

Neurona-sareetan oinarritutako euskararako korreferentzia-ebazpena

Urbizu, Gorka; Soraluze, Ander eta Arregi, Olatz

IXA taldea. Euskal Herriko Unibertsitatea.

gurbizu002@ikasle.ehu.eus

Laburpena

Lan honek euskararako korreferentzia-ebazpenean egindako lanari jarraipena ematea du helburu, korreferentzia-ebazpenerako neurona-sareetan oinarritutako sistema bat eraikiz. Horretarako polonierarako eraikitako sistema bat hartu da abiapuntutzat, eta euskarara egokitu. EPEC-KORREF corpusetik abiatuta, aipamen-bikoteak eta hauen ezaugarriak erauzi dira eta neurona-sarea entrenatu da aipamen-bikoteak korreferenteak ote diren erabakitzeko. Jarraian, neurona-sarearen iragarpenetatik korreferentzia-klusterrak sortu eta ebaluatu egin dira. CoNLL metrikaren % 41,20 puntuko F1 balioa lortu da eta lortutako emaitza baxuak corpusaren tamaina txikiagatik direla ondorioztatu da.

Hitz gakoak: korreferentzia-ebazpena, ikasketa-sakona, neurona-sareak, euskara

Abstract

This work aims to continue with the previous work done in coreference resolution for Basque language, building a system for coreference-resolution based on neural networks. For this purpose, a system for Polish has been used and adjusted for Basque. Mention-pairs and their features were extracted from the EPEC-KORREF corpus, to train the neural network model at the task of deciding if the pair is coreferent or not. After that, the coreference clusters were built from the output of the neural network, and these were evaluated. 41.20% F1 points of the CoNLL metric were obtained and it was concluded that the low results obtained were due to the small size of the corpus.

Keywords: coreference-resolution, deep learning, neural networks, Basque language

1. Sarrera eta motibazioa

Hizkuntzaren prozesamendua (NLP - *Natural Language Processing*) informatika, adimen artifiziala eta hizkuntzalaritza diziplinen arteko arloa da. Hizkuntzaren ulermena behar duten hizkuntzaren prozesamenduko atazetan korreferentzia-ebazpena oinarritzko urratsa da.

Testu bateko bi espresio testualek objektu berbera adierazi edo erreferentziatzen dutenean, bi espresio horien artean korreferentzia erlazio bat dagoela esan ohi da. Testu batean ager daitezkeen espresio testual horien arteko korreferentzia-erlazioak ebaztea helburu duen atazari korreferentzia-ebazpena deritzo (Grishman eta Sundheim, 1995).

Argiago ikusteko azter dezagun ondorengo adibidea:

(1) **Nazio Batuen Erakundea** izan zen **bitartekari** eta **hark** hartu zuen prozesuaren ardura.

Nazio Batuen Erakundea, *bitartekari* eta *hark* espresio testualek NBEari erreferentzia egiten diotenez, korreferenteak direla edo korreferentzia-erlazioa dutela esan dezakegu.

Ataza honetan sarri erabiltzen diren bi termino *entitatea* eta *aipamena* dira. Entitatea mundu errealeko pertsona, objektu edo erakunde bat litzateke; aipamena berriz, entitate bati erreferentzia egiten dion hitz multzo edo espresio testuala da.

Termino horiek hobeto ulertzeko, lehenengo adibidera itzul gaitzen:

(2) [**Nazio Batuen Erakundea**] izan zen [**bitartekari**] eta [**hark**] hartu zuen [**prozesuaren ardura**].

Adibide honetan aipamenak kortxete artean daudenak dira, lehenengo hirurak, [Nazio Batuen Erakundea], [bitartekari] eta [hark] entitate berdinari egiten diote erreferentzia, beraz, korreferenteak dira eta korreferentzia multzoa edo klusterra osatzen dute. Bestalde, [prozesuaren ardura] aipamenak entitate ezberdin bati egiten dio erreferentzia, beraz, ez da gainerakoekin korreferentea.

