



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

**Poesiaren eskantsio automatikoa:
bi hizkuntzen azterketa**

*Manex Agirrezabal, Iñaki Alegria eta
Mans Hulden*

48-55 or.

<https://dx.doi.org/10.26876/ikergazte.ii.01.06>

ANTOLATZAILEA:



ELKARLANEAN:



LAGUNTZAILEAK:



Poesiaren eskantsio automatikoa: bi hizkuntzen azterketa

Manex Agirrezabal¹ eta Iñaki Alegria¹ eta Mans Hulden²

¹ IXA taldea (UPV/EHU). Manuel Lardizabal pasealekua 1, Donostia

² Department of Linguistics. University of Colorado Boulder. Boulder, CO (USA)

Laburpena

Lan honetan poesiaren eskantsioa, hau da, poemetako egitura erritmikoaren erauztea, burutzen dugu automatikoki. Horretarako hizkuntzaren prozesamenduko ohiko teknikak eta sare neuronaletan oinarritutakoak erabili ditugu. Esperimentazioa bi hizkuntzarekin egin dugu, ingelesarekin eta gaztelarekin. Emaizten arabera, sare neuronalekin lortu ditugu emaitza onenak, bi hizkuntzetan. Gainera, ondoriozta daiteke gaztelarazko poesia erregularragoa dela ingelesezkoa baino. Egungo erronka esperimentuok euskarara aplikatzea da.

Hitz gakoak: literatura, poesia, eskantsioa, ikasketa automatikoa, ingelesa, gaztelera, euskara

Abstract

In this work we perform automatic scansion —the act of metrically representing a line of verse— of poetry. We employ typical Natural Language Processing techniques and go further by using current Deep Learning models. We have tested our systems in two languages; English and Spanish. We got our best results with the Deep Learning frameworks, both in English and Spanish. Furthermore, we can conclude that the scansion of poetry in English is more complex than in Spanish. We intend to apply these models on Basque poetry.

Keywords: literature, poetry, scansion, machine learning, English, Spanish, Basque

1 Sarrera

Hona hainbat hizkuntzetako poemen adibide batzuk:

Beti tiriki tauki eztarrian kanta ardoaz ondo bustita etzedin marranta ²	<i>I don't like to brag and I don't like to boast said Peter T. Hooper, but speaking of toast.¹</i>
	<i>de verdes sauces hay una espesura toda de yedra revestida y llena,³</i>

Lehen zutabean dagoen poema osorik irakurrita, erritmo gorabeheratsu (TA-TAN-TA-TAN) konstante bat hauteman daiteke. Bigarren zutabeko lehen adibidea ahoz irakurriko bagenu, TA-RA-TAN moduko soinu bat hautemango genuke (Butler, 1989). Bigarren adibidea, aldiz, gaztelarazko hendekasilabo bat da, beraz, hamaika soinu unitateko lerroa dugu hura, azken aurreko silaba azentudunarekin. Baina, posible al litzateke horrelako egiturak antzematea hizkuntzaren erabateko ezagutza izan gabe? edo, are gehiago, hizkuntzari buruzko inolako informaziorik gabe, topa al daitezke halako patroiak? Hizkuntzaren Prozesamenduaren arloko erronkatzat har dezakegu poemetako patroia prosodikoen hautemate hau.

Uneko hizkuntzari buruzko informaziorik izan gabe egitura prosodiko hau erauzteko, tradizio poetiko ezberdinen azterketa tipologiko bat egitea beharrezkoa dela uste dugu. Bide horretan lehen pausuak emateko ikerlan hau aurkezten dugu, non poesiaren egitura prosodikoa automatikoki aztertzen dugun hizkun-

¹Dr. Seuss-en *Scrambled Eggs Super!*

²Nikolas Ormaetxearen —*Orixe*— Azkeneko Ola-Gizona

³Garcilaso de la Vega-ren *Égloga III*

tzaren prozesamenduko oinarrizko algoritmo batzuk erabilia. Metodo hauek ingelesezko eta gaztelerazko poemetan aplikatu ditugu emaitza onak lortuaz, eta gure egungo erronka lanok euskarara aplikatzea da.

