

GARATERM: EUSKARAREN ERREGISTRO AKADEMIKOEN GARAPENAREN IKERKETARAKO LAN-INGURUNEA¹

Igone Zabala eta Mikel Lersundi

(Universidad del País Vasco/Euskal Herriko Unibertsitatea, UPV/EHU)

Igor Leturia eta Iker Manterola

(Elhuyar Fundazioa)

Gotzon Santander

(GARATERM-2 proiektuko ikertzaile kontratuduna)

AURKIBIDEA

Laburpena

Artikulu honetan erregistro akademikoaren garapenari laguntzeko eta garapena bera ikertzeko diseinatu dugun GARATERM lan-ingurunearen oinarriak eta ezaugarriak deskribatuko ditugu. Gure abiapuntuko onarpena izan da euskararen erregistro akademikoak neurri handian unibertsitateko ikasgeletan garatzen direla; hori dela eta, komunikazio-testuinguru akademiko horietan erabiltzen diren testu espontaneo eta “ikusezin” horiek biltzeko tresnak garatzea izan da gure xede nagusia. GARATERM lan-ingurunea diseinatu dugu, hain zuzen ere, aipatutako testu horiek modu arin eta sistematikoan bildu ahal izateko. Lan-ingurunearen zentro neuralgikoa GARATERM corpus monitorea da, eta horren funtsezko gunea unibertsitateko ikasgeletan erabiltzen diren testu espontaneoek osatzen badute ere, kontrol linguistikoa pairatu duten material didaktikoak eta bestelako testuak (artikuluak, eskuliburuak e.a.) ere bildu nahi ditugu, gure helburua baita iturri desberdinetako testuetan erabilitako hizkuntza-baliabideen arteko azterketa konparatiboak ahalbidetzea. Lan honetan, lan-ingurunearen beharra justifikatu ez ezik, testuen kudeaketarako interfazearen eta kontsulta-interfazeen ezaugarriak ere deskribatuko ditugu.

Gako hitzak: corpus akademikoa, corpusgintzarako tresnak, corpusgintzarako metodologia.

¹ Lan-ingurune hau garatu ahal izateko, 2011 eta 2012 urteetan Unibertsitate-Gizarte ikerketa-proiektu batean elkarlanean aritu gara Elhuyar Fundazioko eta UPV/EHUko ikertzaileak: “GARATERM-2: Linguistikoki prozesatutako corpusetarik terminologia eta neologia erauzteko ingurune baten garapena” (OTRI kodea: US10/01).

1. Sarrera

Azken hamarkadan nabariki aldatu dira hizkera espezializatuen ikerketa eta trataera. Izan ere, laurogeiko hamarkadatik aurrera, esan daiteke espezialitate-hizkeren azterketa aurreko ikuspegi lexiko edota morfosintaktikotik testu eta diskurtsoari begirako ikuspegira iragan dela (Ciapuscio 2003). Zalantzarik ez dago diziplina-alorra eta gaia hizkera espezializatuak elkarren artean bereizteko parametro saihestezinak direla baina, bestalde, egun oro har onarturik dago espezialitate-alor bakarraren barruan aldakortasun handia dagoela komunikazio-egoera desberdinekin lotutako hizkuntza-baliabideen hautapenean eta aukera bakoitzaren maiztasunean. Hori dela eta, esan daiteke espezialitate-hizkerez hitz egiten dugunean unitate birtualez ari garela, alor tematiko batekin lotutako unitateez, eta unitate horien barruan hainbat erregistro² bereiz daitezkeela³ (Lamuela 1983: 55 eta Hudson 1996: 47, besteak beste). Ildo honetatik, espezialitate-hizkerez baino areago hizkera edo erregistro profesionalaz eta akademikoez hitz egiten da egun.

Ikuspegi berriak bat egin du corpus-hizkuntzalaritzaren garapenarekin. Izan ere, corpusgintza konputazionalak, askotariko testu-masa handiak biltzea eta testuinguruaren faktore ugariarekin harremanetan jartzea bideragarri egin duelarik, posible egin du lehenago pentsaezinak ziren hizkuntza-azterketak aurrera eramatea. Erregistro akademiko eta profesionalen karakterizazioa neurri handian eskutik lotuta joan da xede berezietarako hizkeren (LSP⁴) irakaskuntzarekin eta, aipatutako ildotik, esan daiteke egun oro har onartuta dagoela xede profesional edota akademikoetarako hizkeren irakaskuntzak ezinbestekoa duela aldeaz aurretik espezializatutako corpus adierazgarriak eraiki eta erregistro horien karakterizazioa egitea (Biber 1988, 1994, 2006; Swales 1990, 2001, 2004; Parodi 2005, 2007; Cabré eta Gómez de Enterría 2006, besteak beste).

Flowerdew-k (2004) azaldu duenez, erregistro akademikoak ikertzeko ezinbestekoa da horretarako propio diseinatutako corpusak erabiltzea. Izan ere, corpus orokorrek espezializatutak har daitezkeen azpicorpusak eduki baditzakete ere, Flowerdew-ren iritziz, azpicorpus horiek ez dira egokiak izaten erregistro espezializatuak aztertzeke, hainbat arrazoiengatik:

- 1) Corpus orokorrak hizkuntza oro har islatzeko biltzen dira, eta presentzia oso txikia izan ohi dute erregistro akademiko edota profesionalak aztertzeke interesgune izan daitezkeen generoek.
- 2) Corpus orokor batek tamaina egokiko azpicorpus espezializaturen bat duenean ere, zaila izan ohi da bilaketak interesatzen zaizkigun eremuen arabera egin ahal izatea, halako corpusak ez baitira diseinatzen erregistro espezializatuen ikerketaren beharretarako.

² Biber-ek (1995: 1) honela definitzen du erregistroa: 'Any variety associated with particular situational contexts or purposes'. Baina badira aldaerez (*variety*) baino, aldakortasunaz (*variation*) hitz egitea nahiago duten soziolinguistak ere (Ferguson 1994), aldaerak baino, diskurtso motak (*discourse types*) terminoa nahiago dutenak, edo hizkuntza-aldaerak eta erregistroak bereizten dituztenak (Lamuela 1983: 48). Azkenik, hizkuntzalaritzaren zenbait hurbilketatzen genero eta erregistro terminoak sinonimotzat hartzen dira.

³ Ondo islatuta geratzen da ikuspegi hori Duarte-ren (1990) hitzetan:

«Dins de la classificació de les varietats de la llengua, situem òbviament els llenguatges d'especialitat entre als varietats funcionals. De fet els llenguatges d'especialitat es componen de registres diversos aglutinats per una unitat de camp temàtic.»

⁴ Language for Special Purposes

3) Diskurtso mota batzuk oso eskuragaitzak dira, esate baterako, ahozko diskurtsoak eta konfidentzialak izan ohi diren genero ezkutua (*occluded genres* (Swales 1996)), eta ez dira jasotzen ohiko corpusetan.

4) Zenbait corpus orokorrek ez dituzte jasotzen testuak osorik, baizik eta testu zatiak, eta horrek zaildu egiten du zenbait ezaugarri testu atal jakin batzuekin lotzea. Azken batean, corpusak helburu jakin batzuekin eraikitzen dira, eta jomugan hartu diren helburuetarako izaten dira egoki, baina ez aurreikusi gabeko helburuetarako.