Korreferentzia-ebazpena bi azpi-atazatan banatu ohi da, aipamen-detekzioa eta erreferentzien ebazpena (Pradhan *et al.*, 2011). Lehenik aipamen-detekzioa egiten da, testuan zehar entitateren bati erreferentzia egiten dioten espresio testualak identifikatuz. Ondoren, aipamenak multzokatu behar dira, entitate berdinari erreferentzia egiten diotenak klusterretan elkartuz.

Korreferentzia-ebazpen automatikoa garrantzitsutzat jotzen da, oro har, testu ulermen sakona dakarren hizkuntzaren prozesamenduko ataza oro burutzeko (Clark, 2015). Bestek beste, informazio erauzketan, testuen laburpenean, galdera-erantzun sistemetan, sentimenduen analisisian eta itzulpen automatikoan aplikatzen da.

2. Arloko egoera eta ikerketaren helburuak

2.1. Korreferentzia-ebazpena

Orain dela hamarkada batzuk korreferentzia-ebazpena lantzen hasi zenetik, hainbat teknika aplikatu dira ataza ebazteko, erregela bidezko sistemak lehenik, gerora ikasketa automatikokoak eta berriki ikasketa sakonekoak. Erregelatan oinarritutako sistemen artean, esanguratsuen Stanfordeko korreferentzia-ebazpenerako sistema determinista da (Lee *et al.*, 2013). Ordurako, ikasketa automatikoan oinarritutako sistemak hasiak ziren emaitza onak ematen, horien artean BART (*Beautiful Anaphora Resolution Toolkit*) (Versley *et al.*, 2008) sistema aipa dezakegu.

Azken urteetan hizkuntzaren prozesamenduan neurona-sareek arrakasta handia izan dute, eta korreferentzia-ebazpenean ere, artearen egoeran emaitzarik onenak lortzen dituzten sistemek, teknologia hori darabilte. Neurona-sareetan oinarritutako sistema gehienek ikasketa automatikoko sistemen antzera funtzionatzen dute, ikasteko eta sailkatzeko atalak neurona-sareekin ordezkaturik. Sistemek aipamen-bikoteak korreferente edo ez-korreferente gisa edo aipamenak entitateka klusterretan sailkatu ohi dituzte. Aipamenen detekzioa eta ikasketarako ezaugarriak garapen neketsuko erregela bidezko tresnekin lortzen dira. Eredu hau darabilten lan esanguratsuenak Clark eta Manning (2016) eta Wiseman *et al.* (2016) dira.

Lee *et al.* (2017) lanean aldiz, korreferentzia-ebazpenerako muturretik muturrerako neurona-sarea aurkezten dute, aipamen-detekzio automatikorik eta ezaugarriak erauzteko tresnen beharrik gabe. Neurona-sarea gai da ataza burutzeko behar duen informazioa testutik ikasteko, eta hori eginik aipamen-detekzioan eta ezaugarriak erauztean dauden errorearen propagazioa saihesten da. Sistema horren neurona-sareetan oinarritutako arkitektura konplexua aurretik sortutako sistema guztiei gailendu zaie. Lee *et al.* (2018) artikuluan lanari jarraipena ematen diote, aipamenak entitateka multzokatzeko neurona-sareak erabiliz eta neurona-sareen inguruko ikerketetan lortu diren azken berrikuntzak aplikatuz.

Neurona-sareen aurreko korreferentzia-sistemetan ezaugarri semantikoak erabiliz ezagutza semantikoa integrazteko saiakerak egin dira arrakasta handirik lortu gabe. Neurona-sareen bitartez ikasitako hitzen errepresentazio distribuzionalekin berriz, ezagutza semantikoa arrakastaz integratzea lortu da (Clark eta Manning, 2016).

Ingelesa ez den beste hizkuntzetarako ere neurona-sareetan oinarritutako korreferentzia-ebazpenerako sistemak hasi dira garatzen. Clark eta Manning (2016) artikuluan txinerarako sistema aurkezten da, Park *et al.* (2016) lanean korearrerako eta Niton *et al.* (2018) lanean polonierarako.

2.2. Euskararako korreferentzia-ebazpena

Euskararen ezaugarriak erreparatuta, esan daiteke eranskaria (sintagma batzuk postposizioen bitartez osatzen dira), buru-azkena (sintagmen burua amaieran ageri da), ordena librekoa (esaldiaren egitura malgua da) eta pro-drop (ize-nordainak isilik ageri daitezke) hizkuntza dela. Ezaugarri horiek tarteko, korreferentzia-ebazpenerako euskararako tresna propioak garatzeko beharra ikusi da.