Honako egitura jarraitzen du artikulu honek: Bigarren atalean eskantsioa definitzen dugu eta tradizio poetiko ezberdinak aurkezten. Horretaz aparte, poesiaren analisi automatikoaren inguruan egin diren lan batzuk zerrendatzen ditugu. Hirugarren atala lanaren muina dela esan dezakegu, hor aurkezten baititugu lan honetarako erabili ditugun corpusak, metodoak eta egindako esperimenduak. Bukaeran, laugarren atalean, esperimientuen ondorioak eta lanaren ondorio orokorrak jartzen ditugu, etorkizunean egiteko ditugun ideia batzuk proposatzearekin batera.

2 Eskantsioa

Poema lerro batean eskantsioa egitea poema horren egitura erritmikoa erauztea da, azentuak, oinak eta errimak adierazita (Baldick, 2015). Lan honetan, ordea, lerro bakoitzaren azentu sekuentzia soilik inferitzen dugu.

<i>I don't like to brag and I don't like to boast</i>	<i>Beti tiriki tauki</i>
<i>said Peter T. Hooper but speaking of toast</i>	<i>eztarrian kanta</i>
	<i>ardoaz ondo bustita</i>
	<i>etzedin marranta.</i>

2.1 Poesia ingelesez

Hainbat liburu idatzi dira ingelesezko poesiaren prosodiaren inguruan, Halle eta Keyser (1971); Corn (1997); Fabb (1997) eta Steele (1999), adibidez. Ingelesezko poesian silabak oin izeneko multzoetan elkartzen dira. Multzo hauek hainbat silabez osatuta daude, baina ohikoenak bi edo hiru silabako multzoak dira. Oin hauetako bakoitzak gutxienez gailentzen den silaba bat izango du, azentuatua kontsideratuko duguna. Egitura ohikoenak ianbikoa (*bal-loon*), trokaikoa (*jun-gle*), daktilikoa (*ac-ci-dent*) eta anapestikoa (*but I'm tel-ling you Liz*) dira (Baldick, 2015).

Metrika tradizionalaren arabera (Fussell, 1965; Steele, 1999), honelako oinez osatua egongo da lerro metriko oro. Lerroon luzera oin kopuruaren arabera izango da, beraz, trimetro batek hiru oin izango ditu, tetrametro batek lau, pentametro batek bost, etab. (hexametro, heptametro, ...). Ingelesezko poesian metrika arruntena pentametro ianbikoa da, adibidez,

oh change thy thought, that I may change my mind.

non bost azentu argi nabaritzen diren eta TA-TAN multzo bakoitzak oin bat osatzen duen. Poemok orokorrean erregularrak diren arren, ohikoa da aldaketa txiki batzuk egitea egiturotan, helburu estetiko edota artistikoekin.

Grant if thou wilt, thou art beloved of many

Aurreko adibidearekin alderatuta, honetan hasieran TAN-TA-TA-TAN moduko soinu bat antzematen da. Aldaketa honi, literaturan bariazio trokaiko deitzen zaio. Gainera, lerroa ianbikoa izanda, bukaerak tonikoa behar luke izan, baina aldaketa ohikoa da silaba azentudun baten ostean silaba ez-azentudun bat gehitzea lerroaren bukaeran.

2.2 Poesia gaztelaniaz

Gaztelerazko poesian hainbat egitura metriko erabili izan dira (Quilis, 1984; Tomás, 1995; Caparrós, 1999). Lan honetan, corpusaren eskuragarritasuna medio, garai espezifiko batean soilik egin dugu enfasia, Espainiako Urrezko Aroan, alegia. Garai honetan gehien erabilitako metrika hendekasilaboa izan zen, lerro bakoitza hamaika silabez osaturik. Lerroetako azentu sekuentzia nahiko erregularra da eta normalean hamargarren silabak azentua darama. Beste silabek ere azentua izan dezakete, eta nabarmendutako posizio horien arabera, hendekasilabo hauek hainbat motatakoak izan daitezke. Azpikategorizazio hau, ordea, lan honen eremutik kanpo dago.

