

Euskararen erregistro akademikoen garapenaz: hiztegia eta fraseologia¹

Igone Zabala Unzalu*

Zientzia eta Teknologia Fakultatea, Euskal Herriko Unibertsitatea (UPV/EHU)
HiTZ Hizkuntza Teknologiako Euskal Zentroa - Ixa ikerketa-taldea (UPV/EHU)

Laburpena: Erabilera-alor akademikoak funtsezkoak izan dira euskararen biziberritze prozesuan baina, alor horien konkistak hizkuntza-baliabideen garapen eta berregituratze handia eragin badu ere, esparru akademikoetan erabiltzen diren hizkuntza-baliabideek eta horien atzean dauden motibazio funtzionalek oso neurri txikian erakarri dute hizkuntzalarien arreta. Artikulu honetan, erregistro akademikoen garapenean gako diren elementu batzuei erreparatu diegu, bereziki hiztegi eta fraseologia akademikoari. Bestalde, diskurtso akademikoen garapenerako funtsezkoak diren nominalizazio-estrategiez eta gertaera-izenak sortzen dituzten atzizkien banaketa morfopragmatikoaz ere aritu gara.

Hitz gakoak: euskararen erregistro akademikoak, hiztegi akademikoa, fraseologia akademikoa, nominalizazio-estrategiak, atzizki eratorleen banaketa morfopragmatikoa.

1. SARRERA

Azken hamarkadetan aurrera eraman den euskararen biziberritze prozesua arrakastatsua izan dela esateko hainbat arrazoi daude. Izan ere, euskaldunen kopurua nabariki igo da, eta batez ere gazteen artean gertatu da igoera. 2016ko VI. Inkesta soziolinguistikoa egin zenean, 16-44 urteko EAEko biztanleen artean % 73,2 ziren euskaldunak. Gainera, zenbait iker-

* Igone Zabala. Euskal Hizkuntza eta Komunikazioa Saila. Zientzia eta Teknologia Fakultatea (UPV/EHU). Sarriena auzoa, z/g (48940 Leioa). igone.zabala@ehu.eus. <https://orcid.org/0000-0002-1931-4136>.

¹ Lan honek Ministerio de Ciencia e Investigación erakundearen diru-laguntza jaso du HARTAvas ikerketa-proiektuaren barruan (PID2019-109683GB-C22). Coruñaiko Unibertsitateko eta UPV/EHUko taldeek aurrera daramagun HARTAvas (*Herramienta de Ayuda a la Redacción de Textos Académicos en Español y en Vasco*) proiektu koordinatuaren azpiproiektua da HARTAvas. A motako Talde Finkatuaren (Eusko Jaurlaritza: IT-1570-22) ikerlanen barruan ere kokaturik dago.



ketaren arabera (Amorrortu *et al.* 2017), hiztun gazte horietatik % 67 hiztun berriak dira, alegia, euskara etxetik kanpo ikasi dutenak. Zalantza handirik ez dago ezagutzaren igoera horren arrazoietakoa bat behintzat hezkuntzan datzala.

Bestalde, euskara batuaren garapenari esker, euskaldunok hizketa-komunitate bat (*speech community*) osatzen dugu, Hymes-en (1974) eta Labov-en (1974) zentzuan. Ez hori bakarrik, euskararen erabilera-eremuak ere nabariki hedatu dira eta, hedapena posible izateko, eta bai hedapenaren beraren ondorioz ere, erabilera-eremu berrietan aritzeko hizkuntza-baliabideak ere garatu ditu euskarak: eremu-konkista edota elaborazioa (Laurén *et al.* 2002). Euskararen erabilera berri horietako batzuk, adibidez, komunikabideak eta literatura, orokortzat har daitezke, testuek denetariko gaiak jorratzen baitituzte eta hiztunen komunitate osoari zuzenduta baitaude, baina beste erabilera-alor batzuk espezializatutzat hartu behar dira, testuek jorratzen dituzten gaiak espezializazio maila askotakoak izaten direlako, eta solaskideak ere espezialitate-alorrez alorreko adituak izaten direlako: Fisika, Biologia, Medikuntza, Filosofia, Historia, Hizkuntzalaritza, Zuzenbidea...

