

Erantzunen kalifikazio automatikorako lehen urratsak

*Eneko Agirre, Itziar Aldabe, Oier Lopez de Lacalle,
Iñigo Lopez-Gazpio* eta Montse Maritxalar*

Lengoaia eta Sistema Informatikoak saila. Informatika Fakultatea (UPV/EHU)

* inigo.lopez@ehu.eus

DOI: 10.1387/ekaia.14530

Jasoa: 2015-05-23

Onartua: 2015-09-25

Laburpena: Lan honetan erantzunen kalifikazio automatikorako lehen urratsak azaltzen ditugu. Urrats horiek azaltzeko, irakaskuntzaren domeinuko bi sistema deskribatzen ditugu: bata, erantzunak hierarkikoki antolatzen eta kalifikatzen dituena eta bestea, kalifikazioaren azalpena eman ahal izateko erantzunaren zatiak erreferentzia batekin alderatzeko gai dena. Horretaz gain, alorreko teknika erabilienak mahairatuko ditugu, irakaskuntzaren alorreko ataza nazioarteko bat aztertuz.

Hitz-gakoak: Hizkuntzaren prozesamendua, erantzunen kalifikazio automatikoa, adimen artifiziala, ikasketa automatikoa.

Abstract: This work describes the first steps towards automatic response analysis. To explain these steps, we describe two systems from the educational domain: on the one hand, a system that uses hierarchical structures to grade answers; and, on the other hand, a system capable of comparing answer segments in order to try to explain the grading. In addition, we analyze an international educational domain task and describe the principal techniques used.

Keywords: Natural language processing, automatic response analysis, artificial intelligence, machine learning.

1. SARRERA

Hizkuntzaren Prozesamendua (HP) alorrak informatika, adimen artifiziala eta hizkuntzalaritza biltzen ditu eta hainbat hizkuntza-teknologiak arduratzen da [1]. HPko adituek erronka nagusi modura dute hizkuntza sortzeko, ulertzeko zein interpretatzeko gai diren aplikazioak garatzea. Hala eta guztiz ere, aplikazio horiek garatzea ez da lan samurra, hizkuntza uler-

tzeko munduaren ezagutza beharrezkoa baita; jakina da gainera makinek ez dutela inguratzen gaituen munduaren noziorik. Horregatik, hein handi batean, HPk helburu modura hartu du gizakiok jaioberriak garenetik barneratzen dugun ezagutza makinei helaraztea. Lan honetan hezkuntzari edo pedagogiari lotutako alorrari helduko diogu HPko ikerketa-lerroen artean. Horretarako, testu-ulermena lantzeko baliagarriak diren hizkuntza-teknologietan sakonduko dugu.

Testu-ulermena gizakiok bizi heinean garatzen dugun gaitasun bat da, irakurketaren bidez jasotzen dugun informazioaren erantzule dena. Oro har, testu-ulermena garatzeko galdera itxiak zein galdera irekiak erabili ohi dira. Teknika bakoitzak abantaila zein desabantaila batzuk dakartza. Esaterako, aukera anitzeko galderak (galdera itxien adibide direnak) ezagutza hautemateko oso baliagarriak badira ere, azterketa batzuetan zalantzagarritzat jo da kontzeptuak modu sendoan barneratzeko eraginkortasuna [2-3]. Aldiz, galdera irekiei erantzuteak ahalegin handiagoa eskatzen du eta kontzeptuak argi antolatuak izan behar ditugu eta beraz, ondorioztatu da aukera ematen duela trebakuntza-arroetan ezagutza modu eraginkorragoan finkatzeko [4].

