



IKER
GAZTE
NAZIOARTEKO
IKERKETA EUSKARAZ

I. IKERGAZTE

NAZIOARTEKO IKERKETA EUSKARAZ

2015eko maiatzaren 13, 14 eta 15
Durango, Euskal Herria

ANTOLATZAILEA:
Udako Euskal Unibertsitatea (UEU)

OSASUNA

**Osasun-txosten elebidunak
posible ote?**

*O. Perez de Viñaspre, M. Oronoz
eta J. Patrick*

730-738 or.
<https://dx.doi.org/10.26876/ikergazte.i.101>

ANTOLATZAILEA:



BABESLEAK:



LAGUNTZAILEAK:



Osasun-txosten elebidunak posible ote?

Perez-de-Viñaspre O. eta Oronoz M.¹

Patrick J.^{2,3}

¹ IXA taldea (UPV/EHU). Manuel Lardizabal Pasealekua 1. 20018 Donostia

² Innovative Clinical Information Management Systems (iCIMS). Sydney

³ Health Language Analytics (HLA). Sydney

Laburpena

Artikulu honetan osasun-txosten elebidunak lortzea posible dela erakusten dugu, horretarako, tresna-prototipo bat garatuaz. Prototipoa lortzeko zaintza klinikorako informazioa kudeatzeko sistema bat sortu dugu horretarako prestatuta dagoen software bati esker. Sistema honen oinarrian dauden terminologia-zerbitzariak sortu ditugu euskara eta gaztelaniarako, eta medikuntzara egokitutako ortografia-zuzentzaile bat ere sortu dugu euskararako. Prototipoa eraikitzeo, Donostia Unibertsitate Ospitaleko alta-txostenen bildumako “Digestio-aparatua”ko ereduak erabili dugu. Horrela, euskaraz alta-txostenak idazteko eta modu errazean haren terminologiaren gaztelarako baliokideak lortzeko prototipo bat eraiki dugu.

Hitz gakoak: Hizkuntzaren prozesamendua, termino kliniko, osasun-txosten elebidun, SNOMED CT

Abstract

This paper shows that getting bilingual health-records is possible by means of developing a prototype. We developed the prototype by creating a Clinical Care Information System using a software designed for that purpose. We created the terminology servers for Basque and Spanish that are the base of the system, and we adapted a Basque spell-checker for medicine. We used the model of the “Digestive System” specialization from the Donostia University Hospital discharge summary collection to build the prototype. Thus, we created a prototype to write discharge summaries in Basque and to easily obtain the Spanish translation of its terminological content.

Keywords: Natural Language Processing, clinical term, bilingual health-record, SNOMED CT

1 Sarrera eta motibazioa

Osakidetza, Euskal Autonomia Elkargoan (EAEn) osasun-zerbitzuak eskaintzeaz arduratzen den erakundea da. EAEn euskara eta gaztelania hizkuntzak dira koofizialak baina maiz osasun-langileekin tratua euskaraz izaten dugun arren, haiek osasun-txostenak gaztelaniaz idazten dituzte. Batzuetan arrazoi desberdinak medio (hezkuntzarako erabilitako hizkuntza, osasunarekin lotutako terminoen ezagutza eza ...) erosoago sentitzen direlako egiten dute hori; besteetan berriz, euskaraz osasunaren inguruan idazteko gai izan arren, beraien burua erdi behartuta sentitzen dute gaztelaniaz idaztera: Osakidetzak sistema zentralizatua du eta osasun-langile batek idazten duen txostena, gaixoa tratatzen duten gainerako osasun-langileek ere irakur dezakete. Euskaraz idatzi duenaren ondoren datozen irakurleak euskara ulertzeko gai ez badira, gaixoaren segurtasuna kolokan egon daiteke.

Osakidetzak bere 2013-2019 eperako definitu duen *Osakidetzan Euskararen Erabilera Normalizatzeko 2. Plan*an (Osakidetza, 2013) zera zehazten du: “ahalegin handiak egin behar dira historia klinikoak ere ele bietan idatzita egoteko; eta horretarako, nahitaezkoa da luze gabe azterketa sakon bat egitea, informazioa sortu eta ustatzeko prozesuari eragiten dioten faktoreak ondo ezagutze aldera”. 2014ko abenduan, aditu-talde batek hainbat hilaletan alderdi teknikoak, etikoak, gaixoaren segurtasuna eta

antzeko faktoreak aztertu ondoren, txosten bat aurkeztu zuen gai honen inguruan. Txosten horren alderdi teknikoaren atalean, artikulu honetan aurkezten dugun tresna azaltzen da, osasun-txosten elebidunak erdiautomatikoki sortzea posiblea dela erakusteko. Beste aukera, osasun-txostenak giza itzultzaileek itzultzea, oso garestia da.

Beraz, aipatu berri dugun moduan, artikuluan aurkeztuko dugun tresnaren helburua osasun-txosten elebidunak sortzeko bidean lehen urratsak ematea da, hizkuntzaren prozesamenduko teknikak erabiliz. Hizkuntzaren prozesamendua, informatika eta hizkuntzalaritza konbinatzen dituen alorra da eta hizkuntza aztertzeko edo sortzeko tresna konputazionalak ikertzeaz arduratzen da. Alor honetan garatzen diren tresna ezagunenak ortografia-zuzentzaileak¹, itzultzaile automatikoak² edota informazioaren berreskurapenerako tresnak³ (Google edo Yahoo bilatzaileak, besteak beste) badira ere, hamaika ikerlerro daude zabalik, eta tresna berriak etengabe sortzen eta hobetzen ari dira.

