

Terminologia lantzeko baliabideak EHU

IZASKUN ALDEZABAL, MARIA JESUS ARANZABE, ARANTZA DIAZ DE ILARRAZA, IGONE ZABALA
(UPV/EHU: Ixa Taldea)

Euskarazko erregistro espezializatuak eta, bereziki, terminologia unibertsitatean garatzen dira, neurri handi batean. Izan ere, unibertsitateko klaseetan, klase horietan erabiltzen diren materialetan, gradu edo master amaierako lanetan eta doktore-tesietan, denetariko gai espezializatuaz aritzen dira euskaldunak, ahoz eta idatziz. Gainera, unibertsitatean prestatzen dira etorkizuneko profesionalak, eta, lan-munduan sartuko direnean, unibertsitate-ikasketetan ikasitako terminologia erabiliko dute, neurri handi batean. Nolanahi ere, unibertsitateko komunikazio-egoera espezializatuetan erabiltzen diren testuak eta terminologia ikusezin gertatzen dira maiz deskripzio-ekimenetarako, gehienetan behin-behineko testu idatziak zein ahozkoak erabiltzen baitira irakasleen eta ikasleen arteko komunikazioa bideratzeko. Hain zuzen ere, unibertsitateko komunikazio-egoera espezializatu horietan erabiltzen diren testuak eta terminologia ikusgai egiteko xedeaz sortu zen Terminologia Sareak Ehunduz programa (TSE).

TSE programak 2008an hasi zuen bere ibilbidea, UPV/EHUko Euskal Hizkuntza eta Komunikazioa Saileko zenbait irakasleren eskutik eta Euskara Errektoreordetzaren babespean. Orduz geroztik, UPV/EHUko zenbait gradutako irakasgaietan erabiltzen diren testuak bildu, testu horietan erabiltzen den terminologia erauzi, eta terminologia-hiztegi eleaniztunak osatzen ari gara, programari esker. Uste dugu hiztegi horiek, euskararen normalizazio bidean, espezialitate-alorretako erregistroak garatzen eta finkatzen lagunduko dutela. Testuinguru akademikoetan ez ezik, handik kanpoko jardueretan aritzen diren askotariko erabiltzaileen-tzat baliagarri izateko pentsatuta daude: itzultzaileak, interpreteak, lexikografoak eta hiztegi-argitaratzaileak, lexikoan interesa duten filologoak eta hizkuntzalariak, eta erregistro espezializatuak helburu dituzten hizkuntza-irakasleak, besteak beste.

TSE programa terminologiaren azterketarako hurbilketa berriek ezarri dituzten oinarri teoriko eta metodologikoetan sustraiturik dago, hau da, Wüster-en teoria klasikoari (1979)

egindako kritiketatik goeko hamarkadan sortu diren hiru ildo teoriko nagusietan: sozioterminologia (Gaudin 1993; Boulanger 1991), teoria soziokognitiboa (Temmerman 2000) eta terminologiaren teoria komunikatiboa (Cabr e 1999, 2001). Ildo horiei jarraituz, TSE programan, UPV/EHUko aditu-irakasleek erabiltzen duten terminologia erreala zer-nolakoa den, nola sortzen den eta nola aldatzen den aztertuko da; hori guztia, betiere, irakasle horiek sortu edo erabiltzen dituzten idatzizko testuetan oinarrituta.

Artikulu honetan, TSE programan erabiltzen eta osatzen ari garen baliabideak aurkeztuko ditugu, laburki, erabiltzen dugun metodologian kokatuta: Garaterm corpora, Erauzterm termino-erauzle erdiautomatikoa eta TZOS Terminologia Zerbitzurako Online Sistema. Baliabide horiek diziplinartekotasunean eta modu sinergikoan lan eginda garatu edota egokitu ditugu, alorrez alorreko aditu, hizkuntzalari eta informatikariak elkarlanean arituta.