Ikerketa honetan galdera irekien bideari helduko diogu; izan ere, galdera irekien sorkuntza eta erantzunen kalifikazio automatikoa direlako hizkuntza-teknologiaren beharra duten alorrak [5]. Zehazki, lan honek bi helburu nagusi ditu. Alde batetik, ikasle-erantzunak automatikoki kalifikatzeko gai den oinarritzko sistema bat hartuko dugu, eta berau hobetu daitekeen aztertuko dugu. Horretarako, SemEval 2013ko ikasle-erantzunak nahasmenaren arabera hierarkizatuko ditugu, ataza nagusia hainbat azpiatazatan deskonposatzeko eta kalifikazioen esleipenak egiteko. Beste alde batetik, ikerketa honen helburua ere bada esaldi-antzekotasuna egoera berri batean erabilgarria den ebaluatzea. Hau da, esaldi osoak zuzenean alderatu beharrean esaldiak sintagmen arabera deskonposatu, eta sintagma horien arteko antzekotasunak alderatzen dira. Zehatz-mehatz, esaldien arteko *gainezarpen partziala* da hurbilpen honekin landu nahi duguna. Hain zuzen ere, esaldien barruko segmentuak identifikatzeko eta erreferentzia baten aurka segmentuz segmentu alderatzeko. Abiapuntu modura, onartu dugu segmentu hauen azterketak esaldi-antzekotasuna hobeto hautematen lagunduko digula.

2. IKASLEEN ERANTZUNAK KALIFIKATZEKO TEKNIKAK

HPko teknologia-baliabide gehiago dituzten hizkuntzekin alderatuta urrats hutsalak badira ere, euskarazko galdera irekien sorkuntza automatikoa lantzen hasiak gara ikertzaileok.

Adibidez, erabiltzaileen elkarrekintzan oinarritzen den web-aplikazio baten prototipoa deskribatzen dute Lopez-Gazpiok eta Maritxalar Angla-

dak lan honetan: [6]. Web-aplikazio hori erabiliz erabiltzaileak gai dira testu-fitxategiak sisteman gordetzeko, eta galdera-sortzaile automatiko baten laguntzaz, testuen gaineko galderak sortzeko. Azaleko syntaxian oinarrituta galdera irekiak sortzeko sistema bakuna bada ere [7-9], ikerketa-lerro zabal baten barruan kokatuta dago. Dagoeneko, sistema konplexuagoak ere garatu dira azaleko syntaxiaz gain, hitzen arteko dependentzietan oinarrituta galderak sortzeko [10]. Gauzak horrela, komenigarria izango litzateke, galdera irekiak sortzeaz gain, galdera irekiei emandako erantzunak automatikoki kalifikatzeko gai lirakeen hizkuntza-baliabideak garatea: bai erabiltzaileen ezagutza maila kuantifikatzeko, baita egindako akatsen azalpenak emateko ere.

Oro har, erantzunen kalifikazio automatikoari aurre egiteko bi estrategia nagusi daude: bata, *testu-inferentzia*z baliatuta erantzunean hautesmandako kontzeptuak identifikatu eta erreferentziaren aurka alderatuz kalifikatzea, eta bestea, erantzuna erreferentziarekin alderatu eta hurbiltasun sintaktiko-semantikoaren araberrako kalifikazioa esleitzea. Ondoren azaletik azalduko ditugu bi teknikak (bi tekniken arteko norgehiagoka garbia dago artearen egoeran), baina lan honetan *esaldi-antzekotasunari* helduko diogu. Alde batetik, testu-inferentziaren aldekoek esaldi-antzekotasuna kritikatzan dute, ez direlako gai kontzeptuak behe mailan marra gabe identifikatzeko eta kontzeptuen arteko erlazioak zehaztasunez hautemateko [11]. Bestalde esaldi-antzekotasunaren aldekoek testu-inferentzia kritikatzan dute, esaldi-antzekotasuna modu orokorragoan lan egiteko egokiagoa dela argudiatuz [12]. Esate baterako, testu-inferentziaren ikuspuntutik auto jakin bat ibilgailua dela ondoriozta daiteke, baina ez ibilgailu bat autoa denik (ibilgailu mota ugari egon daitezkeelako). Aitzitik, esaldi-antzekotasunak autoaren eta ibilgailuaren arteko hurbiltasun semantikoaren kalkulua ahalbidetuko liguke.