Lehenik erregelatan oinarritutako aipamen-detektatzaile bat garatu zen, gerora eraikitako sistemek darabiltena (Soraluze *et al.*, 2017b). Ondoren, Soraluze *et al.* (2015) lanean, ingeleserako diseinatutako erregelatan oinarritutako Stanfordeko sistema (Lee *et al.*, 2013) egokitu zem korreferentzia-ebazpena euskarazko testuetan egiteko. Jarraian, hobekuntzak proposatu zituzten, Wikipedia eta Wordnet baliabide semantikoak gehituz (Soraluze *et al.*, 2017a).

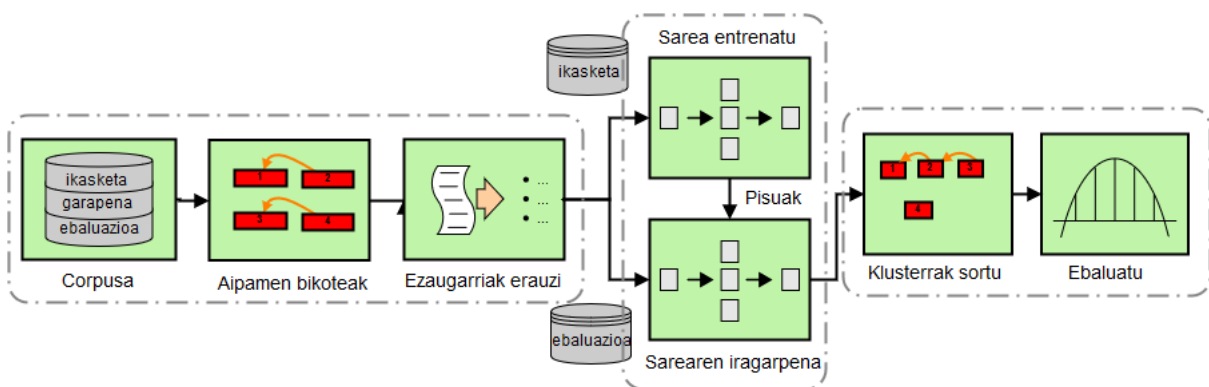
Soraluze *et al.* (2016) lanean, ikasketa automatikoan oinarritutako sistema bat plazaratzen dute, ingeleserako diseinatutako BART korreferentzia-ebazpenerako sistema (Versley *et al.*, 2008) euskararen ezaugarrietara egokituz. Azkenik, Wikipediatik erauzitako ezagutza entziklopedikoa gehitu zaio (Soraluze *et al.*, 2017b).

Lan honek Soraluze (2017) tesi lanari jarraipena ematea du helburu, euskararako korreferentzia-ebazpenerako neurona-sareetan oinarritutako sistema bat eraikiz. Horretarako polonierarako eraikitako sistema bat (Niton *et al.*, 2018) hartu da abiapuntutzat, neurona-sareekin erregelatan oinarritutako sistemaren antzeko emaitzak lortu dituztelako, baliabide aldetik polonierak ingelesak baino corpus txikiagoa duelako eta bi hizkuntzak eranskariak eta ordena librekoak direlako.

3. Ikerketaren muina

Neurona-sareetan oinarrituta euskararako eraiki den korreferentzia-ebazpenerako sistema azalduko da jarraian. Sistema hiru atal nagusitan banatzen da, aurreprozesaketa, neurona-sarea eta ebaluazioa. Ikus 1 irudia.

1. irudia. Korreferentzia-ebazpenerako sistemaren eskema



3.1. Aurreprozesaketa

Aurreprozesaketaren abiapuntua, EPEC-KORREF (Ceberio *et al.*, 2018) corpusa da, 45.000 hitz eta 12.000 aipamen dituena. Corpus honek aipamenak eta korreferentzia erlazioak eskuz anotatuta ditu, korreferentzia-ebazpenerako baliagarriak izan daitezkeen beste ezaugarri askorekin batera. EPEC (Aduriz *et al.*, 2006) corpusaren azpi-corpusa da eta euskarazko *Egunkaria* egunkariko albisteez osatuta dago. Corpusa ikasketa, garapena eta ebaluazioa ataletan banatua dator.