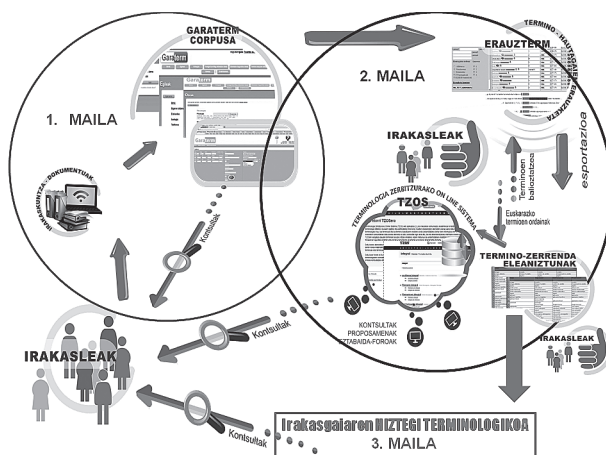
TSE programaren metodologia eta baliabideak

TSE programan, terminologia-lana egiteko abiapuntua EHUko gradu eta graduondokoen euskarazko irakaskuntzan erabiltzen diren testuak dira; zehazki, irakasgaiak emateko erabiltzen diren testuak: irakasgaiaren programa, ariketak, laborategiko praktikak egiteko jarraibideak, eskoletan erabiltzen dituzten aurkezpenen diapositibak, azterketak... Irakasgaia aukeratu dugu lan-unitate gisa, komunikazio-unitate akademikotzat har daitekeen heinean. Helburua, berriz, terminologia erreala deskribatzea da; alegia, aditu-irakasleek komunikazio akademikoan erabiltzen duten terminologia. Horrenbestez, balio espezializatudun hizkuntza-unitateak (eta, haien barruan, terminoak) testuetatik abiatuta biltzen eta aztertzen dira.

TSE programako parte-hartzaileei hiru konpromiso-maila eskaintzen zaizkie, eta urratsez urrats egin ditzakete (ikus 1. irudia). Lehen mailan, irakasgai bateko materiala igotzen dute horretarako diseinatu den Garaterm lan-ingurunea (Zabala *et al.*, 2013). Lan-ingurune hori orotariko corpusetatik ahalik eta informazio linguistikoki gehien erauzi ahal izateko eta testu bakoitzaren gainean ahalik eta informazio xeheena biltzeko diseinatu da. Informazio hori metadatuen bitartez jasotzen da, eta corpusak atzean duen datu-basea antolatzeke erabiltzen da. Lan-ingurune horretara igotako testuak linguistikoki prozesatu ondoren, kontsultagai jartzen dira Garaterm corpusean. Garaterm corpusean, testuen jakintza-alorra ez ezik, linguistikoki kontrolatutako (zuzendutako) testuak eta kontrolatu gabeko testuak (espontaneoak) bereizten dira, irakasleek espontaneoki erabiltzen dituzten hizkuntza-baliabideak zuzentzaileen interbenzioaren ondorio diren erabileretatik bereizi ahal izateko xedez. Horretaz gain, irakasleek berek ekoiztiko testu eta itzulpenak ere bereizten dira.

Bigarren mailan, linguistikoki prozesatutako testuetatik termino-hautagaiak erauzten dira, Erauzterm termino-erauzlearen bitartez (Alegria *et al.*, 2004). Erauztermek (Alegria *et al.*, 2005) hizkuntza-ingeniaritza, metodo estatistikokoak eta teknologia informatiko modernoak integratzen ditu. Haiei esker, termino izan daitezkeen elementu bakun eta konplexuen zerrenda proposatzen du, bakoitzari buruzko informazio estatistikoa ematen du, eta elkarren aldakiak

(ortotipografikoak, morfosintaktikoak edota sintaktikoak) izan daitezkeenak erlazionatzen ditu. Termino-erazuleak erdiautomatikoak direnez, aditu edo terminologoen parte-hartzea behar da proposatzen dituzten hautagaiak benetan terminotzat har daitezkeela egiaztatzeko. Hortaz, TSE programan parte hartzen duten irakasleek erazutako termino-hautagaiak balioztatzen dituzte, eta, ondoren, esportazioa egiten dugu. Lortutako termino-zerrendak gainbegiratu, sinonimia- eta polisemia-kasuak identifikatu, eta euskarazko terminoen beste hizkuntzetako ordainak ematen dituzte aditu-irakasleek, beren espezialitate-alorrean beste hizkuntza batzuetan duten esperientzia linguistikoa baliatuz.



1. irudia: TSE programa

Terminologiaren deskripzioa testuak sortu eta erabiltzen dituzten aditu-irakasleek egiten dutenez, deskripzio aktiboa esaten diogu. Izan ere, terminologiaren deskripzioa testu akademikoaren egileek berek egiten dutenez, deskribatutako terminologiari eragiteko zilegitasuna ere badute. Kontuan hartu behar da, gainera, TSE programa formazio linguistikorako programa modura diseinatu dela eta irakasleek erabiltzen duten terminologiaren kalitate linguistikoa ere eragin nahi duela.