Gaztelerazko poesiaren erronka handienetako bat silaba laburketen erabilera da, sinalefa gisa ezagutzen dena, non hamaika silaba baino gehiago dituzten lerroak hamaika silabetan ahokatzen diren. Lan honen helburua silaba bakoitzari azentu bat automatikoki esleitzea da, ondorioz, metodo erdi-automatiko bat erabili dugu sinalefak dauden kasuetan lerroko silaba bakoitzari azentu balio bat esleitzeko.

2.3 Poesia euskaraz

Gaur egungo poesian, eta bereziki bertsolaritzan, neurri ezagunik bada, neurri txikiak eta handiak dira. Neurri txikiek lerro bakoitietan zazpi silaba izaten dituzte eta bikoitietan sei. Handiek, ordea, hamar silaba eta zortzi silaba izaten dituzte lerro bikoiti eta bakoitietan, hurrenez hurren. Ez dira hauek, ordea, poesian erabiltzen diren neurri bakarrak. Idatzizko poesian ohikoa da zortziko ertainaren erabilera, non lerro bakoitiek zortzi silaba dituzten eta bakoitiek zazpi. Neurri gehienetan lerro bakoitiek elkarrekin errimatu behar dute.

Ikerlan honetan azentuei erreparatzen diegu eta oraindik ez dugu argi ea euskarazko poesian azentuek eragin nabarmena duten ala ez. Hainbat adituk idatzi izan dute euskal poesia eta haren neurkerari buruz, XVII. mendetik hasita. Hauek irakurtzean ikuspegi kontrajarriak topa daitezke. Batzuen arabera —Oihenart eta aita Onaindia, kasu— euskal poesian erritmoak garrantzia du, eta poema orok nolabaiteko erritmoa izan behar du.

...

Literatur guztiak dabez euren lege ta arauak, olerkigintzan bereziki; euskeran be naitaez izan bear. Lau gauza oneik beintzat gogotan artu bearrak doguz: 1) Igikera (ritmu); 2) etena (cesura); 3) neurria, ta 4) oskide edo azken amaitze bardiña (rima).

...

Onaindia (1961)

Beste batzuk, berriz, euskal poesiagintzan azentuak eraginik ez duela dio. Nikolas Ormaetxea —Orixe— da hori esaten duen poeta bat.

...

Para probar lo poco sensible que es el acento vasco, inténtese colocar acentos gráficos en las silabas que uno crea acentuadas, encárguese el trabajo a cien personas de buen oído y en una página que se someta al análisis, se puede asegurar sin temor, que no habrá dos que coincidan.

...

Ormaechea (1920)

2.4 Eskantsio automatikoa

Azken urteotan eskantsio automatikoaren inguruan lan ezberdinak egin dira. Lan hauetan, hitz sekuentzia bat sarrera gisa jasota, hauek jarraitzen duten azentu sekuentzia itzultzea izan ohi da burutu beharreko ataza. Itzulpen edo transdukzio prozesu hau hainbat modutara egin daiteke:

- Erregeletan oinarrituta: Adituek ezarritako **arauak** jarraituta, hainbat ezaugarri linguistiko kontutan izanda.
- Datuetan oinarrituta: Etiketatutako informazioan oinarrituta, testutik azentuetarako patroiak **automatikoki** ikasita. Ildo honi jarraitu diogu aurkezten dugun lan honetan.