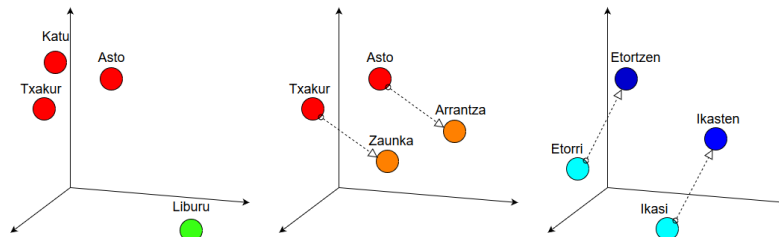
Aurreprozesaketarekin jarraitzeko, corpuseko aipamenak hartu eta bikoteak sortzen dira. Ikasketa atalean, eskuz anotatutako aipamenak erabili dira, eta bikoteak sortzeko, aipamen bakoitza bere aurrekariarekin (bikote positiboak) eta aipamenaren eta aurrekariaren artean daudenean (bikote negatiboak) parekatu da. Ebaluazioa atalean, aldiz, aipamen-detektatzaile automatikoa (Soraluze *et al.*, 2017b) baliatu da testuan zehar entitate bati erreferentzia egiten dieten aipamenak identifikatzeko. Ondoren, aipamen bakoitza aurreko zortziarekin parekatu da, aurrekaria hauen artean topatzen baita kasuen % 97an (Goenaga *et al.*, 2012).

Aipamen-bikoteak sortuta ditugunean, aipamen horien ezaugarriak erauzi behar dira, neurona-sarea entrentzeko sarrera osatuko duten zenbakizko balioak sortzeko.

Ezaugarri horien artean, aipamen bakoitzeko hitzen errepresentazio distribuzionalak dira nagusi. Hitzen errepresentazio distribuzionalak, hitzen esanahia gordetzen duten zenbakiz osatutako bektoreak dira, eta generoa, numeroa, eta beste hainbat ezaugarri biltzen dituzte. 2. irudian ikus dezakegunez, antzeko hitzak espazio berean multzokatzen dira, eta animalia eta hauen hotsen arteko erlazioak (txakur-zaunka, asto-arrantza), eta aditz forma ezberdinenak (etorri-etortzen, ikasi-ikasten) biltzeko gai dira. Horiek, 160 miloi hitzeko corpuseko (Wikipedia + Elhuyar Web Corpus) hitzen testuinguruan agertzen diren hitzetatik ikasi dira, eta 50 dimentsiotako zenbakizko bektoreetan kodetu, word2vec¹ tresnarekin. Honetaz gain, aipamen bakoitzaren numeroa eta pertsona, aipamenen arteko distantzia eta horiek ea hitz/lema/buru berdinak dituzten bezalako ezaugarriak ere erabili dira.

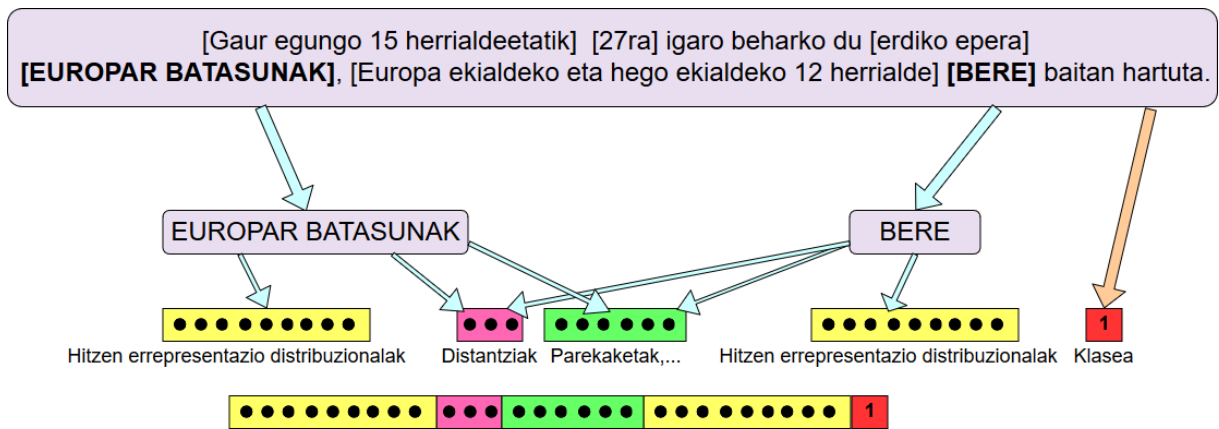
¹<https://github.com/danielfrg/word2vec>