Testu-inferentzia sistemak honela laburbil daitezke: adituaren erantzuna hipotesizat (H) hartzen dute eta ikaslearen erantzuna testutzat (T). Ondoren, H hipotesia T testutik ondoriozta ote daitekeen aztertzen dute, hau da, logika mailan $T \rightarrow H$ formalizatzen dute. Esaldi-antzekotasunean oinarritzen diren sistemek aldiz, adituaren eta ikaslearen erantzunen artean hurbiltasuna edo urruntasuna islatzen duten hainbat ezaugarri kalkulatzan dituzte; era berean, kalifikazioak esleitzeko prozedurak ere ikasten dituzte.

3. SEMEVAL

Ondorengo lerroetan kalifikazio automatikoak emateko gai diren sistemei helduko badiegu, beharrezkoa da SemEval 2013ko 7. atazari erreparatzea [13]. Ataza honek ikasle-erantzunei kalifikazio egokiak esleitzen diz-

kie hainbat zientzia-domeinutatik erauzitako galdera eta erreferentziazko erantzunak kontuan hartuta. Edozein kasutan, ikasle-erantzun guztiei —doitasunaren arabera— kalifikazio kualitatibo bat esleitzen zaie. Kalifikazioak hauetariko bat izan behar du: zuzena, erdizka edo hein handi batean osatu gabea, kontraesana, garrantzirik gabekoa edo domeinuz kanpoko. 1. taulan ikus daitezke adibide gisa, jatorrizko atazatik euskaratutako galdera, bi adituren erreferentziazko erantzunak eta lau ikasleren erantzunak. 2. taulak ikasleei esleitu beharreko kalifikazioak zerrendatzen eta deskribatzen ditu.

1. taula. SemEval 2013ko 7. atazako datu-basetik euskaratutako adibide bat.

Galdera	Zergatik neurtu da 4,5 voltetako tentsio-erorketa A eta B terminalen artean?
Adituaren erantzuna (1)	A terminala polo positiboarekin lotu delako.
Adituaren erantzuna (2)	B terminala polo negatiboarekin lotu delako.
Ikaslearen erantzuna (1)	Ez dakit.
Ikaslearen erantzuna (2)	A eta B terminalak polo berriekin lotu direlako.
Ikaslearen erantzuna (3)	A terminala polo negatiboarekin lotu ez delako.
Ikaslearen erantzuna (4)	A terminala polo negatiboarekin lotu delako.

2. taula. SemEval 2013ko 7. atazan sistemek ikasleen erantzunei esleitu beharreko kalifikazioen zerrenda eta deskribapena.

Kalifikazioa	Deskribapena
Zuzena	Erantzunak adituaren erreferentziaren informazio bera ematen du.
Erdizka	Erantzuna ontzat emateko kontzeptu batzuk aipatzea falta da.
Kontraesana	Erantzunak esplizituki aurka egiten dio adituaren erreferentziazko erantzunari.
Garrantzirik gabe	Domeinuko kontzeptuak erabili dira baina erantzuna ez da esanguratsua.
Domeinuz kanpo	Ez dira galderaren domeinuko kontzeptuak erabili erantzuteko.

Adibide gisa, 1. taulako ikasle-erantzunak kalifikatuko bagenitu, baziko kalifikazioak honako hauek lirateke:

1. ikaslearen kalifikazioa: Domeinuz kanpo.
2. ikaslearen kalifikazioa: Garrantzirik gabe.
3. ikaslearen kalifikazioa: Zuzena.
4. ikaslearen kalifikazioa: Kontraesana.

Argi dago makina batek zerbait gehiago jakin behar duela kalifikazio horiek ematen ikasteko azaleko hitzen esanahia eta sintaxiaren inguruko informazioa baino. Horretarako, hizkuntzaren azterketan eta prozesamenduko teknikan murgiltzea ezinbestekoa da makinak domeinuaren inguruko ezagutza bereganatzeko.