Espezializatutzat har daitezkeen komunikazio-egoera profesional eta akademiko desberdinetan testu-genero desberdinak erabiltzen dira, eta genero desberdinak erregistro desberdinak baliatzeagatik bereizten dira. Hartara, ikerketa bakoitzaren helburuen arabera, mota jakin batzuetako testu-generoak bildu eta inbentariatu dira. Parodi-k (2007) *continuum* bat bezala irudikatzen du genero orokorretatik genero akademiko, profesional eta zientifikoetara doana; *continuum* horren erdigunean kokaturik daude unibertsitateko genero akademikoak, unibertsitatean prestatzen baitira alorrez alorreko profesionalak eta adituak, eta unibertsitatean ikasten baitituzte aditu-profesional modura aritzeko beharrezkoak izango dituzten hizkuntza-baliabideak. Ildo horrekin lotuta, Biber-ek (2006) nabarmentzen du askotan erregistro espezializatuen karakterizazioak genero bakar batean (idatzitako ikerketa-artikuluetan) oinarrituta egoten direla, baina unibertsitateko ikasketetan askotariko genero eta erregistroak erabiltzen direla, esate baterako, irakasgaien irakaskuntza-gidak, ahozko klase magistralak, praktika-protokoloak eta ariketak. Erregistro akademikoak karakterizatzeari begira, funtsezkoa da genero horiek guztiak kontuan hartzea, horietako bakoitzak mota jakin batzuetako hizkuntza-baliabideak ustiatzen baititu.

Erregistro espezializatuen karakterizazioaz orain arte esan duguna, erregistro horien bereizgarri garrantzitsuenetakoa den terminologiari ere zabaldu behar zaio. Izan ere, egun terminologiarako hurbilketa klasikoaren sinplismoa gainditutzat har daitekeelarik, oro har onarturik dago edozein espezialitate-alorretako terminologiak alor horretako adituen diskurtsoaren bitartez garatzen, egonkortzen eta eboluzionatzen joaten direla. Onarpen hori terminologiarako hurbilketa komunikatibo, sozioterminologiko eta soziokognitibo berriek (Cabré 1999, 2001; Boulanger 1991; Guespin 1993; Temmerman 2000) partekatzen duten oinarri teoriko-metodologikoa da, eta adituek erabiltzen duten terminologia eta fraseologia erreala zein den, nola sortzen eta nola aldatzen den ikertzean datza hurbilketa horien interesgunea. Beraz, terminologiari buruzko ikerketak aurrera eramanez ahal izateko ere, ezinbestekotzat jotzen da adituek komunikazio espezializatuan erabiltzen dituzten testuak biltzen dituzten corpusak osatzea eta aztertzea.

Lan honetan euskararen kasuaz ari garelarik, euskara oraindik ere normalizazio-prozesuan murgilduta dagoen hizkuntza izatetik datozen zenbait berezitasun gaineratu behar diegu zerrendatutako onarpen teoriko-metodologikoei. Batetik, euskararen estandarizazio prozesuak hamarkada batzuk baino ez ditu, eta oraindik ere bukatu gabe dago. Bestalde, euskara duela gutxi hasi da erabiltzen esparru akademikoetan eta presentzia txikia du oraindik ere alor profesionalean. Hartara, oraindik ez da nahikoa denbora egon espezialitate-alorrez alorreko erregistroak, eta bereziki, terminologia eta fraseologia erabat garatzeko eta finkatzeko (Elordui eta Zabala 2005; Zabala eta Elordui 2006). Azkenik, biziberritze prozesuan dagoen euskara bezalako hizkuntza gutxituetan, motibazio teoriko hutsaz harantzago joanda, erregistro espezializatuak eta terminologia garatzen eta normalizatzen laguntzeko xedea duten

ekimenek are ezinbestekoago dituzte corpusak, adituek egiten dituzten hizkuntzaren erabilera erreala ezagutuko badituzte. Izan ere, erregistro espezializatuen eta terminologiaren erabilera erreala kontuan hartu gabe egindako interbentzioek baliabide espezializatu horien garapena eta autorregulazioa oztopatzeko edota desorekatzeko arriskua izan lezakete (Zabala et al. 2013).

Gogoeta horietatik abiatu zen, hain zuzen ere, GARATERM egitasmoa (Zabala et al. 2008), eta lehen ondorioa izan zen euskararen erregistro espezializatuen irakaskuntza erregistro horien garapenarekin eta garapen horren ikerketarekin uztartzeko lan-ingurune bat behar zela. Egitasmo horretatik abiatuta emandako urratsen ondorio da, hain zuzen ere, lan honetan aurkezten dugun lan-ingurunea⁵. Nolanahi ere, azpimarratzekoa da ingurune horren diseinu eta inplementazioa funtsezko beste osagai batekin lotuta eraman dela aurrera, alegia, erregistro akademikoaren garapenaren protagonisten inplikazioa posible egiten ari den Terminologia Sareak Ehunduz (TSE) programarekin⁶. Izan ere, euskararen erregistro akademikoaren garapena neurri handian unibertsitateko ikasgeletan gertatzen da, baina garapen horretan erabiltzen diren testuak eskuratzen zailak dira, askotan ahozko testuak edota elaborazio maila baxuko testu-generoak direlako. UPV/EHUko Euskara Errektoreordetzaren babespean 2008-2009 ikasturtean abian jarri zen TSE programari esker, lortzen ari gara zenbait irakaslek beren irakasgaiaren irakaskuntzan erabiltzen dituzten testuak eta terminologia beste irakasleentzat eta hizkuntzalariontzat ikusgai jartzeko prozesuko parte aktibo bihurtzea (Zabala et al. 2011, 2013).

Bigarren atalean, GARATERM lan-ingurunearen justifikazioa eta deskripzio orokorra egingo dugu. Hirugarren atalean, testuen inbentarioa egiteko erabiltzen den interfazeaz, oinarrian duen datu-baseaz eta testuen etiketatzeaz arituko gara. Laugarren atalean, informazio linguistikoa erauzteko eta kontsultak egiteko interfazeak izango ditugu hizpide. Azkenik, bosgarren atalean zenbait ondorio laburbilduko ditugu.

2. GARATERM lan-ingurunearen justifikazioa eta deskripzioa

GARATERM lan-ingurunea euskararen erregistro akademikoaren garapena ikertzeko eta garapen horri laguntzeko xedeaz diseinatu da. Hortaz, haren zentro neuralgikoa GARATERM corpusa da. Atal honetan corpus horren beharra justifikatuko dugu; izan ere, euskarak baditu hainbat corpus eta pentsa liteke ahaleginak alferrik galtzea dela beste corpus bat eraikitzea.

Hasteko, badira hainbat corpus orokor, zenbait testu espezializatu ere bildu dituztenak. Horien berrikustapen azkarra egingo dugu, GARATERM corpusaren ezaugarriak azaldu aurretik.