Komunikazio-alor espezializatuetan aritzeko behar diren hizkuntza-baliabideak, besteak beste, terminologia eta fraseologia, komunikazio-egoera akademikoetan garatzen dira eta, bereziki, unibertsitateetan. Euskal Herriko unibertsitateetan espezialitate-alor guztietako eskolak ematen dira euskaraz eta, horretarako, nahitaezkoa da hiztegi eta fraseologia akademikoa erabiltzea, baita kontzeptu espezializatuak adierazteko terminologia garatzea eta erabiltzea ere. Maila indibidualean, ikasleek unibertsitatean ikasten dituzte beren alorreko adituen komunitateko kide bihurtzeko behar dituzten erregistro akademikoak (Biber 2006): inolako hizkuntzatan ez dago erregistro akademikoen hiztun natiborik. Hori dela eta, unibertsitateko ikasleei ziklo bat bukatzen duten bakoitzean, lan akademiko bat idazteko eta ahoz defendatzeko eskatzen zaie, adituen komunitateko kide bihurtzeko ezagutzak eta komunikazio-gaitasunak garatu dituztela frogatu dezaten: gradu amaierako lana (GrAL), master amaierako lana (MAL) edota doktoretza-tesia.

Nolanahi ere, eremu akademikoen konkistaz eta elaborazioaz ari garenean, indibiduo maila ez ezik, komunitate maila ere hartu behar dugu kontuan (Laurén *et al.* 2002): euskararen erregistro akademikoen garapena alorrez alorreko diskurtso-komunitate espezializatuetan gertatu da eta gertatzen ari da, denboran luzatzen den prozesua baita. Erregistroen garapenean bi motatako prozesuak edo indarrak bereiz daitezke, Cabré-k (2003) terminologiaren normalizazioan bereizten dituenak bestelako hizkuntza-baliabideetara ere zabalduta. Batetik, diskurtso-komunitate espezializatu barruko autorregulazioa dago eta, bestetik, erakunde arauemaileen interbentzioa. Gainera, diskurtso-komunitateen barruko autorregulazioa bi modu-

tara egiten da: komunitateko kideek baliabide jakin batzuk erabiltzeko ados jartzen direlako edo/eta komunitateko kideek egiten duten erabilera dela eta baliabide batzuk finkatzen doazelako. Gure iritziz, funtsezkoa da erakunde arauemaileek diskurtso-komunitate akademikoen barruan gertatzen den garapen eta autorregulazioa kontuan hartzea, bi helbururekin: erregistro akademikoen garapen eta autorregulazioa ez oztopatzeko (Zabala *et al.* 2014) eta, areago, garatutako baliabideen harmonizazioan laguntzeko (Zabala 2021). Erregistro akademikoen garapenerako oztopo izan litekeen interbentzio itsua sailhesteko, oso garrantzitsua da interbentzio hori ez oinarritzea beste erregistro batzuen azterketetatik ateratako ondorioetan. Izan ere, edozein hizkuntza oso desberdin gauzatzen da erabilera-eremu desberdinetan (Seoane & Biber 2021: 3).

...the differences in language use patterns across registers are so significant, in terms of both their quantitative scope and their functional importance, that any claim to describe general patterns of use for an entire language is highly suspect at best [...] it is interesting to note that the history of empirical register research is to a large extent coexistent with the history of empirical corpus-based research. In one respect, this relationship is not surprising: empirical register analysis are essentially not possible without corpus, because we need to have a collection of texts from a register —a corpus— in order to perform an empirical analysis of its linguistic characteristics. But the opposite relationship is not equally necessary: it is certainly possible to build a corpus and carry out corpus analyses that disregard register distinctions.