SemEval 2013ko atazan bederatzi sistemak hartu zuten parte, baina, desberdintasun guztien gainetik, sistema guztiak hiru multzotan sailka daitezke: esaldi-antzekotasunean oinarritutakoak, testu-inferentzian oinarritutakoak, eta, azkenik, aurreko bi metodoen konbinazioak egiten dituzten sistema hibridoak.

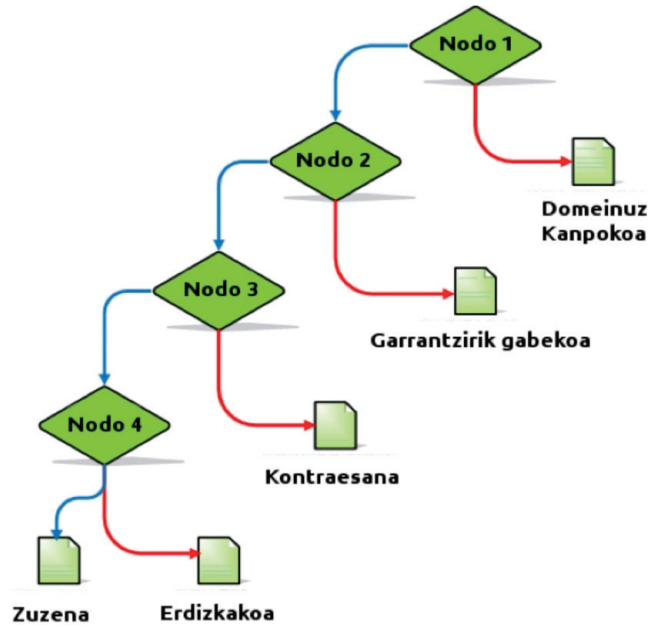
4. KATEGORIA-NAHASMENEAN OINARRITUTAKO SISTEMA

EHU-ALM sistema hau, gai da ezaugarri sintaktiko eta semantikoe-tan oinarrituta ikasleen erantzunak automatikoki kalifikatzeko [14]. Ikas-keta gainbegiratuan oinarrituta dago, eta, guztira, hogeita hamar antzeko-tasun-ezaugarri erabiltzen ditu bukaerako kalifikazioa esleitzeko. SemEval atazako emaitzek aitortzen duten modura sistema eraginkorra da, nahiko emaitza onak lortzen ditu-eta [13].

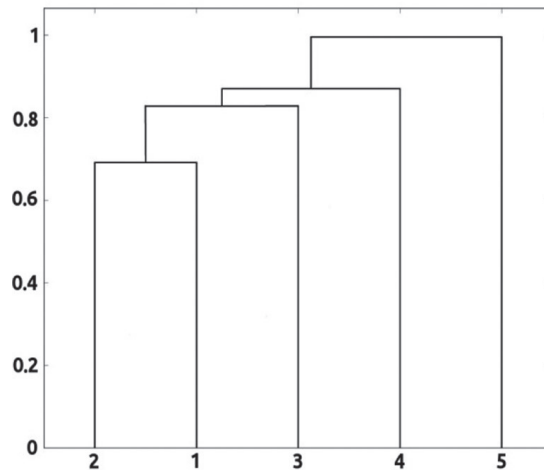
Informatikaren alorrean, atazak zatitu eta errazteko teknika asko erabiltzen dira. Horrek aukera ematen du ataza konplexu bat hainbat azpiata-zatan deskonposatzeko. Ideia horri helduz, eta SemEval 2013ko emaitzak hobetzeko asmoz, ondorengo zalantza plazaratu genuen: ba al da SemEval 2013ko ataza deskonposatzeko aukerarik? Abiapuntu modura onartuko dugu oinarritzko sistema bat gero eta eraginkorrago bihur daitekeela baldin eta SemEval-eko klase-erlazioak modelatzeko —ulertzeko— gai baldin bada; alegia, klase-erlazioak kontuan hartuta sailkapenak egiteko gai izan beharko da (ikus klase-erlazioak 2. taulan).