Hasteko, *XX. mendeko euskararen corpus estatistikoa*⁷ dugu, XX. mendean sortutako orotariko testuen lagin orekatuak biltzen dituen. Corpus honek bilaketak urte-tartean arabera, euskalkien arabera eta testu-motaren arabera egiteko aukera ematen du. Nolanahi ere, testu-motak corpus orokor baten ikuspegitik diseinatu daudelarik, genero akademikoak eta profesionalak ez daude erregistro espezializatuen berariazko ikerketak egiteko beharrezkoa genukeen moduan antolatuta. Bestalde, ez daude identifikatuta jakintza-alorrak eta, beraz,

⁵ Garaterm lan-ingurunearen garapena kokatzen den ikerketa-proiektuak ondoko helbidean kontsulta daitezke: www.garaterm.ehu.es/

⁶ TSE programari buruzko informazio xehea aurki daiteke helbide honetan: http://garaterm.ehu.es/garaterm_ataria/tse-programa/

⁷ <http://www.euskaracorpora.net>

erregistro espezializatuei buruzko ikerketak egiteko ezinbestekotzat jotzen den aldakortasun-faktore hori ezin da kontuan hartu bilaketetan. Azkenik, bistan da XXI. mendea bigarren hamarkadan sartua dela, eta, gainera, iragan den hamarkada gaketzat har daitekeela euskararen garapenaren ikuspegitik, besteak beste, Hiztegi Batuaren ezarpenaren ondorioz.

Bigarrenik, UPV/EHUko Euskara Institutuak eraikitako *Ereduzko Prosa Gaur* (EPG) eta *Ereduzko Prosa Dinamikoa* (EPD)⁸ izeneko corpusak ditugu. Hauek ere, corpus orokorrak dira, baina testu-laginak gabe, testu osoak biltzen dituzte. Bestalde, testuak aukeratzeko erabili den irizpidea testuen eredugarritasuna izan da erabilera errearen adierazgarritasuna baino areago. Bilaketen emaitzak bi ataletan banatuta erakusten dituzte: prentsa eta liburuak. Horrez gain, bilaketa murriz daiteke, urtearen arabera (2000tik aurrekoak edota, urte horretatik aurrera, urtez urte bereizita), edota "itzuliak vs jatorrizkoak" parametroaren arabera. Idazle jakin baten testuen barruko bilaketa murriztua ere eska daiteke. Bistan da kasu honetan ere ez gaudela erregistro akademikoaren garapena aztertzeke baldintzak dituen corpus baten aurrean: hasteko, genero akademikoak ez daude ondo islatuta, eta ezin dira bereizi bestelako generoetatik; bigarrenik, ez dira bereizten jakintza-alorrak eta, azkenik, denetariko erabilera errearen baino, corpusaren egileen arabera "kalitate-ezaugarri" batzuk betetzen dituzten testuak bildu dira⁹.

Hirugarrenik, Euskaltzaindiak garatutako *Lexikoaren Behatokiaren Corpusa*¹⁰ dugu. Komunikabideek sortzen edo plazaratzen dituzten materialak biltzen dituen corpus monitoria da, alegia, azterketa diakronikoak egin ahal izateari begira, etengabe elikatzen doana. Bilaketak generoen arabera eta testuek tratatzen dituzten gaiekin lotutako alor eta azpialorren arabera egin daitezke. Nolanahi ere, corpusaren diseinuan bestelako helburuak egon direlarik, genero, eragile eta jakintza-alor akademikoak ez daude nahikoa ondo islatuta eta bereizita, jomugan ditugun ikerketak egin ahal izateko.

Ezin aipatu gabe utzi berriki garatu diren bi corpus handiak: Euskara Institutuak eraikitako *Egungo Testuen Corpusa* (ETC)¹¹ eta Elhuyar Fundazioak bildutako *Web Corpusen Ataria* (WCA)¹². Bi corpus horiek oso handiak dira, hurrenez hurren, 204 eta 120 milioi hitzetik gorakoak. Nolanahi ere, corpus orokorrak dira hauek ere, eta bilaketak egiteko orduan ezin dira bereizi erregistro akademikoaren garapena ikertzeke ezinbestekoak zaizkigun generoak eta jakintza-alorrak.

Corpus orokorrak alde batera utzita, euskarak badu baita espezializatutzat har daitekeen corpusa ere, alegia, Elhuyar Fundazioak eta IXA ikerketa-taldeak elkarlanean eraikitako *ZT corpusa*¹³. Haren izenak iragarri bezala, zientzia eta teknologiari buruzko testu-bilduma da, eta ikuspegi horretatik egituratuta eta etiketatuta dago. 1990-2002 bitartean argitaratutako zientzia eta teknologiaren alorreko obrak biltzen ditu, jakintza-alorraren eta generoaren arabera sailkatuta. Aztertutako guztien artean iturririk egokiena izan liteke gure ikerketarako

⁸ <http://www.ehu.es/euskara-orria/euskara/ereduzkoa>
<http://ehu.es/ehg/epd/>

⁹ Corpusgintzan funtsezkotzat hartu ohi den irizpide bat da corpusak bakarrik *kanpo-irizpideen* arabera diseinatu behar direla, alegia, testuaren komunikazio-zereginak bakarrik kontuan hartuta eta testuen beraien hizkuntza-xehetasunak (*barne-irizpideak*) kontuan hartu gabe (Clear 1992; Sinclair 2005).

¹⁰ <http://lexikoarenbehatokia.euskaltzaindia.net>

¹¹ <http://www.ehu.es/etc/>

¹² <http://webcorpusak.elhuyar.org/>

¹³ www.ztcorpusa.net

corpus hau baina, hala ere, labur geratzen da gure ikerketak behar dituen zenbait ezaugarri dagokienez. Batetik, giza eta gizarte-zientzietako testurik ez dago jasota corpusean eta, bestalde, testuen espezializazio maila, neurri batean kontuan hartu bada ere, oso maila txikian daude islatuta unibertsitateko irakaskuntzan nagusi diren generoak (ikasle-gidak, apunteak, *Power Point*eko aurkezpenak, praktika-protokoloak, problema-bildumak, azterketak, eta abar).

Euskara aztertzeko eskuragarri ditugun corpusen bana-banako berrikustapen honetan aipatu ditugun oztopoak ez ezik, aipaturako corpusek badituzte jomugan ditugun helburuetarako oztopo diren beste zenbait ezaugarri. Hasteko, corpusgintzan ohikoa den bezala, corpus horietan nolabaiteko kontrol linguistikoa izan duten testu argitaratuak biltzen dira eta, beraz, erabilera erreal espontaneo baina areago kontrol linguistikoaren iragazkitik pasatako erabilerak erakusten dizkigute. Bigarrenik, ez digute ematen testuen ekoizleen aditutasun mailaren berri¹⁴ eta ezaugarri hori funtsezkoa da ikusmiran ditugun ikerketahelburuetarako. Azkenik, informazio linguistikoa erauzteko ezartzen dizkiguten mugak aipatu behar ditugu. Zerrendatutako corpus guztiak lematizatuta egonda, konkordantziak bilatzeko aukera ematen digute, bai eta horien maiztasuna eskaintzen ere. Areago, ETC eta WCA corpus berriek maiztasun handiko konbinazioak bilatzeko aukera ere eskaintzen digute. Bide honetatik, hizkuntza-elementu batzuk, adibidez, termino baten zenbait aldaki, bilatu eta horiek corpusetan dituzten agerraldiak ikusi eta kuantifika ditzakegu. Hizkuntza-egitura jakin batzuen inguruan egindako hipotesiak egiaztatze adibide errealak ere lor ditzakegu horietatik. Nolanahi ere, ez digute eskaintzen beste ikerketa mota batzuk aurrera eramateko bilaketak egiteko aukerarik, adibidez, corpusean edo interesatzen zaigun zatian erabilitako terminologia ezin dugu erauzi; ezin ditugu erauzi gramatika-kategoria jakin bateko elementuak maiztasunaren arabera antolatuta, edota ezin ditzakegu egin corpus-atal desberdinetan erabilitako erregistroak karakterizatze agerkidetzapatroien bilaketak, dimentsio anitzeko azterketak egin ahal izateko. Era horretako azterketek bi eskakizun dituzte: batetik, corpusak linguistikoki egokiro etiketatuta egon behar dira eta, bestetik, kontsulta-interfazea aipaturikoak bezalako galderak egiteko prestatuta egon behar da (ikus, adibidez, Biber et al. 1998; McEnery eta Hardie 2011). Aipaturako azterketen adibideak aurki daitezke Douglas Biberren lanetan (Biber et al. 1998; Biber 2006).