2. irudia. Hitzen errepresentazio distribuzionalak



Guztira, 1.147 dimentsioko bektorea sortu da aipamen-bikoteko (3. irudia), eta honekin batera aipamen-bikotea korreferentea den edo ez (1 edo 0) adierazten duen klasea pasatzen zaio sarreran neurona-sareari ikasteko.

3. irudia. Aipamen-bikotearen bektorea



3.2. Neurona-sarea

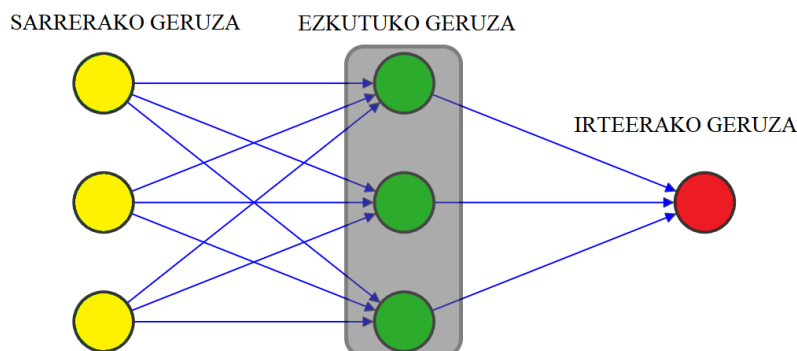
Garuneko neurona-sare biologikoen portaera imitatzen duten neurona-sareak (*neural networks*), ataza ezberdinak burutzeko gai dira adibideetatik ikasiz. Neurona-sareak, sarreran jasotzen duen seinalea prozesatzeko eta bidaltzeko gai diren neuronez osatuta daude. Neuronen arteko konexioek pisu bat dute, eta hauen arabera prozesatuko dira seinaleak. Sarea, dagokion ataza ikasteko, pisuak doitzen joango da, unean ikasteko dituen adibideetara egokituz.

Normalean, neuronak geruzetan antolatzen dira eta ezkutuko geruzak eta irteera geruza bereizten dira. Ezkutuko geruzak sarrera eta irteeraren artean daudenak dira eta kutxa beltz moduan funtzionatzen dute, barruan gertatzen dena erabiltzailearentzat opakua izanik. Irteerako geruza neurona batek edo gutxi batzuk osatzen dute, eta hauek emaitza itzultzen dute. Neurona-sare simple baten arkitekturaren eskema 4. irudian ikus dezakegu.

Lan honetan erabili den neurona-sarea, berriz, ezkutuko 3 geruzadun sare trinko bat da, polonierarako sistemak bezala, 500, 300 eta 100 neurona ditu geruza bakoitzean. Irteerako geruzak, aldiz, neurona bakarra du, 0 eta 1 arteko zenbaki bat itzultzen duena (korreferenteak izateko probabilitate gisa har genezake). Geruza eta neurona kopuruekin proba ezberdinak egin dira, baina ez da hobekuntzarik hauteman. Epoch (iterazio kopurua) eta mini-batch (ikasketan sareari aldiko pasatako instantzia kopurua) hiperparametroak ere doitu dira, 2 eta 64 balioetan finkatuz. Pisuak hasieratzea berriz, ausaz egin da, polonierarako entrenatutako neurona sarearen pixuak erabiliz ez baita hobekuntzarik egon.

Neurona-sarea, ikasketa ataleko aurreprozesatutako instantziekin entrenatuko da. Ikasketaren emaitza neuronon arteko pisuak dira, neurona-sareak aipamen-bikote berriak korreferenteak diren erabakitzeke erabiliko dituenak.

