

EuSnomed: osasun-terminoen euskaratze automatikoa

Olatz Perez-de-Viñaspre, Maite Oronoz

Hizkuntzaren Prozesamendurako IXA taldea (UPV/EHU)

olatz.perezdevinaspre@ehu.eus

Laburpena

Artikulu honetan, osasun-zientzietako terminologia euskaraz lortzeko garatu ditugun metodo automatikoen berri ematen dugu labur-labur. Horretarako, erreferentziazko baliabide terminologiko bat hartu dugu abiapuntutzat (SNOMED CT) eta berori euskaratu dugu. Horretaz gain, etorkizunerako planteatzen ditugun lanak ere aurkezten ditugu, hala nola terminoen balidazioa eta osasun-txostenen itzulpen automatikoa.

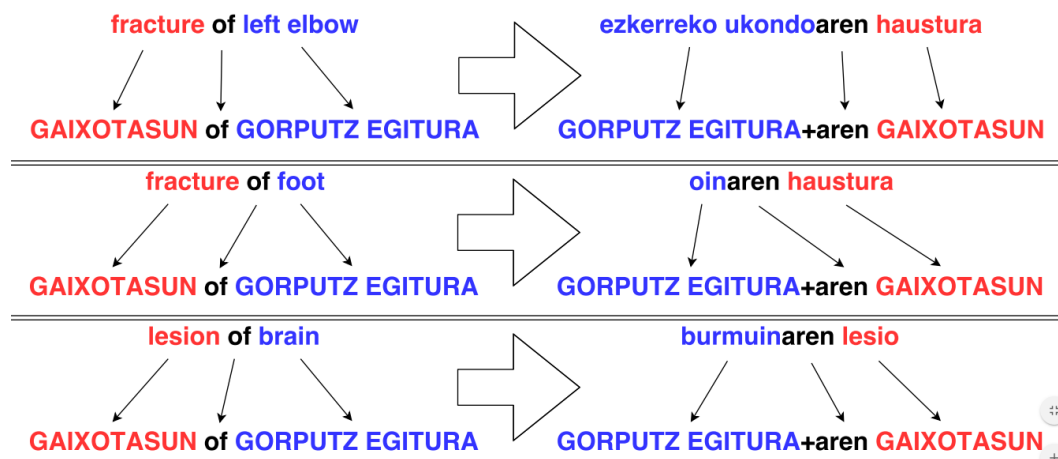
Hitz gakoak: Terminologia, SNOMED CT, Itzulpen automatikoa.

Gaur egunean, osasun-zientzien lan-eremuan euskararen normalizazioan lehenasunezko beharretako bat terminologia espezializatuarena da. Premia horri aurre egiteko, osasun-terminoak automatikoki euskaratzeko sistemak garatu eta ebaluatu ditugu. Horretarako, *Systematized Nomenclature of Medicine – Clinical Terms* (SNOMED CT) terminologia kliniko zabala barnebiltzen duen ontologia hartu dugu abiapuntutzat. SNOMED CT (1) herrialde askotan estandar gisa erabiltzen hasiak dira osasun-txostenak idazteko eta aztertzeke. Terminologia hori gure hizkuntzan izateak ingelesa zein espainiera bezalako hizkuntzekin parekatzen du euskara osasun-alorreko terminoen erabilerari dagokionez. Euskaratzea kudeatzeko EuSnomed deritzon sistema garatu dugu. Ingelesa hartu dugu itzulpenerako oinarritzko hizkuntzatzat, SNOMED CTren jatorrizko bertsioa ingelesez garatu baita.

EuSnomedek lau urrats dauzka terminoen euskarazko ordainak lortzeko (2):

- Lehenengo urratsak baliabide lexikalak erabiltzen ditu SNOMED CTren terminoei euskarazko ordainak zuzenean esleitzeko. Horretarako, besteak beste, Euskalterm banku terminologikoa, Zientzia eta Teknologiaren Hiztegi Entziklopedikoa, Giza AnATOMIako Atlas eta Erizaintzako hiztegiak erabili ditugu.
- Bigarren urratserako, ingelesezko termino neoklasikoak (latinetik zein grezieratik datozenak) euskaratzeko NeoTerm sistema garatu dugu. Sistema horrek morfema neoklasikoen baliokidetzak (ingelesezko «necro-» morfema euskaraz «neko-» litzateke) eta transliterazio-erregelak (ingelesezko «ph» fonema euskaraz «f» litzateke) erabiltzen ditu euskarazko ordainak sortzeko. Adibidez, ingelesezko «radionecrosis» terminoarentzat, euskarazko «erradionekrosi» ordaina sortzen dugu.
- Hirugarrenerako, ingelesezko termino konplexuak (hitz bat baino gehiago dituztenak) euskaratzen dituen KabiTerm sistema garatu dugu. KabiTermek termino konplexuetan agertzen diren habiaratutako terminoen egiturak erabiltzen ditu euskarazko egiturak sortzeko, eta horrela termino konplexuak osatzeko, hurrengo irudian ikus dezakegun moduan.

1. irudia. KabiTermek erabiltzen dituen egituren adibide bat.



- Azken urratsean, itzulpenerako erregelak erabiltzen dituen Matxin (3) itzultzaile automatikoa osasun-zientzien domeinura egokitu dugu, MatxinMed sortuz. Horretarako Matxin osasun-alorrera egokitzeko prestatu dugu, eta, besteak beste, hiztegia zabaldu diogu osasun-zientzietako testuak itzuli ahal izateko. Horrela, matxinek «colon cancer», «kolon-minbizi» itzul dezake, eta ez jatorrian itzultzen zuen «bi puntu minbizia», ingelesezko «colon» hitzaren beste adiera erabiliz.