Arazo horiek gainditzeko asmoz diseinatu dira GARATERM lan-ingurunea eta corpora. Corpusgintzan ohikoa den bezala, hiru eratako erabakiak hartu behar izan ditugu:

- a) Zer eratako testuak bildu
- b) Nola antolatu testu horiek, alegia, zer eratako metadatuak jaso testu bakoitzari buruz
- c) Nola etiketatu testuak
- d) Zer eratako kontsultak ahalbidetu.

Corpusa elikatze lehen iturria Terminologia Sareak Ehunduz programan parte hartzen duten EHUko irakasleen testuek osatzen dute, alegia, irakasle horiek ematen dituzten irakasgaietan erabiltzen dituzten materialek: irakaskuntza-gidak, apunteak, aurkezpenak,

¹⁴ EPG, EPD eta ETC corpusek aukera ematen dute bilaketak testuak "jatorrizkoak vs itzuliak" parametroaren arabera murrizteko, baina ezaugarri hori ez da nahikoa ekoizlearen ezaugarriak egokiro murrizteko. Izan ere, badira espezialitate-alor jakin bateko adituek itzulitako testuak eta badira, esaterako, adituak ez diren kazetariak sortutako jatorrizko testuak, hertsiki testu espezializatutzat ezin har daitezkeenak.

ariketak, praktika-protokoloak eta azterketak. Dokumentu horiek bildu eta ikusgai egiteari ezinbesteko deritzogu, egunero ikasgeletan egiten den euskararen erabilera errealean berri eduki ahal izateko. Testu espontaneoak dira eta, beraz, erroreak eduki ditzakete. Nolanahi ere, funtsezkoak dira unibertsitateko irakasleek egunero erabiltzen duten terminologiaren eta fraseologiaren argazkia lortzeko, bai eta erregistro akademikoaren garapenean egon litezkeen arazo edota oztopoen diagnosis egin ahal izateko ere. Helburua da ahal den eta irakasle gehienek eta ahal den eta irakasgai gehienetako testuak biltzen dituen corpus monitore bat (Sinclair 1982) osatzen joatea. Neurri handian espontaneotzat har daitezkeen testu horietaz gain, nolabaiteko kontrol linguistikoa izan duten beste testu akademiko batzuk ere bildu nahi ditugu: EHUko ikasmaterialen sareko argitalpenak, itzulitako ikasliburuak eta bestelako testuak ere (ahozko azalpenak, hitzaldiak, artikulak, gradu lanak, master-tesiak, doktoretza-tesiak, eta abar). Azken helburua da iturri desberdinetako testu espontaneoetan eta kontrol linguistikoa izan duten testuetan egiten diren hizkuntza-hautuak konparatzeko aukera ematen duen corpus monitorea osatzen joatea.

Nolanahi ere, edozein corpus ahal den eta informazio linguistiko gehien erazteko erabilgarria izango bada, funtsezkoa da testu bakoitzaren gainean ahal den eta informazio xeheena biltzea. Informazio hori metadatuaren bitartez jasotzen da eta corpusak atzean duen datu-basea antolatzeke erabiltzen da. Metadatuak ezinbestekoak dira erregistroen azterketetan corpusetik erazten dugun informazio linguistiko kanpo-faktoreekin lotu ahal izateko. GARATERM corpusaren xedea euskararen erregistro akademikoaren garapena aztertzea delarik, garapen hori gertatzen den komunikazio-egoerak eta horiek bereizten dituzten ezaugarri pragmatikoak (ekoizle motak, komunikazio-kanalak, komunikazioen xedea eta testuaren gaia) eta horiekin lotutako testu-generoak (material didaktikoak, hitzaldiak, artikulak, liburuak, eta abar) identifikatu ditugu. Ekoizle moten artean bereziki garrantzitsua da solaskideen aditutasun maila; izan ere, egun oro har onartuta dago testu bat espezializatutzat hartu ahal izateko testuaren ekoizlea behintzat alorreko aditua izan behar dela, adituek beren espezialitate-alorreko kontzeptuen gaineko kontrola baitute (Cabré eta Gómez de Enterría 2006) eta, hain zuzen ere, sortzen dituzten testuetan ikus baitaiteke beren alorreko kategoria semantikoak nola kontzeptualizatzen dituzten (Temmerman 2000). Testu espezializatuen hartzailak askotariko aditutasun mailakoak izan daitezke eta, horien artean, unibertsitatean espezialitate-alor bateko adituak izateko ikasten ari direnei zuzendutako testuak nahiko espezializazio maila altukotzat har daitezke (Cabré eta Gómez de Enterría 2006). Testuak jatorrizkoak ala itzuliak diren ere jasotzen dugu metadatuaren eta, bereziki, testuaren ekoizlea itzultzailea ala alorreko aditua den, ezaugarri horrek baliabideen hautapenean eragina izan dezakeela baiteritzogu. Gainera, material didaktikoen kasuan, material horiek erabiltzen diren titulazioen eta ikastegiaren berri ere jasotzen dugu, hizkuntza-baliabideen sorreran, zirkulazioan eta finkapenean titulazioaren eta ikastegiaren testuinguruak eragina izan lezakeelakoan. Bestalde, testuak sortu diren urtea ere jasotzen dugu, etorkizuneko erregistro akademikoaren garapenean azterketa diakronikoak egiteko aukera zabaltzeari begira.

Azkenik, testua espontaneo ala linguistikoki kontrolatua den ere jasotzen dugu metadatuaren, eta bai zein den kontrol linguistikoaren ardurak izan duen erakundea ere. Izan ere, gure ikerketetan funtsezko deritzogu garapen eta autorregulazio naturalaren ondorio diren erabilerak eta kontrol linguistikoaren ondorio direnak bereizi ahal izateari.

Corpusa eraikitzeke tresnei dagokienez, corpusgintzan ohikoa den bezala, etiketatzerako moduluak eta kontsultak egiteke moduluak bereizi ditugu. Testuen etiketatzeke dagokionez,

TEI formatu estandarra erabiltzen dugu eta, corpusgintzan ohikoa den bezala, testuen metadatuak etiketatzea (goiburuak), etiketatze formala eta etiketatze linguistikoa egiten zaie testuei corpuseratu aurretik. Prozesu horiek aurrera eramateko, Corpusgile tresna (Areta et al. 2006) gure beharretara egokitu dugu. Batetik, tresna horren hiru moduluak (testuen inbentarioa kudeatzeko moduluak, egitura etiketatze moduluak eta etiketatze linguistikorako moduluak) lan-ingurune bakar batean integratu ditugu. Bestalde, ERAUZTERM termino-erazle erdiautomatikoak (Alegria et al. 2004) ere integratu dugu lan-ingurune berean. Azkenik, inbentariorako eta etiketatzerako moduluen interfazeak eraldatu egin ditugu, edozein erabiltzailerentzako atsegina gerta daitezela.