4. irudia. Neurona-sarearen arkitekturaren eskema



3.3. Ebaluazioa eta emaitzak

Korreferentzia-ebazpenerako atazan, neurona-sarearen zehaztasuna ebaluatzeko, ikasteko erabili ez diren testuetan aplikatzen da sistema.

Neurona-sareak, corpuseko ebaluazioa ataleko instantzientzat iragarpen bat bueltatzen du, 0 eta 1 arteko balio bat, aipamen-bikoteak korreferente diren edo ez erabakitzeko. Jarraian, sistemak korreferenteak izateko probabilitate altua emandako aipamenak kluster berean sartuko dira. Azkenik, lortutako emaitzak ebaluatu dira korreferentzia-ebazpenerako ohikoa den ebaluatzailearekin (Pradhan *et al.*, 2014).

Korreferentzia-ebazpenerako sistemen kalitatea neurtzeko, eta sistema ezberdinen arteko konparaketa egin ahal izateko ebaluazio metrika ezberdinak erabiltzen dira. MUC loturetan oinarritutako metrika da, hau da, urre-patroiko eta sistemak lortutako korreferentzia-loturak konparatzen ditu (Vilain *et al.*, 1995). B^3 metrikak aipamenak konparatzen ditu loturen orde, doitasun eta estaldura globalak, aipamen bakoitzaren balioetatik lortzen dira eta ondorioz aipamenetan oinarritutako metrika dela esaten da (Bagga eta Baldwin, 1998). $CEAF_e$ metrikak ϕ antzekotasun balioa erabiltzen du bi entitateen arteko antzekotasuna neurtzeko, beraz entitateetan oinarritutakoa dela esan daiteke (Luo, 2005). CoNLL metrika, aurreko hiruen batezbesteko aritmetikoa da, eta metrika nagusizat hartzen da. 1. taulan laburbiltzen dira proiektu honetan korreferentzia-ebazpena atazan neurona-sareetan oinarritutako sistemak lortutako emaitzak. MUC metrikari lortutako balio baxua azpimarratzekoa da, sistemak korreferentzia-erlazio gutxi sortzen dituela adierazten duena. Neurri honek eragin handia du CoNLL metrikari, eta ondorioz lortu den F1 balioa % 41,20 puntukoa da.

1. taula. Lortutako emaitzak

	MUC	B^3	$CEAF_e$	CoNLL
Neurona-sareak	9,32	58,40	55,87	41,20

Neurona-sareetan oinarrituta eraikitako sistemarekin lortutako emaitzak, aurretik sortutako erregelatan (Soraluze *et al.*, 2015) eta ikasketa automatikoan (Soraluze *et al.*, 2016) oinarritutako sistemen emaitzak baino baxuagoak dira euskararako. Aipatutako korreferentzia-ebazpenerako sistemekin egindako konparaketa 2. taulan dago ikusgai.

2. taula. Euskararako korreferentzia-ebazpenerako sistemen konparaketa

Sistema	MUC	B^3	$CEAF_e$	CoNLL
Erregelatan oinarritutakoa	42,95	62,97	62,04	55,98
Ikasketa automatikoa	40,98	61,66	60,00	54,21
Neurona-sareak	9,32	58,40	55,87	41,20

4. Ondorioak

Neurona-sareetan oinarritutako euskararako lehen korreferentzia-ebazpenerako sistema eraiki da. Sistema honekin CoNLLn % 41,20 puntuko F1 balioa lortu da, aurretik sortutako erregelatan eta ikasketa automatikoan oinarritutako sistemekin lortutako emaitzak baino baxuagoa.