Tresna garatzen hasi aurretik, egunerokoan osasun-txostenak eskuartean dituzten osasun-langileen behar errealeak zeintzuk diren jakin nahi izan genuen inkesta bat erabiliz. Horretarako, 2014 urteko otsailean osasun alorreko 45 langileri galdeketa bat pasa genien honakoen inguruan: txostenak edozein hizkuntzatan idazten dituztenean zein zailtasun topatzen dituzten ezagutu nahi genuen eta txostenak euskaraz idazteko aukera izanez gero, tresnak zein behar ase beharko lituzkeen (ortografia-zuzentzailea, osasun-arloko terminologia euskaraz, txantiloak e.a.) zerrendatu nahi genuen.

Galdeketa emaitzetatik ateratako ondorioak zerrendatuko ditugu jarraian. Hizkuntzaren erabilerrari dagokionez, 41 osasun-langilek gaixoei komunikazioa euskaraz egiten dute baina 5ek soilik idazten dituzte txostenak euskaraz. Galdetegian osasun-txosten bat euskaraz idaztera gonbidatu genituen inkestatuak eta honako zailtasunak aurkitu zituzten: euskarazko terminologiaren falta (45etik 32 pertsonak), eredu falta (24 pertsonak) eta ohitura falta (15 pertsonak). Euskaraz idazteko beharrezkoak iruditzen zaizkien baliabideen artean galdetuta, lehenetsun handiena eman zieten euskarazko medikuntzaren inguruko terminologia bateratu eta osatu bati (45etik 39 pertsonak), medikuntzara egokitutako ortografia-zuzentzaile bati (30 pertsonak) eta idazketarako eredu edo txantiloak izateari (25 pertsonak). Galdetutakoen % 89 prest agertu zen tresna sortuko balitz erabiltzeko, gainerakoek berriz, bere zalantzak agertu zituzten.

Beharrak eta lehenetsunak zehaztuta, garapenari ekin genion, nahiz eta badakigun haren erabilera normalizatua ez dagoela gure esku. Horixe da, hain zuzen, aztertu nahi izan genuen hurrengo kontua: beste herrialde elebidunetan (edo eleanitzetan), osasun-zerbitzuetan hizkuntzen erabilera nola egiten den. Kanadan, elebitasuna ofiziala den zonaldeetan, gaixoen eta osasun-langileen arteko harremanerako hizkuntza gaixok erabakitzen du (Desjardins, 2003) eta osasun-txostenak hizkuntza horretan idazten dira. Belgikan aldiz, hizkuntza-komunitate bakoitzak bere osasun-zerbitzua kudeatzen duen sistema ez-zentralizatua dauka, eta Bruselaren kasuan, zonalde elebiduna izanik, bi osasun-zerbitzu sistema eskaintzen dira: frantziar komunitatearena eta flandiarrena (Gerkens eta Merkur, 2010). Luxenburgon, herrialde txiki bezain anitza izanik, nahiz eta alemana, frantsesa, italiera, ingelesa eta portugesa hizkuntza zabalduak diren, *lingua franca* moduan frantsesa erabiltzen dute (on Health Care Systems, 1999).

Gurean aldiz, hizkuntza komunitateak nahastuta daude, eta osasun-zerbitzu bateratua daukagu; *lingua franca* gisa gaztelania erabiltzen dugu Hego Euskal Herrian, baina horrek ez ditu herritar guztien hizkuntza-eskubideak bermatzen. Euskara hizkuntza ofiziala izan arren, osasun-langile guztiek ez dute euskara ezagutzen eta euskaldunak diren osasun-langileek, ahozko komunikazioa euskaraz egiten dute eta berehalako itzulpena egin beharra dute oharrak gaztelaniaz idazteko. Edozein kasutan, gaixok oraingoz ez du sistematikoki bere osasun-txostena euskaraz lortzeko aukerarik. Horretarako bidea ireki nahi du tresna-prototipo honek, euskarazko idazketan laguntzeaz gain, osasun-txosten elebidunak automatikoki lortze-bidean itzulpen automatikoaren oinarriak ezartzen baititu.

Osasun-txostenetan erabiltzen den hizkuntzaren gramatikari erreparatuta, egitura sinpleak erabiltzen direla konturatu gara; esaldi konplexuen erabilera oso murrizta da. Gainera, erabiltzen den hiztegia terminologia espezializatuarekin egoten da hertsiki lotuta, eta horrela anbiguotasuna asko gutxitzen da. Hori guztia kontutan harturik, osasun-txostenetan idatziko den testua inplizituki bada ere, "hizkuntza kontrolatu" gisa definitu ahalko genuke, erabiltzen diren egiturak eta esaldiak gehienetan sinpleak eta

¹<http://eu.wikipedia.org/wiki/Ortografia-zuzentzaile>

²http://eu.wikipedia.org/wiki/Itzulpengintza_automatiko

³http://eu.wikipedia.org/wiki/Informazioaren_berreskurapen

haien artean antzekoak baitira. Ezaugarri horiek ezinbestekoak dira gaur egun itzulpen automatikoa berme batzuekin bideragarria izan dadin (Mitamura, 1999; Aikawa *et al.*, 2007).