Deskonposaketaren intuizioa abian jartzeko *clustering hierarkikoa* erabili da. Clustering hierarkikoak aukera ematen du instantzia multzoak —ikasleen erantzunen multzoak— distantziaren arabera antolatatzeko, multzo sakabanatuak banan-banan elkartzuz. Hala, sailkapen-ataza nagusia erabaki-zuhaitz bilaka daiteke ikasleen erantzunen multzoak distantziaren arabera hierarkizatuta. Era honetan, erabaki-zuhaitzak kalifikazio adina hosto izango litzuke, eta, nodo bakoitzean erabakiko genuke kalifikazio jakin bat esleitu ala ez. Horrela, ikaslearen erantzuna zuhaitzean behera barreiatuko litzateke, 1. irudian agertzen den moduan.

Clustering hierarkikoa abian jartzeko EHU-ALM sistemaren emaitzak erabili dira, hain zuzen ere, kontingentzia-matrizea. Kontingentzia-matrizeak sistemak esandakoaren eta benetakoaren arteko erlazioa deskribatzen du (sistemaren hutsegiteak eta asmatzeak deskribatzen dituen matrizea da alegia). Horri esker, ikasleen erantzunak multzo edo cluster moduan ikus daitezke, eta bai multzo horien arteko distantziak kalkulatu ere.



1. irudia. Erabaki-zuhaitzaren irudikapen grafikoa, behin ikasleen erantzunen multzoak hierarkizatuta.



2. irudia. Klase-bikote antzekotasuna oinarritat hartuz clustering hierarkiko aglomeratiboaz lorturiko dendograma. Distantziak Y ardatzean ageri dira eta X ardatzeko balioak honakoak dira: 1 → Erdizka, 2 → Zuzena, 3 → Kontraesana, 4 → Garrantzirik gabe, 5 → Domeinuz kanpo.

Multzoen arteko distantzia kalkulatzeko kontingentzia-matrizearen nahasmena erabili dugu oinarrian, hau da, EHU-ALM sailkatzailea bi multzotan askotan nahastu bada, bi multzoak antzekoak direlako hipotesia. Zenbat eta nahasmen handiagoa, orduan eta multzoak antzekoagoak. Antzekotasun balio horiek erabiliz, clustering hierarkiko aglomeratiboak 2. irudian ikus daitekeen egitura itzultzen digu, intuizio logikoarekin bat datorrena.

Ondorioz, teknika hauek erabiliz hasiera batean 5 klaseko sailkapen-ataza bat zena, orain, 5 ataza bitar bilakatu da.

Emaitzei dagokienez (1.irudian), hainbat saio egin ditugu nodo bakoitzean aritzen den sailkatzailea espezialitzatzeko, baina aitortu beharrekoa da orain arte ez dugula lortu oinarritzko sistemaren emaitzak estatistikoki adierazgarria den modu batean hobetzea.

3. taulan ageri dira oinarritzko sistemaren eta sistema hierarkikoaren emaitzak, SemEval-eko ebaluazio multzoa erabiliz lortutakoak. Ikusten denez, oso hobekuntza xumea da lortutakoa. Argi dago hipotesiarekin lanean jarraitu behar dugula emaitzak hobetu ahal izateko.

3. taula. SemEval 2013ko ebaluazio multzoa erabilia lorturiko F balioak.

Oinarritzko sistema	Sistema hierarkikoa
0.415	0.426

5. ESALDI-SEGMENTUAK LERROKATZEKO SISTEMA

Atal honetan esaldi desberdinen arteko segmentuen lerrokatzeari helduko diogu. Dagoeneko, esana dugu segmentu hauen erabilera eraginkorra dela esaldi-antzekotasun metodo tradizionalen gabezia bati aurre egiteko.

Hau da, 4. atalean ikusi dugun bezala estrategia berri honek aukera ematen du esaldi pareta emanik emaitza zuzenean konputatu beharrean esaldi bikotearen segmentuak elkarrekin parekatu eta esaldien zatikako gainezarpena lantzeko.

