egin da.