2 Arloko egoera eta ikerketaren helburuak

Osasun-txostenen prozesamendu automatikoan aurrera urrats oso garrantzitsuak eman dira, adibidez osasun-txostenen sailkapen automatikoa (Pérez *et al.*, 2015) edota txosten horietan ebakuntza osteko konplikazioen identifikazio automatikoa (Murff *et al.*, 2011). Dena dela, osasun-txosten horien itzulpen automatikoa landu gabeko arloa da orain arteko literaturan. Aztertu ahal izan dugunaren arabera, osasun-txosten elebidunak kasu puntualetan eskuzko itzulpenekin lortu dira orain arte, bai munduan (Kanadaren kasua, aurreko atalean ikusi dugun moduan), bai eta Euskal Herrian ere (gaixoak txostena euskaraz eskatzen duen kasuetan).

Jarraian, artikulua honetan aurkeztuko dugun tresnaren aurrekari izan diren hainbat baliabide deskribatuko ditugu.

Osasun-txostenen prozesamenduari dagokionez, bi tresna aztertu ditugu bereziki: *Clinical Care Information System*⁴ tresna *Innovative Clinical Information Management Systems (iCIMS)*-ek garatutakoa eta *CliniTermServer*⁵ tresna *Health Language Analytics*-ek garatua.

Clinical Care Information System edo CCIS, zaintza klinikorako informazioa kudeatzeko sistemari deritzo iCIMS enpresak. Sistema honek gaixoaren informazioa kudeatzeko aukera zabaltzeaz gain, osasun-arloko langileek euren azterketa eta prozeduren ondorioz idatziz jaso nahi duten informazioa modu antolatu eta egituratuan idazteko aukera ematen du. Gainera, gaixoaren historia klinikoa modu bateratuan gordetzeko aukera ematen du eta modu honetan kudeaketa errazten du. iCIMSek CCIS-ak diseinatzeko softwarea eskaintzen du *formBuilder* txantiloitzaileari esker, eta modu honetan CCIS berri bat sor daiteke ospitale edo zerbitzu ezberdinen beharretara erabat egokiturik. Software hori oso malgu da aldaketak eta eguneraketak egiteko garaian.

CliniTermServer terminologia-zerbitzari⁶ bat da, eta osasun-arloko terminologia eta beste hainbat sasitermino edo zerrenda barnebiltzen ditu. Besteak beste, osasun-txostenetan maiz erabiltzen diren esamoldeak eta izen-sintagmak jasotzen dira, baita salbuespeneko jokabidea dutenak ere. Zerbitzari honek terminologia jasotzeaz gain, berau prozesatzeko softwarea ere eskaintzen du, esaterako, gordeta duen terminologia testuan identifikatzeko gai da, eta terminoarekin loturiko informazioa eskaintzen du.

Euskarazko baliabideak ugariak ez badira ere, baditugu pare bat baliabide bereziki esanguratsuak eskuartearen dugun ataza aurrera eramateko: Donostia Unibertsitate Ospitaleko alta-txostenen bilduma (Joanes Etxeberri Saria V. Edizioa, 2014) eta SNOMED CT (*Systematized Nomenclature of Medicine – Clinical Terms*) (IHTSDO, 2014) sare semantiko eleanitza.

Aipaturiko alta-txostenen bilduman, Donostia Unibertsitate Ospitaleko espezialitate ezberdinetako alta-txostenen ereduak agertzen dira, euskaraz idatzitako alta-txostenen adibideak, hain zuzen ere. Hogeita hamar espezialitate eta berrogeita hiru alta-txosten barnebildurik daude bilduma horretan eta benetan erabiltzen diren egiturak jasotzen dira bertan. Gainera, esfortzu berezia egin dute egileek bertan ageri den euskarazko terminologia zuzena izan dadin.

SNOMED CT, medikuntzaren alorreko ontologia zabala da. Osasun-txostenetan agertzen diren kontzeptuak, deskribapenak eta erlazioak barnebiltzen dituen ontologia da. Medikuntzaren arloan orain artean egin den terminologia eleanitz ulergarrientzat hartzen da. Nolabait esateko, hizkuntza eta sistema desberdinen arteko osasun-txostenen adierazpen eta interpretazio automatikoa eta anbiguotasunik gabea ahalbidetuko duen hiztegi normalizatua da, eta hiztegi-sarreraren arteko harremanak zehaztuta izango ditu. Bere jatorrizko hizkuntza ingelesa da, eta honentzako 300.000 kontzeptu baino gehiago definituak ditu, baita horiei dagozkien 1.000.000 termino edo deskribapen baino gehiago ere. SNOMED CTk gaur egunean erabilera zabala du mundu osoan zehar. Hainbat hizkuntzatarako bertsioak daude eskura edota garatze-lanetan, hala nola, nederlanderako, frantseserako, gaztelaniarako, ... Euskararen bertsioa ere garatze-bidean dago (Perez-de-Viñaspre eta Oronoz, 2014) eta berrikusi gabeko 68.495 terminorekin

⁴<http://www.icims.com.au/solutions/products/>

⁵<http://www.healthll.com.au/product-suite/clinitermsrver/>

⁶Zerbitzaria informatikan zera da: Bezero deitzen diren programen eskaerei erantzuten dien programa edo aplikazio informatikoa.[Wikipedia] Zerbitzu bat eskaintzen duen aplikazioa, azken finean.

lanerako bertsio bat eskuragarri dago egileei eskatuta.