Euskararen erregistro akademikoak nola ari diren garatzen aztertzeko, nahitaezkoa dugu testu-corpus egokiak osatzea, alegia, egin nahi dugun deskripziorako erreferenteak diren testu-generoak biltzen dituzten eta erakunde arauemaileen interbentzioa ez ezik diskurtso-komunitate akademikoen barruko autorregulazioaren emaitzak ere islatzan dituztenak. Helburu horri begira hasi ginen osatzen Garaterm corpusa² 2010. urtean eta, batez ere, unibertsitateko irakaskuntza-materialak eta lan akademikoak (GrAL, MAL, doktoretza-tesiak) biltzen ari gara. *Garaterm* corpusaren azpicorpusetako bat da *HARTAeus* corpusa (Aranzabe *et. al.* 2022), Universidade da Coruña eta UPV/EHUko ikertzaile taldea aurrera eramaten ari garen *HARTAesvas* proiektu koordinatuaren barruan osatu dena. *HARTAeus* euskarazko corpusa³ espainolezko *HARTAnoveles* corpusarekiko (Villayandre

² Ikus Zabala *et al.* (2013) Garaterm corpusaren kontsulta-interfazeak helbide honetan daude: <http://garaterm-corpusa.ixa.eus/>.

³ *HARTAeus* corpusean kontsultak egiteko, *Garaterm* corpusaren kontsulta-interfaze aurreratuan «Non» sakatu eta «HARTAeus» corpuseko kontsultak iragaz daitezke. Kontsultak domeinu nagusien arabera, eta Garaterm corpusak eskaintzen dituen bestelako parametroen arabera ere iragaz daitezke.

2018) konparagarria izateko osatu da. 295 GrAL eta 104 MAL bildu dira: guztira 3.285.098 hitz. Bi hizkuntzetako corpusak konparagarriak izateko, lau hipereremutan banatu dira testuak: Arteak eta Humanitateak, Biologia eta Medikuntza, Zientzia Fisikoak eta Gizarte Zientziak. Hipereremu horien barruan, jakintza-alor xeheagoak ere bereizi dira. *HARTAeus* corpusak aukera ematen digu aztertzeke unibertsitateko ikasketak egin dituzten ikasleek, alegia, alorrez alorreko aditu berriek, unibertsitatean garatu dituzten ikerketa- eta komunikazio-gaitasunak egiaztatzeke idazten dituzten testuetan erabiltzen dituzten hizkuntza-baliabideak. Euskararen erregistro akademikoen garapenaren ikerketari ekiteko corpus egokia dela deritzogu, corpusetan oinarritutako erregistroen ikerketarako oinarrien arabera (Biber & Conrad 2009: 10):

There is not a correct level on which to identify a register. Rather, it depends on the goals of your study. You may want to characterize the register of academic prose, a very general register. Or you may be interested in only research articles, a more specific register within academic prose. Or you might focus on medical research articles, or even only the methods sections of experimental medical research articles. All of these can be considered registers, differing in their level of generality...

Erregistro akademikoetan garrantzi berezia eman ohi zaie alorrez alorreko terminologiaren garapen eta normalizazioari. Izan ere, terminologia funtsezkoa da kontzeptu espezializatuak adierazteko, eta testu espezializatuen bizkar-hezurra osatzen duela ere esan daiteke.⁴ Nolanahi ere, lan honetan euskararen kasuan gutxitan aztertu diren beste hizkuntza-baliabide batzuei erreparatuko diegu: lexiko akademikoari eta fraseologiari.

Hizkuntzaren azterketarako hurbilketa estrukturalistek, besteak beste gramatika sortzaileak, hizkuntza-osagaietan, eraketa-erregelatan, erregela horien errekurribotasunean eta hizkuntzaren sortzailetasunean jartzen dute arreta. Hori dela eta, hitz anitzeko segidak HAS (*multi-word sequences MWS*) bazterreko adierazpide bakantzat hartzen dituzte. Baina corpus-hizkuntzalaritzaren eskutik egun gero eta hedatuago eta onartuago dagoen ideia da HASak direla idatzizko eta ahozko diskurtsoen eraikuntza-blokeak. HASak dira finkatutako edo erdifinkatutako hitz-segida errepikariak (Siyanova-Chanturia 2015). Testu-corpusetan oinarritutako azterketa enpirikoek agerian utzi dute hizkuntza errepikakorragoa dela hizkuntzalaritzaren hurbilketa estrukturalistek eta sortzaileek aurreikusitakoa baino, eta HASak edozein hizkuntzarako hiltun natiboek ekoizten dituzten idatzizko eta ahozko diskurtsoen % 50tik gora direla (De Coxk *et al.* 1998). Estimazio horiek idatzizko diskurtsoak biltzen dituen corpus batean egiaztatu dituzte ikertzaile batzuek (Erman &

⁴ Euskararen terminologiaren garapenaz, ikus Zabala (2019).