Kontsultei dagokienez, aldiz, bi irteera ditu GARATERM corpusak. Batetik, ERAUZTERM termino-erazle erdiautomatikoaren emaitza (hautatutako corpus zati jakin bateko termino-hautagaien zerrenda) erakusten digun interfazea dugu; bestetik, ohiko kontsulta linguistikoak egiteko hiru interfaze garatu dira: arrunta, aurreratua eta aurreratua. Hurrengo bi ataletan xehekiago deskribatuko ditugu testuen inbentariorako eta prozesatzerako moduluak eta kontsulta-interfazeak.

Web-atari bat garatu dugu (www.garaterm.ehu.es) GARATERM egitasmoari eta Terminologia Sareak Ehunduz programari buruzko informazio guztia interesa duen orenen esku jartzeko, eta erregistro akademikoaren garapena eta ikerketa uztartzeko erabiltzen ditugun tresna eta baliabide guztien sarrerak gune bakar batean biltzeko (1. irudia):

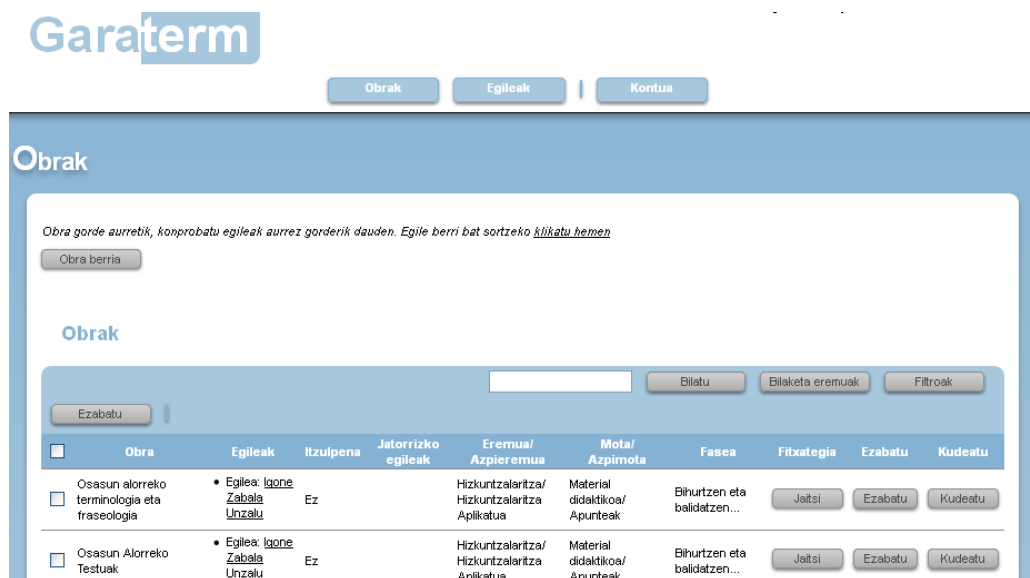


1. irudia. GARATERM: euskararen erregistro akademikoaren garapena ikertzeko lan-ingurunea. Egitasmoari buruzko informazioa eta garatutako tresnak eta baliabideak biltzen dituen web-ataria.

3. Testuen inbentarioa eta prozesatzea

Aurreko ataletan aurreratutako dugu bezala, GARATERM lan-ingurunearen oinarrian Corpusgile tresna (Areta et al. 2006) corpusak modu sistematikoan antolatze eskaintzen dituen metodologia eta teknologia erabili dira. Corpusgile hainbat moduluz osatuta dago: testuen inbentarioa kudeatzeko moduluak, egitura etiketatze moduluak eta etiketatze linguistikorako moduluak. GARATERM proiektu honetan testuen inbentarioa egiteko moduluak nabariki hobetu dugu eta sortu nahi den corpus modular eta dinamikoaren beharretara egokitu dugu. Izan ere, erregistro akademikoaren garapena aztertzea posible egingo digun corpus

monitorea eraikitzeak corpusa etengabe elikatzea eskatuko digu. Bestalde, erregistroen karakterizazioa eta dimentsio anitzeko azterketak egitea ahalbidetuko digun corpusa eraikitzeak, askotariko metadatuak jaso behar ditugu testu bakoitzari buruz. Hartara, koste handikoa gertatuko da testuen inbentarioaren prozesua eta, proiektuari beregaintasuna emateari begira, beharrezkotzat jo dugu testuak kargatzeko prozesua aurrera eramateko interfazea ezagutza tekniko handirik ez duten erabiltzaileentzako egokitzea. Xede hori lortzeko, askotariko formatuen (*html*, *txt*, *Latex*, *Msoffice*, *Openoffice*, eta *pdf*) bihurketa automatikoa inplementatu da. Bestalde, testuak eta metadatuak kargatzeko interfazea intuitiboago eta atseginago egin da. Erabiltzaile arruntek beren kontua kudeatzeko eta beren testuak igotzeko eta kudeatzeko aukera ematen dien karga-interfazerako sarrera dute (2. irudia).



2. irudia. Erabiltzaile arruntentzako karga-interfazerako sarrera.

Interfaze horren bitartez, egileei buruzko datuak eta obra egokiro datu-baseratzeko metadatuak sartzen dira. Hasteko, testua beste corpus batetik hartu den ala ez zehazten da. Bigarrenik, jakintza-eremua eta azpieremua aukeratzen dira, UNESCOren sailkapenean oinarritutako zerrenda batetik. Hirugarrenik, egilearen datuak jasotzen dira: rola (egilea, gainbegiratzailea, itzultzailea, lan akademikoen tutorea/zuzendaria); rol xehea (aurkezlea, bozeramailea, elkarrizketa-egilea, elkarrizketatua, hizlaria, idazlea) eta espezializazioa (aditu irakaslea, aditu profesionala, ikertzailea, kazetaria) (ikus 3. irudia).

3. irudia. Testuak eta metadatuak kargatzeko interfazea: jakintza-eremua eta egileen datuak.

Testuen inbentarioarekin jarraiki, jakintza-eremuaren eta egileen datuak sartu ondoren, testuen beraien karakterizazioarako datuak txertatzen dira. Hasteko, testuaren generoa kargatzen da: artikulua, hitzaldia, ikerketa-lan akademikoa, ikerketa-txostena, jarduera profesionala, komunikabideetako testua, kongresuetako akta, liburua, material didaktikoa edo material lexikografikoa. Aukeratutako generoaren arabera, metadatu jakin batzuk sartzeko aukerak zabaltzen dira. Adibidez, generoan “material didaktikoa” balioa aukeratuz gero, azpigeranoan ondoko balio hauek zabalduko dira: apunteak, ariketa zuzenduak, ariketak, aurkibidea, azterketa, irakaskuntza-gida, irakaskuntzarako barne-txostena, *Power point* aurkezpena eta praktika-protokoloa). Horretaz gain, material didaktikoei dagozkien metadatuak sartzeko gelaxkak ikusaraziko dira: irakasgaiaren izenburua, erakundea, ikastegia, titulazioa, gaia, irakasmila eta ikasturtea. Beste genero bat aukeratuz gero, bestelako eremuak eta balioak zabalduko zaizkigu (4. irudia).