Ikerketa-lan honen helburu nagusia, lau baliabide horiek uztartuz, osasun-txostenak euskaraz idazteko tresna baten prototipoa garatzea da. Tresnaren garapenaz, modu erdiautomatiko batean osasun-txosten elebidunak euskaraz eta gaztelaniaz idaztea posiblea dela frogatu nahi izan dugu, eta etorkizunerako ikerketa bide bat ireki dugu.

3 Baliabideak eta metodoak

Aurreko atalean azaldu dugun moduan, ondorengoak dira lan hau aurrera eramateko uztartu ditugun baliabide nagusiak:

- *Clinical Care Information System* edo CCIS
- *CliniTermServer*
- SNOMED CT eleanitza
- Euskarazko *Donostia Unibertsitate Ospitaleko Alta Txostenak* lana

Aurretik esan bezala, CCIS sistema bat sortzeko iCIMS enpresaren *formBuilder* txantiloiz-sortzailea erabiltzen da, eta horixe bera erabili dugu artikulu honetan aurkeztu dugun prototipoaren (CCIS sistemaren) hezurdura diseinatzeko. Diseinu honetan, lehen txantiloia osatzeko Donostia Unibertsitateko Ospitaleko “digestio-aparatua” espezialitaterako alta-txosten eredua erabili dugu.

Alta-txostena idazteko sistemak duen txantiloia zati bat ikus dezakegu 1 Irudian. Bertan ikus daitekeen bezala, atal eta azpialal ezberdinak definituta daude (“Ospitaleratzeko arrazoa”, “Aurrekariak” e.a.) horiek guztiak alta-txostenen bildumatik jasotako erduei jarraituz. Atal batzuk hautazkoak izanik (ikus “Proba osagarrien laburpena” atala), horiek klik erraz batean agertu edo desagertuko dira, alta-txostena betetzen ari den osasun-langileak beharretara egokitutako alta-txostena idatz dezan.

Gainera, osasun-langileen gomendioei jarraituz, atal batzuei eduki lehenetsia esleitu diegu. Eduki lehenetsi honek gaixoaren egoera “normala” adierazten du. Adibidez, “Aurrekariak” atalean, egoera normala “Ez du alergia-rik. Ez du ohiko tratamendurik.” izango litzateke, hau da, alteraziorik gabeko egoera. Modu honetan, osasun-langileak egoera normaletik at dagoen osasun-egoera topatzen duenean bakarrik aldatu beharko du eremuaren edukia, idazketa-lana sinplifikatuz. Eskakizun hori, langileen denbora-kudeaketari erantzuten dion behar bat da; izan ere, gaixoari eskaini ahal dioten denbora laburra da, eta denbora-galeratzat hartzen dute osasun-txostenen atal guztiak bete behar izatea, egoeraren alteraziorik ez baldin badago, behintzat.

Hiztegi elebidunean termino baten bilaketa egiteko eremua definitu dugu (ikus 1 Irudian “Hiztegi-kontsulta” aukera). Dena dela, funtzionalitate hori ez da oraindik CCIS tresnan eskaintzen

Prototipoaren azalpenekin jarraitu ahal izateko, aurrenik *CliniTermServer* terminologia-zerbitzariaren gainean egindako ekarpenak azalduko ditugu. Ikusi dugun bezala, CCIS sistemako txantiloian testu-kutxak daude. Testu-kutxa horietan automatikoki ezagutzen dira osasun-terminoak *CliniTermServer* terminologia-zerbitzariari esker.

CliniTermServer terminologia-zerbitzaria ingelesezko bakarrik eskaintzen du *Health Language Analytics* enpresak, eta guk berau euskararako eta gaztelaniarako egokitu dugu. Terminologia zerbitzari hori beste hizkuntzetara egokitzeko, hizkuntzaren menpekoak diren prozesamendurako tresnak behar izan ditugu, tokenizatzailea eta lematizatzailea, hain zuzen ere.

Tokenizatzailea testuko hitzak eta puntuazio-markak bereizi eta banatzeko tresna da, hau da, testua tokenetan banatzen duen tresna. Ikus dezagun (1) adibidea (azpian) zeinetan jatorrizko testua tokenizatzean puntuazio-markak hitzetatik bereizi egiten dira, eta horrela, esaldiko elementu bakoitza banatu egiten den.

- (1) **Jatorrizko esaldia:** “Disuriagatik etorri da 28 urteko emakumea.”
Tokenizatua: “Disuriagatik etorri da 28 urteko emakumea .”
Lematizatua: “disuria etorri izan 28 urte emakume .”