Lehen urrats hauek emateko SemEval 2015eko lerrokatze-ataza antolatu, eta, gainera, bertan parte hartu dugu [15].

Hitz gutxitan, atazak ondokoa eskatzen du: esaldi pareta emanik, esaldi hauek segmentutan deskonposatzea, eta segmentu hauei antzekotasun-balioa eta kategoria esleitzea. Hona hemen egin beharrekoaren adibide bat, hau ere jatorrizko datu-basetik euskaratuta:

4. taula. SemEval 2015eko atazatik euskaratutako adibide bat.

Esaldi 1	Esaldi 2
Kirol-auto gorri bat zirkuituan barrena lasterketan.	Kirol-auto gorria pista batean azkar ibiltzen.
Segmentatu eta segmentuak parekatu	
[Kirol-auto gorri bat] ← Baliokidea: 5 → [Kirol-auto gorria]	
[-parekatzerik ez-] ← Parekatzerik gabe: NIL → [azkar]	
[lasterketan] ← Antzekoa: 3 → [ibiltzen]	
[zirkuituan barrena] ← Antzekoa: 3 → [pista batean]	

Ataza honetan parte hartzeko erabilitako sistema hainbat moduluren konbinazioa da. Lehenik eta behin, Sultan et al. autoreek garatutako parekatze-sistemaz baliatzen gara token isolatuen arteko lerrokatzeak identifikatzeko [16]. Informazio hau Agerri et al. autoreek garatutako sintagma-banatzaillearen emaitzarekin konbinatzen dugu [17], era honetan *token*-parekatze isolatuak *chunk*-parekatze bilakatuz. Ur sakonetan murgildu gabe, esan daiteke chunkak eta sintagmak baliokideak direla, eta tokenak eta hitzak ere hala direla. Era honetan, chunk bat hainbat tokenez osatua egongo da.

Bi chunk parekatzean pisu bat esleitzen zaio lotura honi, eta pisu hau bi chunken barnean parekatutako token kopuruaren proportzionala da. Ondorioz, chunk biren artean zenbat eta token gehiago egon, orduan eta pisu handiagoa izango du haien loturak; hain zuzen ere, lotura optimoak aurkitzea da hurrengo zeregina. Lotura optimoak aurkitzeko, gure sistemak informatikaren optimizazio-arloko algoritmo bat erabiltzen du, algoritmo hungariarra deritzona hain zuzen [18]. Algoritmo honek, kostu-matrize bat emanik, lerroetako eta zutabeetako elementu guztiak elkartzen ditu, kostua minimizatuz. Lerroetan eta zutabeetan esaldi parearen chunkak ipintzen baditugu eta kostu gisa chunken arteko loturaren pisuaren alderantzizkoa, loturen pisua maximizatzeke arazoa ebatz daiteke elkar-keta optimoak eskuratuz [19] (kontuan izan pisuen alderantzizkoa minimizatzean arazo duala ebazten ari garela, hau da, loturen pisuak maximizatzen ari gara).

Ondorioz, sistemak esaldi parearen chunkak lerrokatzeke aukera optimo bilatzen du. 5. taulako emaitzak adierazi modura, sistema hau eraginkor izan da SemEval 2015eko lerrokatze-atazan.

5. taula. SemEval 2015eko ebaluazio multzoa erabilia lorturiko emaitzak. Espazio-muga dela-eta, ebaluazio datu-base bakar baten emaitzak aurkezten dira.

Sistema	Alineazioa	Alineazioa eta mota	Alineazioa eta balioa	Alineazioa, mota eta balioa
Baseline	0.8388	0.4328	0.721	0.4326
Azaldutako sistema	0.8846	0.6557	0.8085	0.6159
Beste sistemen arteko maximoa	0.887	0.6143	0.7968	0.5964
Beste sistemen arteko batezbestekoa	0.8193	0.5004	0.7197	0.4748

6. ONDORIOAK

Lan honek helburu modura hartu du irakaskuntzarekin zerikusia duten HPko teknologietan sakontzea. Honetarako, gainera aztertu ditugu erantzunei kalifikazio automatikoak esleitzeko gai diren oinarritzko teknikak eta sistemak: bai esaldi-antzekotasunak erabiltzen dituzten sistemak, baita testu-inferentzian oinarritzen direnak ere.