Korreferentzia-ebazpenerako erregelatan oinarritutako sistemek corpus txikia erabili arren, emaitza onak eman ditzakete, sistema adimendunak darabilen ezagutza eskuz sortutako erregelekin osatzen baita. Ikasketa automatikoko sistemek, erregelatan oinarritutakoek baino corpus handiagoa eskatzen dute, sistemak datuetatik ikasten duelako, baina corpus ez oso handiekin ere emaitza onak eman ditzakete. Ikasketa sakonean ordea, neurona-sareak entrenatzeko datu gehiago behar da, lan honetan erabili den EPEC-KORREF euskarazko corpora txikia da, eta lortu diren emaitza kaxkarren erantzule nagusia dela ondorioztatu da. Kokatzearen, polonierazko corpora EPEC-KORREF corpora baino 10 aldiz handiagoa da eta neurona-sareekin erregelatan oinarritutako sistemaren antzeko emaitzak lortu dituzte; ingeleserako erabili ohi den corpora euskarazkoa baino 20 bat aldiz handiagoa da eta artearen egoeran dauden emaitzak lortu dira. Corpora handituta, CoNLL metrikan % 50a gainditzea espero daiteke, beste teknologien emaitzetara gerturatuz.

5. Etorkizunerako planteatzen den norabidea

Lan ildo honi jarraipena emateko, lehenik eta behin corpusaren tamaina handitzeko beharra ikusten da, gutxienez corpusaren tamaina bikoiztu beharko litzateke, hobekuntzarik dagoen ikusteko, eta ahal dela EPEC corpus osoa (300.000 hitz) etiketatu. Etiketatze hau modu erdi-automatikoan egin daiteke, aipamenak eta korreferentzia erlazioak aipamen detektatzailea eta Standfordeko sistema erabiliz etiketatu eta ondoren zuzenduz.

Bestalde, azken urtean neurona-sareen arkitektura konplexuagoak erabiliz, hobekuntza nabarmenak lortu dira hainbat hizkuntzatan. Lee *et al.* (2018) lanean aurkeztzen den arkitekturaren moduko bat eraikitzeak, euskararako ere emaitzak hobetu ditzakeela uste da.

Azkenik, euskarazko corpora handitzeaz eta arkitektura aldatzeaz gain, hizkuntza arteko sistema bat eraiki daiteke (Kundu *et al.*, 2018; Cruz *et al.*, 2018); baliabide askoko hizkuntza batean (ingelesean adibidez) korreferentzia-ebazpenerako sistema bat entrenatuz, eta ondoren sistema hau euskarako korreferentzia-ebazpenerako baliatuz. Hizkuntza batetik ikasitako ezagutza bigarren hizkuntza bati aplikatuta emaitza onak lortu baitira hizkuntzaren prozesamenduko ataza ezberdinetan.

6. Erreferentziak

- Aduriz, Itziar, Maxux Aranzabe, J Arriola, Aitziber Atutxa, Arantza Diaz-De-Ilarraza, Nerea Ezeiza, Koldo Gojenola, Maite Oronoz, Aitor Soroa, eta Ruben Urizar, 2006. Methodology and steps towards the construction of epec, a corpus of written basque tagged at morphological and syntactic levels for the automatic processing.
- Bagga, Amit, eta Breck Baldwin. 1998. Entity-based cross-document coreferencing using the vector space model. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics-Volume 1*, 79–85. Association for Computational Linguistics.
- Ceberio, Klara, Itziar Aduriz, Arantza Díaz de Ilarraza, eta Ines Garcia-Azkoaga. 2018. Coreferential relations in basque: the annotation process. *Journal of psycholinguistic research* 47.325–342.
- Clark, Kevin, 2015. Neural coreference resolution.
- , eta Christopher D Manning. 2016. Improving coreference resolution by learning entity-level distributed representations. *arXiv preprint arXiv:1606.01323*.
- Cruz, André Ferreira, Gil Rocha, eta Henrique Lopes Cardoso. 2018. Exploring spanish corpora for portuguese coreference resolution. In *2018 Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, 290–295. IEEE.
- Goenaga, Iakes, Olatz Arregi, Klara Ceberio, Arantza Díaz De Ilarraza, eta Amane Jimeno. 2012. Automatic coreference annotation in basque. In *I. HENDRICKX, S. KÜBLER & K. SIMOV (arg.), TLT11 Proceedings of the Eleventh International Workshop on Treebanks and Linguistic Theories. Lisboa: Edições Colibri*, 115–126. Citeseer.
- Grishman, R, eta B Sundheim. 1995. Coreference task definition. version 2.3. In *Proceedings of the Sixth Message Understanding Conference (MUC-6), Columbia, Maryland, USA*, 335–344.