Lematizatzaileak aldiz, hitzen oinarritzko forma ematen du, hau da, hiztegi-tako sarrera izan daitekeen hitzaren forma. Aurreko adibideari jarraiki (1. adibidea), hitzen deklinabidea galdu egiten dela ikus de-

1 Irudia: “Digestio-aparatua” espezialitaterako garaturiko CCIS sistema

Abizenak: Ten	
ALTAREN TXOSTEN KLINIKOA	
DIGESTIO-APARATUA	
OSPITALERATZEKO ARRAZOIA Anemia duen 43 urteko gizona.	Hiztegi-kontsulta Sarrera:
AURREKARIAK Ez du alergiairik. Ez du ohiko tratamendurik.	
EGUNGO HISTORIA Duela 3 hilabetetik ahultasuna sentitzen du. Egun batzuetatik hona gorotzak beltzak direla iruditu zaio. Oheburuko medikuari kontsultatu ostean, odol-analisia egin diote eta anemia duela ikusi da. Ez du sabeleko minik sentitzen, eta ez du izan eginkarien erritmo-alaketarik.	
AZTERKETA FISIKOA Tentsio arteriala (TA), 130/80 mmHg; bihotz-maiztasuna (BM), 70 tau/min; tenperatura, 36,7C. Kontziente eta orientatua dago. Hidratazioa eta perfusioa egokiak dira. Eupneko. Larruazal eta mukosetako kolore normala.	
Burua eta lepoa. Ez dago aurkikuntza patologikorik Biriken auskultazioa. Biriketako murmurio normala. Sabela. Biguna eta zanpagarria. Haztapenean ez du minik sentitzen, eta ez dago masa edo megallarik. Heste-soinu normalak. Gorputz-adarrak. Ez du edemarik. Pulsu periferiko normalak. Ondeste-ukipena. Ez da masarik ukitzen. Eskularruaren hatza melenaz zikindu da.	
PROBA OSAGARRIEN LABURPENA ☐ ANALITIKA. ☒ ELEKTROKARDIOGRAMA (EKG). Erritmo sinusal normala. ☐ BULARREKO ERRADIOGRAFIA.	

zakegu (“Disuriagatik” hitzaren “disuria” lema, adibidez), eta aditzen kasuan euren erroa berreskuratzen da (“da” aditzaren erroa “izan” litzateke).

CliniTermServer egokitzeko hizkuntza bakoitzerako tokenizatzailea eta lematizatzailea txertatu behar izan ditugu: gaztelaniaren kasuan hizkuntzaren prozesamendurako Freeling kode irekiko liburutegiaren 3.1 (Padró eta Stanilovsky, 2012) bertsioa erabili dugu eta euskararako IXA taldeak garaturiko Eustagger etiketatzaile/lematizatzailea (Ezeiza *et al.*, 1998). Informatikaren alorrean liburutegi bat programen kodea idazteko eskuragarri dauden prozedura edo programen multzoa da, eta hori kode irekikoa izanik, libreki erabiltzeko edota aldatzeko aukera ematen du, kostu ekonomikorik gabe, betiere jatorriari erreferentzia eginik. Hau da Freeling kasua.

Hizkuntza prozesatzeko tresnak erabiltzeaz gain, eduki terminologikoz elikatu behar izan ditugu *CliniTermServer*eko gaztelaniako eta euskarazko zerbitzariak. Gaztelaniarako sortutako zerbitzariarentzat, SNOMED CTren nazioarteko banaketaren gaztelaniazko bertsioa jaso dugu, 2014ko apirilaren 30ko banaketa⁷, hain zuzen ere. Euskarazko zerbitzarirako aldiz, 2. atalean aipatutako SNOMED CTren garatze

⁷<http://www.nlm.nih.gov/research/umls/licensedcontent/snomedctfiles.html>

fasean dagoen euskarazko bertsioa erabili dugu (Perez-de-Viñaspre eta Oronoz, 2014).

Prestatu ditugun zerbitzari terminologikoei, bi funtzionalitate nagusi eskaintzen dizkigute: i) testuan agertzen diren SNOMED CTren termino edo deskribapenak identifikatu eta ii) SNOMED CTren kontzeptu baten identifikadorea jaso eta honen adierak eman. Aipatu moduan, hiru zerbitzari ezberdin prestatu ditugu: ingeleserako (HLA enpresak eskainia), euskararako eta gaztelaniarako (guk egokituak). Lan honetarako, hizkuntza ezberdinetako zerbitzariak konbinatu ditugu, eta euren artean komunikatzeko SNOMED CTren kontzeptu-identifikadorea erabili dugu, kontzeptu bat identifikatu, eta beste hizkuntza batean ordainak emateko. Honen adibideak 2 Irudian ikus ditzakegu, euskarazko zerbitzariak testuan dauden terminoak identifikatu, eta gaztelaniazko zerbitzariak horien gaztelaniazko baliokideak ematen dituenak: “alergiarik” deskribapenerako “*disposición alérgica*” edota “tentsio arteriala” terminorako “*presión sanguínea*”.