Urteotan aurkeztu diren lanen artean, arauetan oinarritutakoak Logan (1988); Gervas (2000); Hartman (2005); Plamondon (2006); McAleese (2007); Navarro-Colorado (2015) eta Agirrezabal *et al.* (2016b) ditugu. Geroz eta entzute handiagoa dute datuetan oinarritutako metodoek, etiketatutako informazioaren eskuragarritasuna dela eta. Hauen artean Hayward (1996); Greene *et al.* (2010); Hayes *et al.* (2012); Agirrezabal *et al.* (2016a) eta Estes eta Hench (2016) azpimarratu ditzakegu.

3 Corpusak, metodoak eta esperimentuak

3.1 Corpusak

Datuetan oinarritutako sistemen garapenerako datu etiketatuak izatea ezinbestekoa da. Horretarako bi corpus erabiltzen ditugu, bata ingelesez eta bestea gazteleraz. Ingeleseko lanetarako Virginiako unibertsitatean garatutako “*For Better For Verse*” proiektuaren (Tucker, 2011) emaitza izan den poesia corpusa⁴ erabili dugu. Corpus honetan 78 poema daude eta guztira 1.100 poema lerro. Eskantsioa egiterako orduan, lerro batzuk hainbat analisi izan ditzakete, eta hauek corpusean horrela daude (hainbat aukerarekin). Gaztelerazko esperimentuetarako, lehenago aipatu gisa, Espainiako Urrezko Aroko corpus bat erabili dugu (Navarro-Colorado *et al.*, 2016).⁵ Etiketatutako corpusa 135 sonetoz osatuta dago eta gutxi gorabehera 2.000 lerro ditu.

3.2 Metodoak

Hizkuntzaren prozesamenduan ohikoak diren teknikak aplikatu ditugu datuotatik patroiak ikasi eta aurretik ikusi gabeko poemetan aplikatu ahal izateko. Aurretik garatutako Gervas (2000); Hartman (2005) eta Agirrezabal *et al.* (2016b) sistemen emaitzak erabiltzen ditugu oinarri-lerrotzat (lehen gazteleraz eta beste biak ingelesez). Erabilitako teknikak perzeptroia (*Perceptron*) (Freund eta Schapire, 1999), Markoven eredu ezkutuat (*Hidden Markov Models*) (Rabiner, 1989), ausazko eremu baldintzatuak (*Conditional Random Fields*) (Lafferty *et al.*, 2001) edota epe laburreko memoria luzedun sare neuronal errekurrentek (*Recurrent Neural Networks with Long Short-Term Memory*) (Lample *et al.*, 2016) dira.

Aurrerago ikusiko den moduan, emaitza onenak ausazko eremu baldintzatuekin, CRFekin, eta sare neuronaletan oinarritutako sistemekin, NNekin, lortu ditugu. CRFek aurreikuspena egiteko sekuentziako elementu bakoitzari buruzko informazio osagarria (informazio morfosintaktikoa, azentu lexikala, ...) jasotzen dute eta informazio horren arabera, predikzio bat proposatuko du. Predikzio hori, ordea, ez da beste proposamen guztiekiko independentea izango, *Viterbi* izeneko algoritmo bat erabiltza irteerako sekuentzia optimoa proposatuko baitu ereduak. Adibidez, hona sistemak proposa lezakeen irteera bat pentametro ianbikoan idatzita dagoen poema batentzako:

-	1.0	0.0	1.0	0.0	0.45	0.0	1.0	0.0	1.0	0.0
+	0.0	1.0	0.0	1.0	0.55	1.0	0.0	1.0	0.0	1.0
MAX	-	+	-	+	+	+	-	+	-	+

Zutabe bakoitzeko balioak silaba azentuduna edo ez-azentuduna izateko probabilitateak lirateke. Kasu bakoitzean silaba bakoitzaren azentu probabilitate maximoa hartuz gero, hirugarren lerroan agertzen den proposamena egingo litzateke. Baina, bosgarren silabaren predikzioa ez da besteetakoa bezain segurua. Gainera, pentametro ianbikoaren izaeragatik, ++ sekuentziak dira ohikoak, beraz sekuentzia hoberena bilatzen duen algoritmoak, 5. silaba ez toniko gisa proposatuko luke seguruenik.