Azpimarratu beharrekoa da TSE programan aintzat hartzen dugula terminoak aldakorrek direla ondo garatutako hizkuntzetan ere. Izan ere, terminologia erreala aldakorra da diskurtsoaren baldintzetara eta komunikazio-egoeretara egokitzeko. Ondorioz, unibertsitateko testuetan aurkitzen diren aldaki eta sinonimo gehienak deskribatzen dira TSE programako termino-zerrendetan. Normalizazio-bidean dauden hizkuntzetan, espero izatekoa da aldakortasun asistematikoa handiagoa egotea hizkuntza normalizatuetan baino, baina azterketa enpirikoetan ikus daiteke badagoela komunikazio-egoeretara egokitzeko aldakortasun funtzionala ere (Elordui eta Zabala 2005). Hortaz, aldakortasunaren deskripzioaren helburu nagusia aldaki asistematikoa aldaki funtzionaletatik bereizteko informazio nahikoa izatea da. Izan ere, hizkuntza

normalizatuen egoerara hurbilduz joateko, funtsezkoa da aldaki asistematikoak harmonizatzea, aldakortasun funtzionala elikatzeari begira.

Behin terminologia deskribatuta, Terminologia Zerbitzurako Online Sistema (TZOS) datu-basean jartzen da kontsultagai. Unibertsitatean erabiltzen den terminologiaren deskripzio horren helburua aldakiak harmonizatzea da, eta ez aldakietako bat edo batzuk estandarizatzea. Terminologia denboran zehar ere aldatzen da, areago euskaraz, normalizazio bidean dagoen hizkuntza bat den heinean. Hortaz, TSE programaren oinarrietako bat hau da: ikuspegi diakronikotik eta dinamikotik abiatu behar dugu terminologia harmonizatzeke. TZOSen jasotzen den terminologia testuetatik erauzia denez, sarrera bakoitzean ikus daitezke iturri-testuen jakintza-alorra eta azpialorra, ikastegia, titulazioa eta irakasgaia, eta testuen egilea. Horretaz gain, TZOSek aukera ematen du euskarazko sarreretan klik eginez Garaterm corpusean ikusteko zer testuingurutan erabili den terminoa.

Bukatzeko, TSE programaren hirugarren mailaren helburua deskribatutako terminologiaren alderdi semantikoetan sakontzea da. Terminoen semantika lantzeko, bi baliabide erabiltzen ditugu: etiketa semantikoak eta defnizioak. Etiketa semantikoen bidez, irakasgai bateko terminologia kategoria semantikoen arabera sailkatzen da, eta, defnizioen bitartez, aldiz, terminoek irakasgai bakoitzean hartzen duten berriarazko esanahia uzten da agerian.

Emaitzak

2008az geroztik, askotariko emaitzak lortu dira. Corpusaren tamainari dagokionez, 15.409.839 hitzeko corpus bat osatu da, eta 300en bat irakaslek hartu dute parte. 155 hiztegi eleaniztun inguru osatu dira, eta 29.196 sarrera daude jadanik kontsultagai TZOSen. Hizkuntzaz hizkuntza kargatutako denominazioen kopuruak, aldiz, zerrenda honetan ikus daitezke:

- euskara: 33.784
- gaztelania: 31.105
- ingelesa: 23.713
- frantsesa: 11.713
- latina: 823

Oraindik ere alde handia dago espezialitate-alor batzuetatik beste batzuetara bildutako testuei eta terminologiari dagokienez. Hortaz, oraingo xedea biltzen ari garen baliabideen estaldura eta alorrez alorreko oreka pixkana-pixkana handitzea da.