2 Irudia: “Digestio-aparatua” espezialitaterako garaturiko CCIS sistema

OSPITALERATZEKO ARRAZOIA		Hiztegi-kontsulta
Anemia duen 43 urteko gizona.		Sarrera
☐ 271737000 anemia Ez du alergiarik. Ez du ohiko tratamendurik.		
☐ 609328004 disposición alérgica ☐ 415178003 proceso ☐ 276239002 tratamiento ☐ 119270007 tratamiento		atik hona gorotzak beltzak direla iruditu zaio. Oheburuko anemia duela ikusi da. Ez du sabeleko minik sentitzen, eta ez
AZTERKETA FISIKOA		
Tentsio arteriala (TA), 130/80 mmHg; bihotz-maiztasuna (BM), 70 tau/min; tenperatura, 36,7C. Kontziente eta orientatua dago. Hidratazioa eta perfusioa egokiak dira. Eupneiko. Larruazal eta mukosetako kolore normala.		
☐ 75367002 presión sanguínea		

Itzulpen prozesua nola egiten den ikusteko, azter dezagun “Ez du alergiarik” esaldia 2 Irudian. Euskarazko esaldia tokenizatu ostean, esaldiaren tokenak euren lemarekin (“Ez izan alergia”) ordezkatu dira. Ondoren, lema guztiak SNOMED CTren kontra parekatzen ditu zerbitzariak. Kasu honetan, “alergia” terminoak bakarrik jaso du SNOMED CTren kontzeptu-identifikadorea (“609328004”). Kontzeptu-identifikadore hori gaztelaniazko zerbitzariari pasatakoan, “*disposición alérgica*” terminoa itzuli digu, hau da, gaztelaniazko SNOMED CTn kontzeptu horri loturik dagoen hobetsitako terminoa.

Terminologia-zerbitzarien konbinazio honi esker, osasun-langileek esanguratsuak diren terminoak aukeratu ahal izango dituzte, eta kontzeptu bati dagozkion ordain guztien artean egokia zein den aukeratzeko, hau da, desanbiguatzeko aukera izango dute, itzulpen automatikoan akats gehien sortzen duen ataza sinplifikatuz.

Lan honen helburuari jarraiki (osasun-txostenak euskaraz idazteko tresna baten prototipoa garatzea), euskarazko idazketan lagungarria izango den baliabide bat ere garatu dugu: medikuntzarako ortografia-zuzentzaile.

Ortografia-zuzentzaileek maiz eraginkortasuna galdu egiten dute hizkuntza oso teknikoa erabiltzerakoan. Izan ere, terminologia espezializatua ez da hiztegi orokorretan agertzen, eta zuzentzaileak hiztegi horietatik elikatzen dira. Hortaz, testu oso espezializatua idaztean, testuaren zati handi bat gorritz azpimarratuta agertzen zaigu, akatsak balira bezala, nahiz eta termino horiek domeinuan zuzenak izan. Arazo honi aurre egiteko, Xuxen (Agirre *et al.*, 1992) zuzentzailea oinarri hartu dugu eta berau osasunaren domeinuko hiztegi espezializatueta hitz-sarrerekin elikatu dugu. Prozesu honetan, 32.728 lema gehitu ditugu Xuxenen hiztegian eta XuxenMed izena eman diogu egokitzapen honetatik sortutako zuzentzaileari.

Xuxen zuzentzaile orokorrak alta-txosten batean egiten dituen zuzenketak ikus ditzakegu 3 Irudian, eta 4 Irudian aldiz medikuntzara egokitutakoak egiten dituen zuzenketak. Ikus daitekeenez, Xuxen orokorrak (3 Irudia) medikuntzari loturiko hitzak okertzat hartzen ditu (“sinusal”, “gastroskopia”, “behazunbidea” eta “Adenokartzinoma”), eta XuxenMedek (4 Irudia), aldiz, zuzentzat ditu.

3 Irudia: Xuxen ortografia-zuzentzaileak zuzendutako testua

☒ ELEKTROKARDIOGRAMA (EKG).	Erritmo <u>sinusal</u> normala.
☒ GASTROSKOPIA.	Anestesia topikoa jarri ostean, <u>gastroskopia</u> ahotik barrena sartuta: hestegorriak kalibre eta mukosa normala du; urdailaren <u>zabalgarritasuna</u> eta uzurtasuna normalak dira; ez da odol-aztarnarik ikusten, ez eta lesiorik ere; piloro normala. Duodeno-
☒ ABDOMENEO OTA.	Gibela, barea, pankrea, behazuna, <u>behazunbidea</u> , giltzurrunen-gaineko guruina eta giltzurrunak normalak. Itsuaren paretaren gizontze bat ikusten da, neoplasia itxurakoa. Ez da adenopatiarik ikusten.
☒ ANATOMIA PATOLOGIKOA.	<u>Adenokartzinoma</u> .

4 Irudia: XuxenMed medikuntzako ortografia-zuzentzaileak zuzendutako testua

☒ ELEKTROKARDIOGRAMA (EKG).	Erritmo sinusal normala.
☒ GASTROSKOPIA.	Anestesia topikoa jarri ostean, gastroskopia ahotik barrena sartuta: hestegorriak kalibre eta mukosa normala du; urdailaren <u>zabalgarritasuna</u> eta uzurtasuna normalak dira; ez da odol-aztarnarik ikusten, ez eta lesiorik ere; piloro normala. Duodeno-
☒ ABDOMENEO OTA.	Gibela, barea, pankrea, behazuna, behazunbidea, giltzurrunen-gaineko guruina eta giltzurrunak normalak. Itsuaren paretaren gizontze bat ikusten da, neoplasia itxurakoa. Ez da adenopatiarik ikusten.
☒ ANATOMIA PATOLOGIKOA.	Adenokartzinoma.