Sare neuronalak eragiketa sinple asko egiten dituzten ereduak dira. Guk erabiltzen dugun Lample *et al.* (2016) laneko ereduak bi mailatan egiten du lan. Lehenik eta behin hitz bakoitza modelatzen du bere karaktereen arabera, horrekin hitzaren zenbakizko errepresentazio bat sortuz (hainbat zenbaki izan ohi ditu honek). Honekin, hitz bakoitzaren informazioa (sasi-informazio linguistikoa) zenbaki sekuentzia baten bidez adieraziko da, lehen mailako informazioa. Esaldian daukagun hitz bakoitzarentzako, aurreikuspen bat sortu behar dugu eta horretarako uneko hitzaz haratago, testuinguruaren informazioa izateak berebiziko garrantzia du. Adibidez, ondorengo adibideko lehen kasuan,

to be or not to be || I must be sure a bout

be aditza *to* eta *or* hitzen artean egonda, nabaria da azentuduna behar duela izan. Bigarren kasuan, aldiz, *be* aditza *must* eta *sure* hitzen artean egonda, beste hitzak gailentzen dira. Sare neuronalean, beraz, hitz bakoitzaren informazio linguistikoa ezkerretik dauden hitzen informazioa eta eskuinean dauden hitzen informazioa gehitzen zaio eta guzti horren konbinaziotik, emaitza bat sortzen da, ea hitzaren azentu

⁴<http://prosody.lib.virginia.edu/>

⁵<https://github.com/bncolorado/CorpusSonetosSigloDeOro>

patroia zein den. Hau dena eragiketa sinpleen konbinazioz egiten da. Eragiketa hauez haratago, CRF geruza bat ere gehitzen zaio eta honi esker irteera sekuentzia optimoa lortzen da.

3.3 Esperimentuak

Poema lerro bat emanda haren azentu patroia erauztea da gure ataza. Berau bi modutara egin daiteke: sarrera lerroa silabetan banatu eta silaba bakoitzari azentu bat esleituta (*syllable-to-stress*, S2S), edo hitz bakoitzari bere azentu patroia asignatuaz (*word-to-stress-pattern*, W2SP). Hona bi adibide:

	Poema lerroa	Azentu patroia
S2S	There fore my verse to cons tan cy con fined,	[+] [-] [-] [+] [-] [+] [-] [+] [-] [+]
W2SP	Therefore my verse to constancy confined,	[+ -] [-] [+] [-] [+ -] [-]

Esaldiko hitzak silabetan banatuta, hitzaren egitura ikusgaitz geratzen denez, sare neuronaletako esperimentu batzuetan hitz-muga gehitu dugu (*word boundary*, WB), hitzaren informazioa nolabait sekuentzian txertatzeko. Informazio hau lagungarria da, adibidez, *therefore* eta *forehead* bezalako hitzetan, zeinetan *fore* silaba bietan agertzen den, baina bukaeran agertzen denean azentuatu gabea den.

Ikasketa automatikoko tresnak erabiltzerakoan ohikoa da aztertzen den instantzia bakoitzari buruzko informazio osagarria erauztea. Perzeptroien eta ausazko eremu baldintzatuen kasuan bi atributu konfiguraziorik egin ditugu esperimentuak. Konfiguraziooi buruzko informazio osatua, hau da, atributu bakoitzaren informazioa eta justifikazioa, Agirrezabal *et al.* (2016a) lanean topa daiteke.

Teknika eta konfigurazio ezberdinak ebaluatzeko, metodo ezberdinak erabil daitezke. Datu kopurua oso handia ez denean, gure kasuan bezala, balidazio gurutzatua (*K-fold Cross-Validation*) erabiltzea da ohikoena. Balidazio gurutzatua datu multzoa k zatitan banatzen da. Behin zati horiek eginda, $k - 1$ zati erabiltzen dira eredu bat ikasteko eta ebaluaziorako bat gordetzen da. Hau k aldiz egiten da, eta asmatze-tasaren batzbestekoa itzultzen da. Gure kasuan, 10 zatitan banatu dugu gure datu-multzoa.