Honetaz gain, esana dugu zergatik utzi ditugun testu-inferentzia teknikak albo batera, esaldi-antzekotasun teknikak aukeratuta. Izan ere, testu-inferentziak ez bezala, esaldi-antzekotasunak simetrikoa eta graduatua den baliokidetasun bat itzultzen digu, gure helburuetarako egokiagoa.

Ikerketaren muinari dagokionez, alde batetik, ikasleen erantzunak nahasmenaren arabera hierarkikoki multzokatu eta honen arabera kalifikazioak esleitzeko gai den sistema bat aztertu dugu, baina ikusi dugu sistema ez dela gai izan oinarritzko sistemaren emaitzak modu esanguratsuan gaintzeko. Beste alde batetik, esaldiaren ulergarritasuna behatzeko bako segmentuak parekatzeko saioak egiten dituen bigarren sistema bat deskribatu dugu; sistema honek nahiko emaitza onak eskuratu ditu, ataza berritzaile batean. Sistema honen laguntza dugula, espero dugu ikasleen erantzunak kalifikatzeko eraginkortasuna hobetzea, baina baita eman ahal dugun *feedbacka* eraginkorragoa eta ulergarriagoa izatea ere.

Feedback erabilgarria automatikoki sortzea oso baliotsua da testu-ulermena lantzeko aplikazioen jardueretan Hona adibide bat: tutore-sistema adimendunak. Izan ere, ikasleari zuzenketa edo laguntza egokia emateko aukera dago, eta ondorioz, ikaskuntza-prozesua hobetu egiten da.

7. ESKER ONAK

Ikerketa hau unibertsitateko irakasleak eta doktoreak trebatzeko diru-laguntza baten bitartez finantzatu da; Ekonomia, Nekazaritza eta Kul-

tura Ministerioak esleitu (FPU13/00501). IXA ikerkuntza taldeari¹ ere eskertu nahi diogu, taldeak eskaintzen duen master-bukaerako proiektu gisa zehaztu baitziren lan honen hastapenak.

8. BIBLIOGRAFIA

- [1] JURAFSKY, DAN eta JAMES H., MARTIN. 2000. *Speech and Language Processing*. Pearson Education India.
- [2] DAVIES, PHIL. 2002. «There's no confidence in multiple-choice testing». *Proceedings of 6th CAA Conference*. <https://dspace.lboro.ac.uk/2134/1875>.
- [3] CONOLE, GRÁINNE eta WARBURTON, BILL. 2005. «A Review of Computer-assisted Assessment». *Research in Learning Technology*, **13**, 17-31.
- [4] KARPICKE, JEFFREY eta L. ROEDIGER, HENRY. 2008. «The critical importance of retrieval for learning». *Science*, **319**, 966-968.
- [5] ALDABE, ITZIAR; LOPEZ DE LACALLE, MADDALEN; MARITXALAR, MONTSE; MARTINEZ, EDURNE eta URÍA, LARRAITZ. 2006. «Arikiturri: an automatic question generator based on corpora and NLP techniques». In *Intelligent Tutoring Systems, Springer*. 584-594.
- [6] LOPEZ-GAZPIO, IÑIGO eta MARITXALAR, MONTSE. 2013. «Web Application for Reading Practice». *IADAT-e2013: Proceedings of the 6th IADAT International Conference on Education*, ISBN: 978-84-935915-3-3, 74-78.
- [7] LOPEZ-GAZPIO, IÑIGO. 2014. «Bota galdera!». *Elhuyar: zientzia eta teknika*. **307**, 48-50.
- [8] ALDABE, ITZIAR; GONZALEZ-DIOS, ITZIAR; LOPEZ-GAZPIO, IÑIGO; MADRAZO, ION eta MARITXALAR, MONTSE. 2013. «Two Approaches to Generate Questions in Basque». *Procesamiento del lenguaje natural*, **51**, 101-108.
- [9] LOPEZ-GAZPIO, IÑIGO eta MARITXALAR, MONTSE. 2013. «Seneko: galderak automatikoki sortuz testuak lantzeko aukera ematen duen aplikazioa». *Euskal Herriko Unibertsitatea (UPV/EHU), ADDI dokumentuen bilduma*, <http://hdl.handle.net/10810/10186>.
- [10] MADRAZO AZPIAZU, JON eta MARITXALAR, MONTSE. 2013. «Hizkuntzaren prozesamendurako teknikak irakaskuntza arloan: galdera sortzaile automatikoa». *Euskal Herriko Unibertsitatea (UPV/EHU), ADDI dokumentuen bilduma*, <https://addi.ehu.es/handle/10810/10702>.
- [11] OMER, LEVY; ZESCH, TORSTEN; DAGAN, IDO eta GUREVYCH, IRYNA. 2013. «Recognizing partial textual entailment». *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, **1**, 451-455.