Testu mota

Generoa: Azpigeneroa:

Material didaktikoa

Irakasgaiaren izenburua: Erakundea:

Ikastegia: Titulazioa:

Gaia: Irakasmalla:

Ikasturtea: Oharrak:

Adibidez 2012/2013 kurtsoarako, 2012. jari

4. irudia. Testuak eta metadatuak kargatzeko interfazea: testu-generoekin lotutako eremuak.

Azkenik, testuen jatorriari eta kontrol linguistikoari buruzko datuak jasotzen dira (5. irudia).

Testuaren jatorria

Itzulpena: Jatorrizko izenburua:

Jatorrizko hizkuntza: Jatorrizko erakundea:

Jatorrizko urtea:

1. Jatorrizko egilea: Jatorrizko espezializazioa:

Kontrol linguistikoa

Zuzendua: Erakunde gainbegiratzaila:

Baimena

Baimena: Baimen mota:

5. irudia. Testuak eta metadatuak kargatzeko interfazea: jatorria eta kontrol linguistikoa.

Testuen inbentariarako interfazeak aukera gehiago eskaintzen dizkie eskumen gehiago dituzten erabiltzaileei (erabiltzaile aurreratua, administratzaile arrunta eta administratzaile aurreratua). Adibidez, ikasleen rola eta testu-generoak kargatzeko aukera dute eskumen gehiago dituzten erabiltzaileek.

Testuen inbentarioa osatu ondoren, prozesatze linguistikoa egiten da. Prozesu hori ez dago erabiltzaile guztien eskuetan, eta bakarrik eraman ahal dute aurrera horretarako eskumenak dituzten administratzaileek. Interfazeak (6. irudia) aukera ematen du testuak loteka antolatzeko, eta etiketatze formala eta linguistikoa abian jartzeko. Bestalde, aukera ematen du lematizazio automatikoa eskuz gainbegiratzeko: prozesamendu-hiztegian ageri ez diren hitzei lema bat eta gramatika-kategoria bat esleitzeko, bai eta entitateen izenak, ortografia ez-estandarra eta bestelako akatsak normalizatzeko ere. Eskuzko lan hori egiteko aukera izatea funtsezkoa da testu espontaneoekin lan egiten dugunean, lematizazio- eta etiketatze-errore ugari egoteak emaitzen esangarritasuna nabariki baldintzatu baitezake. Gainera, beste corpus batzuetan biltzen ez diren generoak eta jakintza-alorrak jasotzeak prozesamendu-hiztegietan

ez dauden lema ezezagun ugari agertzea dakar: komenigarria da lema horiek guztiak hiztegitratzen joatea testuen lematizazioa gero eta zehatzagoa izan dadin.

Azkenik, interfaze berean testu-loteak osatu eta ERAUZTERM termino-erazle erdiautomatikora bidaltzeko aukera ere ematen du GARATERM lan-inguruneak. Erregistro akademikoan deskripzioan protagonismo berezia hartzen du terminologiak, baina orain arte testuen corpuseratzea eta terminologia-erazketa prozesu independente eta neketsuak izan dira. Orain inbentariatutako testuak zuzenean termino-erazlerra bidaltzeko aukera gauzatu delarik, nabariki arindu dira corpuseko terminologia aztertzeke eta ustiatzeko zereginak.

The screenshot shows the 'Garaterm' web interface. At the top, there is a navigation bar with buttons for 'Obrak', 'Egileak', 'Hiztegiak', 'Foreign/Corr/Reg etiketatze lotak', 'Etiketatzeko linguistiko lotak', 'Termino eraztuen lotak', 'Kontua', and 'Erabiltzaileak'. Below this, the main content area is titled 'Obrak' and contains a search bar and several filter buttons: 'Ezabatu', 'Bihurtu eta balidatu', 'Hiztegitatu', 'Foreign/Corr/Reg etiketatze', 'Etiketatu linguistikoki', 'Terminoak eraztu', 'Data-baseratu', and 'Kendu datu-basetik'. A table displays the results of a search, with columns for 'Jabea', 'Obra', 'Egileak', 'Itzulpena', 'Jatorrizko egileak', 'Eremua/ Azpierrezua', 'Mota/ Azpimota', 'Fasea', 'Fitxategia', 'Ezabatu', and 'Kudeatu'. The table lists five entries, all related to 'Kalkulus' (Calculus) by 'Patri Anzulo Martin'.

Jabea	Obra	Egileak	Itzulpena	Jatorrizko egileak	Eremua/ Azpierrezua	Mota/ Azpimota	Fasea	Fitxategia	Ezabatu	Kudeatu
<input type="checkbox"/> 1. Ccpenmso	Kalkulus PS	• Egileak: Patri Anzulo Martin	Ez		Matematika/ Analisis eta analisis funtzionala	Material didaktikoa/ Azterketa	Bihurtzen eta balidatzen...	Jatsi	Ezabatu	Kudeatu
<input type="checkbox"/> 2. Ccpenmso	Kalkulus PDF	• Egileak: Patri Anzulo Martin	Ez		Matematika/ Analisis eta analisis funtzionala	Material didaktikoa/ Azterketa	Bihurtzen eta balidatzen...	Jatsi	Ezabatu	Kudeatu
<input type="checkbox"/> 3. Ccpenmso	Kalkulus DVI	• Egileak: Patri Anzulo Martin	Ez		Matematika/ Analisis eta analisis funtzionala	Material didaktikoa/ Azterketa	Bihurtzen eta balidatzen...	Jatsi	Ezabatu	Kudeatu
<input type="checkbox"/> 4. Ccpenmso	Kalkulus 7	• Egileak: Patri Anzulo Martin	Ez		Matematika/ Analisis eta analisis funtzionala	Material didaktikoa/ Azterketa	Bihurtzen eta balidatzen...	Jatsi	Ezabatu	Kudeatu
<input type="checkbox"/> 5. Ccpenmso	Kalkulus 6	• Egileak: Patri Anzulo Martin	Ez		Matematika/ Analisis eta analisis funtzionala	Material didaktikoa/ Azterketa	Bihurtzen eta balidatzen...	Jatsi	Ezabatu	Kudeatu

6. irudia. Testuen prozesamendurako interfazea.

Laburbilduz, esan dezakegu lan-ingurune honen bitartez urrats garrantzitsua eman dela, orain arte modu independentean erabiltzen ziren moduluak ingurune arin eta atsegin batean integratuz. Horrela, testuen corpuseratze prozesua nabariki arintzea eta sistematizatzea lortu da.

4. Kontsulta-interfazeak

Bi mailatako kontsulta-interfazeak ditu GARATERM corpusak. Batetik, ERAUZTERM bitartez erazitako termino-hautagaiak adituek balioztatzeke erakusten dituen interfazea dago (7. irudia). Interfaze horri esker irakasgai jakin batean erabiltzen den terminologia aztertu eta glosategi eleaniztunak sor daitezke, Terminologia Zerbitzurako Online Sistemari (TZOS)¹⁵ (Arregi et. al. 2010) kontsultagai jartzeari begira edota bestelako helburuetarako.