Konfigurazio eta teknika ezberdinen emaitzak 1. taulan aurki daitezke, aurretik garatutako sistemekin alderatuta.

	Ingelesa		Gaztelera	
	Silaba AT (%)	Lerro AT (%)	Silaba AT (%)	Lerro AT (%)
Gervas (2000)	-	-	-	88.73
ZeuScansion (Agirrezabal <i>et al.</i> , 2016b)	86.78	26.21	-	-
Scandroid (Hartman, 2005)	89.78	42.95	-	-
Perceptron ₁₀ (S2S)	84.86	29.32	74.54	0.31
Perceptron ₆₄ (S2S)	89.34	43.36	92.25	40.78
HMM (S2S)	90.43	49.88	92.57	45.40
CRF ₁₀ (S2S)	89.66	50.16	85.30	19.20
CRF ₆₄ (S2S)	91.41	55.30	93.37	57.00
NN (S2S)	91.26	55.28	95.13	63.68
NN hitz mugarekin (S2S)	92.96	61.39	98.74	88.81
NN (W2SP)	89.39	44.29	98.96	91.57

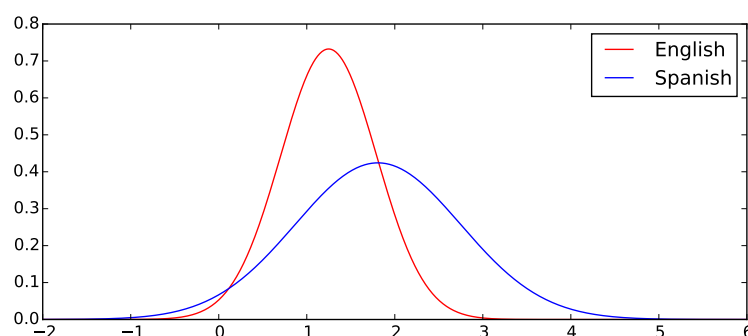
1 Taula: Lehen hiru lerroetan oinarri-lerroen asmatze-tasa (AT) ikus daiteke. Asmatze-tasa poemako silabaka (*Silaba AT*) edo lerro osoka (*Lerro AT*) ebaluatzen da. Bigarren multzoan Agirrezabal *et al.* (2016a) lanean ingelesezko esperimentuak daude, gaztelera egin ditugun esperimentuekin batera. Azken hiru lerroetan sare neuronalekin egin ditugun proben emaitzak ikus daitezke.

4 Emaitzak eta etorkizunerako lana

Emaitzen taulan ikus daitekeen moduan, sare neuronaletan oinarritutako sistemek ematen dituzte emaitza onenak, bai ingelesez eta baita gaztelera ere. Taula horretatik hainbat ondorio plazara ditzakegu.

Agirrezabal *et al.* (2016a) lanean adierazi genuen Perzeptroiean eta CRFetan erabiltzen ditugun 10 atributuak poesiaren analisi prosodikorako egokiak ziren atributuak zirela, bereziki interesgarriak hizkuntzarekiko agnostikoak ziruditelako. Esperimentuotan, gazteleraz probak egin ostean, ikusi dugu ingelesez nahiko emaitza onak ematen dituztela haien sinpletasuna kontutan hartuta. Gaztelerazko datuetan, ordea, emaitzak ez dira horren onak izan eta horrek iradokitzen digu atributuok ez direla nahikoak hizkuntzarekiko independenteak diren sistemak eraikitzeke. Dena den, hau baieztatzeko hizkuntza gehiagorekin egin beharko genituzke esperimentuok.