Warren 2000), eta agerian utzi dute haien corpusean bildutako diskurtso horien % 52,3 HASak direla. Hortaz, esan daiteke hiztunok HAS errepikari ugari erabiltzen ditugula hizkuntza erabiltzen dugunean. Areago, egile batzuek diote testu akademikoak idazteko unitate fraseologiko erdiautomatikoak erabiltzen dituzten finkatutako diskurso-eskemak baliatzen ditugula (Paquot 2018). Hortaz, konbinazio lexiko akademikoak (KLA) helburu aplikatuko corpus-hizkuntzalaritzako lan askoren ikergaia izan dira. KLAak garatzeko eta finkatzeko, nahitaezkoa da erabilera partekatua egotea diskurtso-komunitate akademikoetan. *HARTAeus* corpora erabili dugu, hain zuzen ere, erregistro akademikoaren bereizgarriak diren KLAak euskarazko lan akademikoetan garatu eta finkatu ote diren egiaztatzeko. Izan ere, KLAak erregistro akademikoaren garapenaren adierazletzat har daitezke.

HARTAvas proiektuan, lexiko akademikoa eta konbinazio lexiko akademikoak (KLA) erauzteko, hiru modulu garatu dira *Erauzterm* termino-eraztean: lexiko akademikoa erauzteko modulua, kolokazioak erauzteko modulua eta diskurtso-formulak erauzteko eta diskurtso-funtzioa esleitzeko modulua (Aranzabe *et al.* 2022). KLAk testu akademikoetan maiztasun handiz azaltzen diren hitz (edo morfema) multzoak dira, semantikaren ikuspuntutik konposizionalak zein ez-konposizionalak izan daitezkeenak, eta diskurtso akademikoetan ohikoak diren funtzio erretorikoak estaltzen dituztenak: helburuak azaltzea, adibideak ematea, emaitzak alderatzea... Fraseologiak hitz anitzeko unitateak aztertzen ditu, eta unitate fraseologikotzat hartzeko erabiltzen dituen irizpideak dira konposizionaltasuna (gardentasunarekin nahastu behar ez dena)⁵ eta osagaien hautapen askea / baldintzatua (Mel'čuk 2015).

Nolanahi ere, euskara hizkuntza eranskaria da, eta ingeleseko eta beste hizkuntza flexibo batzuetako hitz anitzeko unitateetan deskribatu diren zenbait fenomeno hizkuntza eranskarietan morfema mailan gertatzen direla erakutsi dute hainbat lanek (Durrant 2013, Nam *et al.* 2016, Öksüz 2019): elementu jakin batzuen maiztasun handiko kookurrentzia, finkatutako elementu-multzoak (*bundles*) eta lexikoaren eta gramatikaren arteko uztardura. Izan ere, hizkuntza eranskarietan postposizioak eta menderagailuak maiz gauzatzen dira morfema lotu modura eta, beraz, *lexical bundle* esaten zaien konbinazioak ez ezik, *morphemic bundle* deritzenak ere hartu behar dira kontuan (Nam *et al.* 2016). HARTAvas proiektuan bi KLA mota aztertu ditugu: kolokazioak (*ondorioak atera*, *emaitzek erakutsi*, *lagin adierazgarri*, *emaitza esangarri*) eta diskurtso-formulak (*ondorio modura esan daiteke*). Diskurso-formula polilexikoak (*lexical bundles*) ez ezik (*lan ho-*

⁵ Erabat gardenak diren adierazpideak konposizionalak ere badira, baina alderantzizkoa ez da egia. Adibidez, gaztelaniazko *con respecto a* '-i dagokionez' erabat konposizionala da, baina *respectar* aditzaren esanahia ez dakiten hiztunek ezin dute ondorioztatu adierazpidearen esanahia, ikasten ez badute (Alonso-Ramos *et al.* 2017).

nen helburua da, oro har, labur esanda), diskurtso-formula monolexiko polimorfemikoak ere hartu ditugu kontuan: *laburbilduz*, *-i dagokionez*, *-z gain*).