XuxenMed Firefox nabigatzaileko gehigarri moduan prestatu dugu, eta hobetze prozesuan egonik, oraindik publikoki zabalduta ez badugu ere, gurekin harremanetan ipinita banatzeko prestutasun osoa dugu.

4 Ondorioak eta etorkizunerako bidea

Osasun-txostenak euskaraz idazteko tresna baten lehen prototipoa aurkeztu dugu artikulu honetan. Horretarako hainbat tresna egokitu edota sortu ditugu: euskararako eta gaztelaniarako terminologia-zerbitzariak, “Digestio-aparatua” espezialiterako alta-txostenen idazketarako interfazea eta medikuntzara egokitutako ortografia-zuzentzaile bat, besteak beste. Lan honek ikerketa bide berri bat ireki digu, eta etorkizunean osasun-txostenetan aldibereko itzulpen automatikoa landu ahal izango dugu.

Izan ere, prestatutako prototipoaren terminoen gainean desanbiguazioa egitea bermatzen dugu eta atal bakoitzari gaixoaren egoera normala deskribatzen duen edukia aldeztu aurretik prestatu dugu. Hori horrela izanik, hizkuntza (sasi-)kontrolatua eta terminologia desanbiguatua duten osasun-txostenenak idazteko oinarria ezarri dugu.

Garatu gabe gelditu zaizkigun funtzionalitateei aurre egin beharko diegu aurrenik, hiztegiaren garapena eta Donostia Ospitaleko Alta Txostenen bilduma oinarri harturik espezialitate guztietara prototipoa zabaltzea, hain zuzen ere. Espezialitate bakoitzerako atalak eta azpiatalak zehaztu beharko dira, eta txostenetan atal bakoitza derrigorrezkoa edo hautazkoa den erabaki beharko dute adituek. Atalak finkatu ostean, bakoitzarentzako lehenetsitako testua zehaztea izango da adituen esku utziko dugun hurrengo ataza. Modu honetan, CCIS sistema Donostia Ospitaleko espezialitate guztietarako zabalduta ahalko dugu, eta haietan guztietan alta-txostenak idazteko prest izango dugu aplikazioa.

Aipatu dugun bezala, hiztegiaren funtzionalitatea ez dago iCIMS-ek eskaintzen duen softwarean aurreikusia eta honek softwarearen inguruan gogoeta egitera eramaten gaitu. Izan ere, iCIMS-eko *form-Builderek* prototipoa sortzeko abantaila asko eskaintzen badizkigu ere, eramangarritasuna, aldakortasuna

eta berehalakotasuna besteak beste, ez da guk garatu nahi dugun proiekturako diseinatua izan. Hau da, iCIMS-eko softwaretik ideiak jaso eta sistema bera osatu nahi dugu, itzulpen automatikoa ahalbidetzeko, hiztegi dinamiko baten aurka kontsultak bermatzeko, terminologia berria identifikatu ahal izateko. . .

Lan honetarako euskarazko terminologia-zerbitzariaren termino-identifikatzailea sortzeko, ingeleserako garatutako teknologia berrerabili dugu. Nahiz eta euskararako tokenizatzaile zein lematizatzaileak erabili, ez dugu metodoa ukitu. Kontuan izaten badugu ingelesaren eta euskararen gramatikak oso ezberdinak direla, honek terminoen identifikazioan eragina izan beharra dauka derrigorrez. Adibidez, ingelesean preposizioek ematen duten informazioa (“of”, “with”, “s”,...) euskaraz hitzen deklinabidearen bidez adierazten da (“-ren”, “-ko”, “-rekin”,...). Kasu honetan, euskarazko hitzen lematizazioan informazio hori guztia galdu egiten da, nahiz eta ingelesean informazioa mantendu. Hori horrela izanik, ezinbestean landu beharko dugun atala da, eta IXA taldeak eta Elhuyar fundazioak elkarlanean sortutako Erauzterm (Gurrutxaga *et al.*, 2005) terminologia-erauzlea⁸ eredu harturik metodoa hobetu beharko dugu.

Euskara normalizazio-prozesuan dagoen hizkuntza izanik, termino berriak etengabe sortzen dira. Gure nahia da, prototipoak automatikoki identifikatu ez dituen terminoak terminologia-zerbitzarira eskuz gehitzeko aukera ematea. Horrela, termino edota adiera berriak eskuz identifikatu ahal izango ditugu eta euskarazko SNOMED CT aberastu.

Euskarazko SNOMED CTren bertsioa oraindik lantzen dihardugu tesi-lanean, eta prototipoaren funtzionamendu zuzenerako ezinbesteko baliabidea izango da. Izan ere, une honetan dituen 68.495 euskarazko terminoek oso bertsio mugatua osatzen dute ingelesezko 1.000.000 terminoekin alderatuz gero. Hortaz, euskarazko bertsioa automatikoki osatzeaz gain, eta honen adituen berrikuspen eta zuzenketaz gain, prototipo honen erabilera SNOMED CT bera elikatzeke aukera ezin hobea izan daiteke erabiltzaileek erabiltzen dituzten terminoak integratu ahal izateko.