BIBLIOGRAFIA

- Alegria, Iñaki; Gurrutxaga, Antton; Lizaso, Pili; Saralegi, Xabier; Ugartetxea, Sahats & Urizar, Ruben (2004). «An XML-Based Term Extraction Tool for Basque», *4th International Conference on Language Resource and Evaluations (LREC)*. Lisboa.
- Gurrutxaga, Antton; Saralegi, Xabier; Ugartetxea, Sahats & Alegria, Iñaki (2005). «Erauzterm: euskarazko terminoak erauzteko tresna erdiautomatikoa», *Euskera zientifiko-teknikoa: normalizaziotik homologazinora, IX. Jardunaldiak*, Mendebalde Kultura Alkartea, 371.-411. or.
- Boulanger, Jean-Claude (1991). «Une lecture socio culturelle de la terminologie», *Cahiers de linguistique sociale*, 18, 13.-30. or.
- Cabré, María Teresa (1999). «Elementos para una teoría de la terminología: hacia un paradigma alternativo», María Teresa Cabré, (arg.): *La terminología. Representación y comunicación*, Bartzelona: Institut Universitari de Lingüística Aplicada. Pompeu Fabra Unibertsitatea, 69.-92. or.
- (2001). «Sumario de principios que configuran la nueva propuesta teórica», María Teresa Cabré & Judit Feliu (arg.): *La terminología científico-técnica*, Bartzelona: Institut Universitari de Lingüística Aplicada, Pompeu Fabra Unibertsitatea.
- Elordui, Agurtzane & Zabala, Igone (2005). «Terminological Variation in Basque: Analysis of Texts of Different Degrees of Specialization», *SKY Journal of Linguistics* 18, 71.-91. or.
- Gaudin, Francois (1993). *Socioterminologie: des problèmes sémantiques aux pratiques institutionnelles*, Rouen: Université de Rouen, Publications de l'Université de Rouen, 182.
- Temmerman, Rita (2000). *Towards New Ways of Terminology Description. The sociocognitive Approach*, Amsterdam/Philadelphia: John Benjamins.
- Wüster, E (1998 [1979]): *Introducción a la teoría general de la terminología y a la lexicografía terminológica*, María Teresa Cabré (arg.), Bartzelona: Institut Universitari de Lingüística Aplicada, Pompeu Fabra Unibertsitatea. [itzulpena. Jatorrizko liburua: Wüster E, (1979) *Einführung in die Allgemeine Terminologielehre und Terminologische Lexikographie*. 2 Vol. Viena].
- Zabala, Igone; Lersundi, Mikel; Leturia, Igor; Manterola, Iker & Santander, Gotzon (2013). «GARATERM: euskararen erregistro akademikoen garapenaren ikerketarako lan-ingurunea», Xabier Alberdi & Pello Salaburu (arg.), *Ugarteburu Terminologia Jardunaldiak (V). Terminologia naturala eta terminologia planifikatua euskararen normalizazioari begira*. Bilbo: UPV/EHUren Argitalpen Zerbitzua, 98.-114. or.

Resumen

Los registros especializados, y especialmente la terminología, se desarrollan en la universidad. En ella, los hablantes de euskera tratan todo tipo de temas especializados en diversas situaciones comunicativas. Sin embargo, los textos y la terminología empleados en dichas comunicaciones especializadas no tienen visibilidad en las iniciativas descriptivas, y el programa Tejiendo las Redes de Terminología (Terminologia Sareak Ehunduz o TSE por su sigla en euskera) nació en el año 2008 precisamente con el objetivo de visibilizarlos. El presente artículo resume, a partir de la metodología empleada en el programa, los recursos que se crean y se emplean en la actualidad: el entorno de trabajo y corpus Garaterm; Erauzterm, la herramienta semiautomática de extracción terminológica; y TZOS, el sistema online de servicio terminológico.

Résumé

C'est à l'Université du Pays Basque que l'on développe les registres spécialisés concernant la langue basque, en particulier ceux ayant trait à la terminologie. Les bascophones utilisent l'euskara pour traiter de tout genre de sujet spécialisé dans des situations de communication tant orale qu'écrite. Toutefois, les textes et la terminologie employés dans ces communications spécialisées n'apparaissent pas souvent dans les activités descriptives. Aussi, le programme *TSE*, *Terminologia Sareak Ehunduz* (Tissant des Réseaux de Terminologie) est-il né en 2008, dans le but précisément de les rendre visibles. Dans cet article, nous présentons, à partir de la méthodologie appliquée dans le programme TSE, le résumé des moyens créés et employés actuellement dans ledit programme, à savoir, le corpus et le travail réalisé autour de GARATERM, ERAUZTERM, l'outil semi-automatique d'extraction terminologique et TZOS, système online de service terminologique.

Abstract

The development of specialized registers of Basque, and especially of terminology, takes place, to a large extent, in university contexts. Basque speakers work on a great variety of specialized topics in every kind of oral and written communication situation. However, the texts and terminology used in such situations often remain "hidden" for descriptive tasks. The Weaving Terminology Networks (Terminologia Sareak Ehunduz, TSE) program was created in 2008 with the aim of making these hidden language usages visible. In this article, we briefly describe the resources we are currently using and developing within the framework of our established methodology: the GARATERM work-environment and corpus, the semiautomatic term extractor Erauzterm, and the Online System for Terminology Service (Terminologia Zerbitzurako Online Sistema, TZOS).