- Kundu, Gourab, Avirup Sil, Radu Florian, eta Wael Hamza. 2018. Neural cross-lingual coreference resolution and its application to entity linking. *arXiv preprint arXiv:1806.10201* .
- Lee, Heeyoung, Angel Chang, Yves Peirsman, Nathanael Chambers, Mihai Surdeanu, eta Dan Jurafsky. 2013. Deterministic coreference resolution based on entity-centric, precision-ranked rules. *Computational Linguistics* 39.885–916.
- Lee, Kenton, Luheng He, Mike Lewis, eta Luke Zettlemoyer. 2017. End-to-end neural coreference resolution. *arXiv preprint arXiv:1707.07045* .
- , Luheng He, eta Luke Zettlemoyer. 2018. Higher-order coreference resolution with coarse-to-fine inference. *arXiv preprint arXiv:1804.05392* .
- Luo, Xiaoqiang. 2005. On coreference resolution performance metrics. In *Proceedings of the conference on human language technology and empirical methods in natural language processing*, 25–32. Association for Computational Linguistics.
- Niton, Bartłomiej, Pawel Morawiecki, eta Maciej Ogródniczuk. 2018. Deep neural networks for coreference resolution for polish. In *LREC*.
- Park, Cheoneum, KyoungHo Choi, Changki Lee, eta Soojong Lim. 2016. Korean coreference resolution with guided mention pair model using deep learning. *ETRI Journal* 38.1207–1217.
- Pradhan, Sameer, Xiaoqiang Luo, Marta Recasens, Eduard Hovy, Vincent Ng, eta Michael Strube. 2014. Scoring coreference partitions of predicted mentions: A reference implementation. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 30–35, Baltimore, Maryland. Association for Computational Linguistics.
- , Lance Ramshaw, Mitchell Marcus, Martha Palmer, Ralph Weischedel, eta Nianwen Xue. 2011. Conll-2011 shared task: Modeling unrestricted coreference in ontonotes. In *Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task*, 1–27. Association for Computational Linguistics.
- Soraluze, Ander, 2017. *Korreferentzia-ebazpena euskarazko testuetan*. University of The Basque Country tesia.
- , Olatz Arregi, Xabier Arregi, eta Arantza Díaz de Ilarraza. 2017a. Enriching basque coreference resolution system using semantic knowledge sources. In *Proceedings of the 2nd Workshop on Coreference Resolution Beyond OntoNotes (CORBON 2017)*, 8–16.
- , Olatz Arregi, Xabier Arregi, eta Arantza Díaz de Ilarraza. 2017b. Improving mention detection for basque based on a deep error analysis. *Natural Language Engineering* 23.351–384.
- , Olatz Arregi, Xabier Arregi, Arantza Diaz de Ilarraza, Mijail Kabadjov, eta Massimo Poesio. 2016. Coreference resolution for the basque language with bart. In *Proceedings of the Workshop on Coreference Resolution Beyond OntoNotes (CORBON 2016)*, 67–73.
- , Olatz Arregi, Xabier Arregi, eta Arantza Díaz de Ilarraza. 2015. Coreference resolution for morphologically rich languages. adaptation of the stanford system to basque. *Procesamiento del Lenguaje Natural* .
- Urbizu, Gorka, 2018. Neurona-sareetan oinarritutako korreferentzia-ebazpen automatikoa. <http://hdl.handle.net/10810/29107>.
- Versley, Yannick, Simone Paolo Ponzetto, Massimo Poesio, Vladimir Eidelman, Alan Jern, Jason Smith, Xiaofeng Yang, eta Alessandro Moschitti. 2008. Bart: A modular toolkit for coreference resolution. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Demo Session*, 9–12. Association for Computational Linguistics.
- Vilain, Marc, John Burger, John Aberdeen, Dennis Connolly, eta Lynette Hirschman. 1995. A model-theoretic coreference scoring scheme. In *Proceedings of the 6th conference on Message understanding*, 45–52. Association for Computational Linguistics.
- Wiseman, Sam, Alexander M Rush, eta Stuart M Shieber. 2016. Learning global features for coreference resolution. *arXiv preprint arXiv:1604.03035* .

7. Eskerrak eta oharrak

Lan hau Urbizu (2018) gradu amaierako lanetik eratorria da.