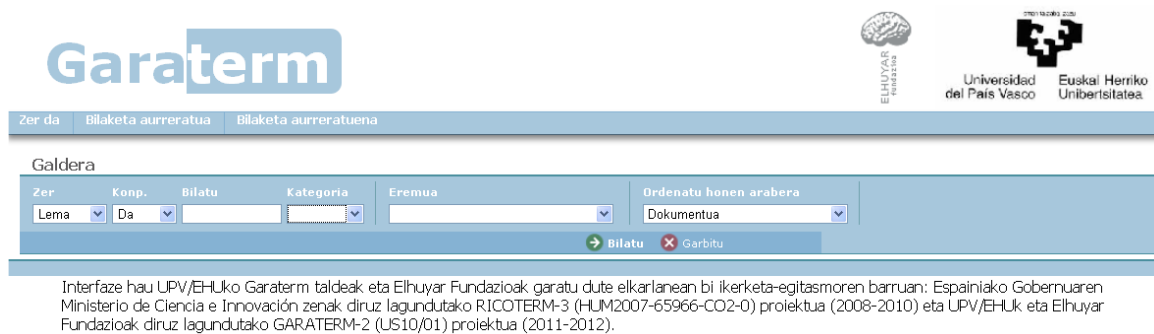
¹⁵ <http://tzos.ehu.es>.

Id.	Forma	Eredu	Alor.	Maiz.	Neur.	Test.	Ter.?
1	programazio-lengoaia	NN	Inform.	14 (14)	100.00	KWIC	✓✗
	ADA programazio-lengoaia	NNN	Inform.	3		KWIC	✓✗
	pascal programazio-lengoaia	NNN	Inform.	1		KWIC	✓✗
	estreinako programazio-lengoaia	AprepNN	Inform.	1		KWIC	✓✗
	programazio-lengoaia Ada	NNN	Inform.	1		KWIC	✓✗
	programa programazio-lengoaia	NNN	Inform.	1		KWIC	✓✗
	programazio-lengoaia agintzaile	NNApos	Inform.	1		KWIC	✓✗
	programazio-lengoaia unibertsal	NNApos	Inform.	1		KWIC	✓✗
2	sistema formal	NApos	Inform.	19 (17)	100.00	KWIC	✓✗
3	datu-mota	NN	Inform.	13 (13)	88.55	KWIC	✓✗

...en sintaxiari dagokionean, formulak idazteko alfabetoa ADA programazio-lengoaian egokituko diogu eta, besteak beste, honako sinbolo hau...
 ...sleitzen dio, aurretik aldagaiak zuen balioa ezabatuz. ADA programazio-lengoaian horrela adierazten ditugu asignazioak...
 ...erazioen kontrol-egiturak erabiltzea da. Atal honetan, ADA programazio-lengoaian de ? nituta dauden kontrol-egitura iteratibo batzuk la...

7. irudia. ERAUZTERM termino-erazle erdiautomatikoak termino-hautagaien balioztatzerako duen interfazea.

Bestalde, ohiko kontsultak egiteko interfazeak daude. Erabiltzaile arruntek kontsulta-interfaze arruntean (8. irudia) eta aurreratuan egin ditzakete bilaketak eta, erabiltzaile aurreratuek eta administratzaileek, aldiz, hirugarren interfaze bat ere badute eskura (interfaze aurreratuena). Kontsulta-interfaze arruntean, lema eta forma ez ezik, lema- eta forma-hasierak eta -bukaerak bila daitezke, eta bilaketak murriz daitezke gramatika-kategoriaren arabera eta jakintza-eremuaren arabera. Bestalde, emaitzak hainbat parametroren arabera ordenatzeko aukera dago.



8. irudia. GARATERM corpusaren kontsulta-interfaze arrunta.

Interfaze aurreratuan, arruntean egiten den bilaketaz gain, hainbat hitzen edo lemaren konbinazioak bila daitezke. Bestalde, bilaketak murrizteko interfaze arruntean eskuragarri ez dauden hainbat parametro erabil daitezke: azpicorpusa, azpierenua, titulazioa, irakasgaia, generoa, azpigeneroa eta urtea. Emaitzak hainbat parametroren arabera ordenatuta eska daitezke eta testuinguruak eta kopuruak, edo kopuruak bakarrik eska daitezke (9. irudia).

9. irudia. GARATERM corpusaren kontsulta-interfaze aurreratu.

Azkenik, eskumenak dituzten erabiltzaileek bakarrik erabil dezaketen kontsulta-interfaze aurreratuena dugu (10. irudia), zeinak kategoria jakin bateko elementu guztiak bilatzeko eta kuantifikatzeko aukera ematen duen. Interfaze aurreratuena interfaze aurreratuak eskaintzen dituen parametroen arabera ez ezik, kontrol linguistikoaren arabera eta jatorrizkoa/itzulpena parametroen arabera ere murrizteko aukera ematen du.

10. irudia. GARATERM corpusaren kontsulta-interfaze aurreratuena.

5. Ondorioak

Edozein hizkuntzaren erregistro akademikoak ikertzeko, ezinbestekoa da helburu horretarako propio diseinatutako corpusak eraikitzea. Euskararen erregistro akademikoak oraindik ere garapenean daudelarik, beharrezkotzat jo dugu garapen horren jarraipena egitea posible egiten digun corpus monitore bat diseinatzea eta corpusa etengabe elikatzen joatea ahalbidetzen digun lan-ingurune arin eta dinamikoa garatzea. Euskararen erregistro akademikoaren garapena neurri handian unibertsitateko ikasgeletan gertatzen da, eta lehentasunezkotzat jo dugu unibertsitate-irakaskuntzarako erabiltzen diren dokumentuak (ikasle-gidak, apunteak, *Power point* aurkezpenak, praktika-protokoloak, azterketak, problemak e.a.) biltzea. Espontaneotzat eta ez oso landutzat har daitezkeen dokumentu

horietaz gain, kontrol linguistikoa pairatu duten dokumentuak eta dokumentu itzuliak ere jasoko dira GARATERM corpusean, azterketa konparatiboak posible egiteko asmoz.

GARATERM lan-ingurunean corpusgintzarako orain arte modu independentean erabiltzen ziren Corpusgile eta ERAUZTERM tresnen moduluak ingurune arin eta atsegin batean integratu dira. Horrela, testuen corpuseratze prozesua nabariki arintzea eta sistematizatzea lortu da. Bestalde, hainbat kontsulta-interfaze diseinatu dira mota desberdinetako erabiltzaileei begira. Bereziki, kontsulta-interfaze aurreratuen euskararen erregistro akademikoen garapena ikertzeko aukera emango diguten dimentsio anitzeko azterketa konparatiboak egin ahal izateko prestatu da.