Garatutako lau urratsak ebaluatu ditugu metodo ezberdinak erabiliz. Alde batetik, mediku eta hizkuntzalariekin osatutako aditu talde txiki batekin egin dugu lehenengo bi urratsen ebaluazioa. Bestetik, Medbaluatoia kanpainaren barruan osasun-zientzietako euskal komunitateari esker egin dugu azkeneko bi urratsetako sistemen ebaluazioa. Bereziki NeoTermen eta KabiTermen emaitzak izan dira azpimarragarriak, metodo erabat automatikoak erabiliz, kalitatezko euskarazko ordainak sortzen baitituzte.

Sortutako terminologiaren balioa erakusteko bi aplikazio garatu ditugu (4): XuxenMed osasunera egokitutako zuzentzaile ortografikoa eta osasun-txostenak euskaraz idazteko laguntza-prototipoa. Xuxen zuzentzaile ortografikoari hiztegia zabaldu diogu osasun-zientzietako terminoak ezagutu ditzan, oker faltsuak ekidinez. Osasun-txostenak euskaraz idazteko laguntza-prototipoari dagokionez, osasun-langileen lana errazte aldera zaintza klinikorako informazioa kudeatzen duen sistema bat sortu dugu, zeinak aurreidatzitako edukia erakusten duen eta idatzitako terminoak ezagutu eta itzultzeko gai den. Prototipo horrek muga handiak ditu, eta lanean hasi gara beste prototipo malguago eta ahaltzuago baten garapenean.

2. irudia. Osasun-txostenak euskaraz idazteko lehenengo prototipoaren adibide bat.

OSPITALERATZEKO ARRAZOA		Hiztegi-kontsulta
Anemia duen 43 urteko gizona.		Sarrera <input type="text"/>
AUI	<input type="checkbox"/> 271737000 anemia	
Ez du alergiairik. Ez du ohiko tratamendurik.		
EGI	<input type="checkbox"/> 609328004 disposición alérgica	
Du	<input type="checkbox"/> 415178003 proceso	atik hona gorotzak beltzak direla iruditu zaio. Oheburuko
mé	<input type="checkbox"/> 276239002 tratamiento	anemia duela ikusi da. Ez du sabeleko minik sentitzen, eta ez
du	<input type="checkbox"/> 119270007 tratamiento	
AZTERKETA FISIKOA		
Tentsio arteriala (TA), 130/80 mmHg; bihotz-maiztasuna (BM), 70 tau/min; temperatura, 36,7C. Kontziente eta orientatua dago. Hidratazioa eta perfusioa egokiak dira. Eupneiko. Larruzal eta mukosetako kolore normala.		
Bu	<input type="checkbox"/> 75367002 presión sanguínea	

Euskarazko terminologia finkatzea ezinbestekoa zaigu tresna automatikoen ikerketan jarraitu ahal izateko, eta beharrezkoa dugun hurrengo urratsa SNOMED CTren bertsio balidatua da. Osasun-alorreko langileen inplikazioa eta euskararen alde lan egiteko gogoia izugarria izan da Medbaluatoian, eta gogo, gurari eta behar hori baliatu nahi dugu ingelesezko termino bakoitzarentzako sortzen diren euskal ordain guztien artean, komunitateari erabileran egokiak iruditzen zaizkion euskarazko terminoak aukeratzeko.

Horretaz gain, itzulpen automatikoan aurrerapausoak eman ahal izateko, osasun-txosten elebidunak behar-beharrezkoak ditugu. Orain, osasun-txosten gehienak gazteleraz sortzen dira osasun-langile guztiak ez baitira euskaldunak. Sistema informatikoen osasuneko hizkuntza ikas dezaten, txostenak bai euskaraz eta bai gazteleraz behar ditugu (txosten paraleloak deritzegu). Asmoa dugu Osakidetzarekin proiektu bat definitzeko hori lortu ahal izateko. Eta berriz ere, osasun-alorreko euskal komunitate gogotsu eta parte-hartzailearen laguntza eskatuko dugu.

Bidea egiten ari gara. Jarrai dezagun elkarrekin.

Erreferentzia bibliografikoak

1. Uddin Z, Maennle D, Russell K, Boltri JM. Monoclonal gammopathy of undetermined significance: Using risk stratification to guide follow-up. J Fam Pract. 2015 Jul;64(7):E5-E12. IHTSDO, I.H.T.S.D.O. SNOMED CT Starter Guide. Technical report, International Health Terminology Standards Development Organisation. 2014.
2. Perez-de-Viñaspre O., Oronoz M. "SNOMED CT in a language isolate: an algorithm for a semiautomatic translation". BMC Medical Informatics and Decision Making 2015, 15(Suppl 2):S5
3. Mayor A, Alegria I, Diaz de Ilarraza A, Labaka G, Lersundi M, Sarasola K. Matxin, an Open-source Rule-based Machine Translation System for Basque. Machine Translation. 2011;25:53–82. doi: 10.1007/s10590-011-9092-y. 10.1007/s10590-011-9092-y
4. Perez-de-Viñaspre O., Oronoz M., Patrick J. Osasun-txosten elebidunak posible ote? Ikerketa 2015, Durango