Emaitzak aztertuta, hitz mugak poemetako egitura prosodikoaren inferentzian garrantzi handia duela ondorioztatzen dugu, bereziki gazteleraz. Horren justifikazioa izan daiteke ingelesezko hitzek batazbestean silaba gutxiago dituztela gazteleraz baino, 1. irudian ikus daitekeen bezalaxe. Gainera, badirudi sare neuronaletan oinarritutako ereduak hitzen egitura fonologikoa ondo modelatzen dutela, baina hau enpirikoki frogatzeko esperimentu gehiago beharko lirateke.



1 Irudia: Batazbesteko silaba kopurua hitz bakoitzeko ingelesez eta gazteleraz erabili ditugun corpusetatik kalkulaturak.

Hemen aurkeztu dugun lan honek poesia analizatzaile orokor eta ez gainbegiratu baten lehen adreilu izan nahi du.

Badira hainbat aspektu etorkizunean landu nahiko genituzkeenak. Lehenik eta behin, lanok euskarazko corpus batean aplikatu nahiko genituzke, horrela euskarazko poesiaren analisi prosodikoa egin ahal izateko eta honetan bide berriak jorratu ahal izateko. Horretarako poema bilduma bat metrikoki etiketatzeko bidean gabiltza, ingelesezko corpusaren antzera. Oinarri gisa, Patri Urkizuren euskal poesia antologia (Urquizu Sarasua, 2009) hartu dugu.

Astigarraga *et al.* (2013); Agirrezabal *et al.* (2013) lanetan bertsoen sorkuntzan egin dugu lan, baina orain arte azentuarekin erlazionatutako aspektuak kontutan izan gabe. Aspektuok bertsoen sorkuntzarako garrantzitsuak izan litezkeela deritzogu, beraz, garatuta dauzkagun poesia sortzaileetan orain arte ikasitakoa aplikatu genezake.

Erreferentziak

- AGIRREZABAL, MANEX, IÑAKI ALEGRIA, eta MANS HULDEN. 2016a. Machine Learning for Metrical Analysis of English Poetry. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics, Osaka, Japan*, 772–781.
- , BERTOL ARRIETA, AITZOL ASTIGARRAGA, eta MANS HULDEN. 2013. POS-tag based poetry generation with WordNet. *Proceedings of the 2013 European Workshop on Natural Language Generation* p. 162.
- , AITZOL ASTIGARRAGA, BERTOL ARRIETA, eta MANS HULDEN. 2016b. ZeuScansion: a Tool for Scansion of English Poetry. *Journal of Language Modelling* (4).3–28.
- ASTIGARRAGA, AITZOL, MANEX AGIRREZABAL, ELENA LAZKANO, EKAITZ JAUREGI, eta BASILIO SIERRA. 2013. Bertsobot: the first minstrel robot. *6th International Conference on Human System Interaction, Gdansk*. 129–136.