Bibliografia

- ALEGRIA, I.; GURRUTXAGA, A.; LIZASO, P.; SARALEGI, X.; UGARTETXEA, S.; URIZAR, R. (2004): "An Xml-Based Term Extraction Tool for Basque", *4th International Conference on Language Resources and Evaluations (LREC)*, Lisbon.
- ARETA, N.; GURRUTXAGA, A.; LETURIA, I.; POLIN, Z.; SAIZ, R.; ALEGRIA, I.; ARTOLA, X.; DIAZ DE ILARRAZA, A.; EZEIZA, N.; SOLOGAISTOA, A.; VALVERDE, A. (2006): "Structure, Annotation and Tools in the Basque ZT Corpus", *International Conference on Language Resources and Evaluations (LREC 2006)*, Genoa.
- ARREGI, X.; ARRUARTE, A.; ARTOLA, X.; LERSUNDI, M.; SANTANDER, G.; UMBELLINA, J. (2010): "TZOS: Terminologia Zerbitzurako Online Sistema", in ALBERDI, X. & SALABURU, P. (argit.): *Ugarteburu Terminologia Jardunaldiak: euskararen garapena esparru akademikoetan*, Bilbo: UPV/EHUko Argitalpen Zerbitzua, 136-153 or.
- BIBER, D. (1988): *Variation across speech and writing*, Cambridge: Cambridge University Press.
- BIBER, D. (1994): "Using register-diversified corpora for general language studies", in ARMSTRON, S (argit.): *Using large corpora*, Cambridge: The MIT Press, 180-201 or.
- BIBER, D. (1995): *Dimensions of register variation. A cross-linguistic comparison*, New York: Cambridge University Press.
- BIBER, D. (2006): *University Language. A corpus based study of spoken and written registers*, Amsterdam: John Benjamins Publishing Company.
- BIBER, D.; CONRAD, S; REPPEN, R. (1998): *Corpus Linguistics. Investigating language Structure and Use*, Cambridge: Cambridge University Press.
- BOULANGER, J. C. (1991): "Une lecture socio culturelle de la terminologie", *Cahiers de linguistique sociale*, 18, 13-30 or.
- CABRÉ, M. T. (1999): "Elementos para una teoría de la terminología: hacia un paradigma alternativo", in CABRÉ, M. T. (argit.): *La terminología. Representación y comunicación*, Barcelona: Institut Universitari de Lingüística Aplicada. Universitat Pompeu Fabra, 69-92 or.
- CABRÉ, M. T. (2001): "Sumario de principios que Configuran la nueva propuesta teórica", in CABRÉ, M. T. & FELIU, J. (argit.): *La terminología científico-técnica*, Barcelona: Institut Universitari de Lingüística Aplicada. Universitat Pompeu Fabra.
- CABRÉ, M. T.; GOMEZ DE ENTERRÍA, J. (2006): *La enseñanza de los lenguajes de especialidad. La simulación global*, Madrid: Gredos.
- CIAPUSCIO, G. (2003): *Textos especializados y terminología*, Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.

- CLEAR, J. (1992): "Corpus sampling", in LEITNER, G. (argit.): *New directions in English language corpora*, Berlin: Mouton de Gruyter, 21-31 or.
- DUARTE, C. (1990): "El llenguatge administratiu i jurídic. Els llenguatges d'especialitat", *Revista de Llengua i Dret*, 14, 93-109 or.
- ELORDUI, A.; ZABALA, I. (2005): "Terminological Variation in Basque: Analysis of Texts of Different Degrees of Specialization", *Sky Journal of Linguistics*, 18, 71-92.
- FERGUSON, Ch. (1994): "Dialect, Register, and Genre: Working Assumptions About Conventionalization", in BIBER, D. & FINEGAN, E. (argit.): *Sociolinguistic Perspectives on register*, NY, Oxford: Oxford University Press, 15-30 or.
- FLOWERDEW, L. (2004): "The argument for using English specialized corpora to understand academic and professional language", in COMMOR, U. eta UPTON, Th. A. (argit.): *Discourse in the Professions*, Amsterdam/Philadelphia: John Benjamins Publishing Company, 11-33 or.
- GUESPIN, L. (1993): "Normaliser ou standardiser?", in GAMBIER, Y. & GAUDIN, F. (argit.): *Le Langage et l'homme*, Vol. XXVIII, 4, De Boeck Université, 213-222 or.
- HUDSON, R. A. (1996): *Sociolinguistics*, Cambridge: Cambridge University Press.
- LAMUELA, X. (1983): "Enquadrament sociolinguistic del llenguatge jurídic-administratiu", *Revista de Llengua i Dret*, 3.
- Mc ENERY, T. eta HARDIE, A. (2011): *Corpus Linguistics*, Cambridge: Cambridge University Press.
- PARODI, G. (2005): *Discurso especializado e instituciones formadoras*, Valparaiso: Ediciones Universitarias de Valparaiso, Pontificia Universidad Católica de Valparaiso.
- PARODI, G. (2007): "El discurso especializado escrito en el ámbito universitario y profesional: Constitución de un corpus de estudio", *Revista Signos*, 40, 63, 147-178 or.
- SINCLAIR, J. (1982): "Reflections on computer corpora in English language research", in JOHANSSON, S. (argit.): *Computer Corpora in English Language Research*, Bergen: Norwegian Computing Centre for the Humanities, 1-6 or.
- SINCLAIR, J. (2005): "Corpus and Text - Basic Principles", in WYNNE, M. (argit.): *Developing Linguistic corpora: a Guide to Good Practices*, Oxford: Oxbow Books, 1-16 or.
- [Interneteko helbidea <http://ahds.ac.uk/linguistic-corpora>; 2012-07-11an irakurria]
- SWALES, J. (1990): *Genre analysis. English in academic and research setting*, Cambridge: Cambridge University Press.
- SWALES, J. (1996): "Occluded genres in the academy", in VENTOLA, E. & MAURANEN, A. (argit.): *Academic Writing*, Amsterdam: John Benjamins, 45-58 or.
- SWALES, J. (2001): "Metatalk in American academic talk", *Journal of English Linguistics*, 29: 34-54 or.
- SWALES, J. (2004): *Research Genres: exploration and Applications*, Cambridge: Cambridge University Press.
- TEMMERMAN, R. (2000): *Towards New Ways of Terminology Description. The sociocognitive Approach*, Amsterdam/Philadelphia: John Benjamins.
- ZABALA, I. eta ELORDUI, A. (2006): "Specialized Discourse and the Linguistic System of Basque", in FERNANDEZ, B. & LAKA, I. (argit.): *Andolin gogoan. Essays in honour of Professor Eguzkitza*, Bilbo: UPV/EHUko Argitalpen Zerbitzua, 895- 914 or.
- ZABALA, I.; AIERBE, A.; ALDEZABAL, I.; ARREGI, X.; ARRIOLA, J. M.; ELORDUI, A.; ELOSEGI, A.; ELOSEGI, K.; EZEIZA, J.; GARCIA, J.; GARCIA, I.; LERSUNDI, M.; SAN MARTIN, I.; UGARTEBURU, I. (2008): "GARATERM: Diskurtso akademiko-profesionalaren didaktika eta garapena uztartzeko tresna informatikoen diseinu eta integrazioa helburu duen

proiektua”, in SALABURU, P. & UGARTEBURU, I. (argit.): *Espezialitate Hizkerak eta Terminologia III: Espezialitate-hizkeren didaktika eta komunikazioa*, Leioa: UPV/EHUKo Argitalpen Zerbitzua: 211-219 or.

ZABALA, I.; SAN MARTIN, I.; LERSUNDI, M.; ELORDUI, A. (2011): “Graduate Teaching of Specialized Registers in a Language in the Normalization Process: Towards a comprehensive and Interdisciplinary Treatment of Academic Basque”, in MARUENDA-BATALLER, S. & CLAVEL-ARROITIA, B. (argit.): *Multiple Voices in Academic and Professional Discourse: Current Issues in Specialised Language Research, Teaching and New Technologies*. Newcastle: Cambridge Scholars Publishing, 208-218 or.

ZABALA, I.; SAN MARTIN, I.; LERSUNDI, M. (2013): “Linguistic and sociolinguistic factors that influence the detection, implantation and circulation of natural terminology in academic uses of Basque”, in DURY, P.; DE HOYOS, J.C.; MAKRI-MOREL, J.; MANIEZ, F.; RENNER, V. & VILLAR DÍAZ, M. B. (argit.): *La néologie en langue de spécialité. Neology in specialized languages. La neología en lengua de especialidad*. Lyon: Publications du CRTT.

[AURKIBIDEA](#)