¹ <https://ixa.si.ehu.es/>

- [12] MOHLER, MICHAEL; RAZVAN, BUNESCU eta MIHALCEA, RADA. 2011. «Learning to grade short answer questions using semantic similarity measures and dependency graph alignments». In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics*, **1**, 752-762.
- [13] DZIKOVSKA, MYROSLAVA; RODNEY D. NIELSEN; BREW CHRIS; LEACOCK, CLAUDIA; GIAM, DANIL; BENTIVOGLI, LUISA; CLARK, PETER; DAGAN, IDO eta T. DANG, HOA. 2013. «Semeval-2013 task 7: The joint student response analysis and 8th recognizing textual entailment challenge». *Second Joint Conference on Lexical and Computational Semantics (*SEM)*, **2**, 263-274.
- [14] ALDABE, ITZIAR; MARITXALAR, MONTSE eta LOPEZ DE LACALLE, OIER. 2013. «EHU-ALM: Similarity-feature based approach for student response analysis». *Second Joint Conference on Lexical and Computational Semantics (*SEM)*, **2**, 580-584.
- [15] AGIRRE, ENEKO; BANEJA, CARMEN; CARDIE, C., CER DANIEL; DIAB, MONA; GONZALEZ-AGIRRE, AITOR; GUO, WEYWEY; LOPEZ-GAZPIO, IÑIGO; MARITXALAR, MONTSE; MIHALCEA, RADA; RIGAU, GERMAN; URIA, LARRAITZ eta J., WIEBE. 2015. «SemEval-2015 task 2: Semantic textual similarity, English, Spanish and pilot on interpretability». In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*.
- [16] MD ARAFAT, SULTAN; BETHARD, STEVEN eta SUMNER, TAMARA. 2014. «Back to basics for monolingual alignment: Exploiting word similarity and contextual evidence». *Transactions of the Association for Computational Linguistics*, **2**, 219-230.
- [17] AGERRI, RODRIGO; BERMUDEZ, JOSU ETA RIGAU, GERMAN. 2014. «Ixa pipeline: Efficient and ready to use multilingual NLP tool». In *Proceedings of the 9th Language Resources and Evaluation Conference*, **1**, 26-31.
- [18] W. KUHN, HAROLD. 1955. «The Hungarian Method for the assignment problem». *Naval Research Logistics Quarterly*, **2**, 83-97.
- [19] AGIRRE, ENEKO; GONZALEZ-AGIRRE, AITOR; LOPEZ-GAZPIO, IÑIGO; MARITXALAR, MONTSE; RIGAU, GERMAN eta URIA, LARRAITZ. 2015 «UBC: Cubes for English Semantic Textual Similarity and Supervised Approaches for Interpretable STS». In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*.