- BALDICK, CHRIS. 2015. *The Oxford Dictionary of Literary Terms*. Oxford University Press.
- BUTLER, FRANCELIA. 1989. Seuss as a creator of folklore. *Children's Literature in Education* 20.175–181.
- CAPARRÓS, JOSÉ DOMÍNGUEZ. 1999. *Diccionario de Métrica Española*. Alianza Editorial.
- CORN, ALFRED. 1997. *The Poem's Heartbeat: A Manual of Prosody*. Copper Canyon Press.
- ESTES, ALEX, eta CHRISTOPHER HENCH. 2016. Supervised Machine Learning for Hybrid Meter. *on Computational Linguistics for Literature* 1–8.
- FABB, NIGEL. 1997. *Linguistics and Literature: Language in the Verbal Arts of the World*. Blackwell. 2nd ISBN: 9780631192435.
- FREUND, YOAV, eta ROBERT E SCHAPIRE. 1999. Large Margin Classification Using the Perceptron Algorithm. *Machine learning* (37).277–296.
- FUSSELL, PAUL. 1965. *Poetic Meter and Poetic Form*. McGraw Hill.
- GERVAS, PABLO. 2000. A Logic Programming Application for the Analysis of Spanish Verse. In *Computational Logic—CL 2000*, 1330–1344. Springer.
- GREENE, ERICA, TUGBA BODRUMLU, eta KEVIN KNIGHT. 2010. Automatic Analysis of Rhythmic Poetry with Applications to Generation and Translation. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, 524–533. Association for Computational Linguistics.
- HALLE, MORRIS, eta SAMUEL JAY KEYSER. 1971. English stress: Its form, its growth, and its role in verse.
- HARTMAN, CHARLES O., 2005. The Scandroid 1.1.
- HAYES, BRUCE, COLIN WILSON, eta ANNE SHISKO. 2012. Maxent Grammars for the Metrics of Shakespeare and Milton. *Language* (88).691–731.
- HAYWARD, MALCOLM. 1996. Analysis of a Corpus of Poetry by a Connectionist Model of Poetic Meter. *Poetics* 24.1–11.
- LAFFERTY, JOHN, ANDREW MCCALLUM, eta FERNANDO CN PEREIRA. 2001. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data.
- LAMPLE, GUILLAUME, MIGUEL BALLESTEROS, SANDEEP SUBRAMANIAN, KAZUYA KAWAKAMI, eta CHRIS DYER. 2016. Neural Architectures for Named Entity Recognition. In *Proceedings of NAACL-2016, San Diego, California, USA*. Association for Computational Linguistics.
- LOGAN, HARRY M. 1988. Computer Analysis of Sound and Meter in Poetry. *College Literature* 19–24.
- MCALKEESE, G., 2007. *Improving Scansion with Syntax: an Investigation into the Effectiveness of a Syntactic Analysis of Poetry by Computer using Phonological Scansion Theory*. Open University tesia.
- NAVARRO-COLORADO, BORJA. 2015. A Computational Linguistic Approach to Spanish Golden Age Sonnets: Metrical and Semantic Aspects. *Computational Linguistics for Literature* p. 105.
- , MARIA RIBES LAFOZ, eta NOELIA SÁNCHEZ. 2016. Metrical Annotation of a Large Corpus of Spanish Sonnets: Representation, Scansion and Evaluation. In *Proceedings of the Language Resources and Evaluation Conference*.
- ONAINDIA, AITA. 1961. Euskal poesi neurkera. *Olerti* (2).83–91.
- ORMAECHEA, NICOLÁS. 1920. De la música y letra popular al acento.
- PLAMONDON, MARC R. 2006. Virtual Verse Analysis: Analysing Patterns in Poetry. *Literary and Linguistic Computing* 21.127–141.
- QUILIS, ANTONIO. 1984. *Métrica Española*. Ariel Barcelona.
- RABINER, LAWRENCE R. 1989. A Tutorial on Hidden Markov Models and selected Applications in Speech Recognition. *Proceedings of the IEEE* 77.257–286.
- STEELE, TIMOTHY. 1999. *All the Fun's in how you Say a Thing: an Explanation of Meter and Versification*. Ohio University Press Athens.
- TOMÁS, NAVARRO TOMÁS. 1995. *Métrica Española*. Ed. Labor.

TUCKER, HERBERT F. 2011. Poetic Data and the News from Poems: A for Better for Verse Memoir. *Victorian Poetry* 49.267–281.

URQUIZU SARASUA, PATRICIO. 2009. Poesía vasca. antología bilingüe.

5 Eskerrak eta oharrak

Eskerrak eman nahi dizkiogu Euskal Herriko Unibertsitateari ikerketa lan hau diruz laguntzeagatik. Lan honen oinarria CoLing (*International Conference on Computational Linguistics*) kongresuan izan genuen artikulu bat izan da. Eskerrak ematen dizkiegu ere artikulu hau berrikusi eta hobekuntzak proposatu dizkien berrikusleei. Azkenik, Udako Euskal Unibertsitateari ematen dizkiogu eskerrak, *IkerGazte* bezalako bilkura bat antolatzeagatik.