Osasungintza euskalduntzeko lan handia egiteke badago ere, hemen aurkeztu dugun lana argi-izpi bat izan daiteke. Oraindik orain lana hastapenetan badago ere, oinarri itxaropentsua ezartzen digu etorkizunean osasun-txosten elebidunak gure osasun-sistemaren egunerokoan erabili ahal izateko, eta osasun-langile eta gaixo euskaldunon hizkuntza-eskubideak bermatuak izan daitezen.

Erreferentziak

- AGIRRE, ENEKO, IÑAKI ALEGRIA, XABIER ARREGI, XABIER ARTOLA, ARANTZA DÍAZ DE ILARRAZA, MONTERRAT MARITXALAR, KEPA SARASOLA, eta MIREN URKIA. 1992. Xuxen: A Spelling Checker/Corrector for Basque based in Two-Level Morphology. In *Proceedings of NAACL-ANLP'92*, 119-125. Povo Trento. 1992..
- AIKAWA, TAKAKO, LEE SCHWARTZ, RONIT KING, MONICA CORSTON-OLIVER, eta CARMEN LOZANO. 2007. Impact of controlled language on translation quality and post-editing in a statistical machine translation environment. *Proceedings of the MT Summit XI* 1-7.
- DESJARDINS, L., 2003. Le santé des francophones du Nouveau-Brunswick. Petit-Rocher, Société des Acadiens et des Acadiennes du Nouveau-Brunswick.
- EZEIZA, NEREA, ITZIAR ADURIZ, IÑAKI ALEGRIA, JOSE MARIA ARRIOLA, eta RUBEN URIZAR. 1998. Combining Stochastic and Rule-Based Methods for Disambiguation in Agglutinative Languages. In *COLING-ACL'98. Pgs. 380 - 384. Vol 1. Montreal (Canada). August 10-14, 1998*.
- GERKENS, SOPHIE, eta SHERRY MERKUR. 2010. Belgium: Health system review. *Health Systems in Transition* 12.1-266.
- GURRUTXAGA, A., X. SARALEGI, S. UGARTETXEA, eta I. ALEGRIA. 2005. Erauzterm: euskarazko terminoak erauzteko tresna erdiautomatikoa. In *Mendebalde Kultur Alkartea, IX. Jardunaldiak: Euskera zientifiko-teknikoa. ISBN 84-931882-5-5*.
- IHTSDO, INTERNATIONAL HEALTH TERMINOLOGY STANDARDS DEVELOPMENT ORGANISATION. 2014. SNOMED CT Starter Guide. February 2014. Technical report, International Health Terminology Standards Development Organisation.
- JOANES ETXEBERRI SARIA V. EDIZIOA (ed.) 2014. *Donostia Unibertsitate Ospitaleko alta-txostenak*. Donostiako Unibertsitate Ospitalea, Komunikazio Unitatea.

⁸Terminologia erauzketaren helburua corpus bat oinarri izanik, automatikoki termino garrantzitsuak eskuratzean datza. [Wikipedia]

- MITAMURA, TERUKO. 1999. Controlled language for multilingual machine translation. In *Proceedings of Machine Translation Summit VII, Singapore*, 46–52.
- MURFF, HJ, F FITZHENRY, ME MATHENY, eta ET AL. 2011. Automated identification of postoperative complications within an electronic medical record using natural language processing. *JAMA* 306.848–855.
- ON HEALTH CARE SYSTEMS, EUROPEAN OBSERVATORY. 1999. Luxembourg: Health system review. *Health Systems in Transition* .
- OSAKIDETZA (ed.) 2013. *Osakidetzan Euskararen Erabilera Normalizatzeko 2. Plana (2013-2019)*.
- PADRÓ, LLUIS, eta EVGENY STANILOVSKY. 2012. Freeling 3.0: Towards wider multilinguality. In *Proceedings of the Language Resources and Evaluation Conference (LREC 2012)*, Istanbul, Turkey. ELRA.
- PÉREZ, ALICIA, KOLDO GOJENOLA, ARANTZA CASILLAS, MAITE OROÑOZ, eta ARANTZA DÍAZ DE ILARRAZA. 2015. Computer aided classification of diagnostic terms in spanish. *Expert Systems with Applications* 42.2949 – 2958.
- PEREZ-DE-VIÑASPRE, OLATZ, eta MAITE OROÑOZ. 2014. SNOMED CT in a language isolate: an algorithm for a semiautomatic translation. *BMC Medical Informatics and Decision Making (Argitaratzeke)* .

5 Eskerrak eta oharrak

Lan hau ezinezkoa izango litzateke Health Language Analytics (HLA) eta Innovative Clinical Information Management Systems (iCIMS) enpresek eskuzabalki emandako laguntza eta tresnarik gabe, eskerrik beroenak bertako langileei.

Osasungoa Euskalduntzeko Erakundea ere eskertu nahiko genuke, inkestan parte hartu, eta eskaintako laguntza guztiagatik, eta bereziki Karlos Ibarguren eta Felix Zubiari, lanaren garapenean sortutako zalantza guztiak argitu, eta beti hor egoteagatik.

Lan hau Eusko Jaurlaritzak emandako beka biri esker garatu da: BFI-2011-389 doktoretza-aurreko programa eta EP-2014-1-173 Egonlabur programa.