Artikulu honetan, *HARTA*eus corpusetik lortutako datu batzuk eskainiko ditugu. Bigarren atalean lexiko akademikoaz eta kolokazioez arituko gara. Hirugarren atalean, diskurtso-formula akademikoak izango ditugu hizpide. Laugarren atalean, nominalizazioaz eta gertaera-izen eratorrien banaketa morfopragmatikoaz arituko gara. Azken atalean ondorioak laburbilduko ditugu.

2. LEXIKO AKADEMIKOA ETA KOLOKAZIOAK

Lexiko akademikoaren zerrendek testu akademikoen hiztegi-elementu bereizgarriak biltzen dituzte: hizkuntza baten maiztasun handieneko lexikoan ageri ez diren edo testu akademikoetan maiztasun bereizgarriaz erabiltzen diren hiztegi-elementuak. Hortaz, lexiko-zerrenda akademikoak diziplina artekoak izaten dira, eta diziplina bakoitzaren berariazko terminoak alde batera uzten dituzte (Cohead 2000, Drouin 2007, Gardner & Davies 200; García Salido 2021). Hasiera batean, halako zerrendak biltzeko helburua izan zen hizkuntzen helburu akademikoetarako irakaskuntzarako ikastaroak eta materialak prestatzea, baina berrikiago, testu akademikoen idazketarako laguntza-tresnak elikatzeko ere osatzen ari dira horrelako zerrendak (besteak beste, Frankenberg *et al.* 2018, Granger & Paquot 2015, García Salido *et al.* 2018). *HARTA*vas proiektuaren azken helburua da testu akademikoen idazketarako laguntza-tresna elikatzeko baliabide lexikoak biltzea.⁶ Hortaz, lexiko akademikoa hiztegi-elementu akademikoen erabilerari buruzko datuak lortzeko, zenbait elementuren erabileran egon litezkeen arazoak detektatzeko eta kolokazio akademikoak identifikatzeko erabiltzen ari gara.

Euskarazko lexiko akademikoa erauzteko, Elhuyarrek Internetetik bildutako *Dabilena* corpus orokorra (300.217.903 hitz) erabili da kontraste-corpus modura. Corpusaren lau hiperemuetan eta testuen gutxieneko portzentaje batean (testuen % 20tik gora) ageri diren hautagaiei corpusen arteko alderaketak egiteko erabiltzen den *log-likelihood* neurria aplikatu zaie. Ondoren, eskuz aukeratu dira lexiko akademikoaren zerrendarako errelevantzat hartu diren elementuak. Horretaz gain, *log-likelihood* negatiboa duten elementu

⁶ *HARTA*vas proiektuaren helburu xehea da euskarazko baliabideak biltzea, baina *HARTA*esvas proiektu koordinatuaren helburu orokorra da gaztelaniarako eta euskararako baliabideak kontsulta-tresna bakar batean integratzea. Une honetan oraindik garapenean badago ere, *HARTA/TAILA* tresna helbide honetan dago kontsultagai: <http://harta.ixa.eus/>.

batzuk ere berreskuratu dira, argi eta garbi elementu lexiko akademikoak direla ondorioztatu baita, eta *log-likelihood* negatiboa izateko arrazoren bat aurkitu baita. Adibide bat da *pauso* / #*pausu* / *urrats* hirukotea. *HARTAeus* corpuseko testuetan *pausu* lemak ia *pauso* lemak beste agerraldi ditu eta *log-likelihood* positiboa (1.176,65), baina agerraldi horiek guztiak ortografia-akatsak dira, ez baitira erabili ‘atseden’ adierazteko, baizik eta *urrats* lemaren sinonimo modura. *Dabilena* corpusean askoz ere agerraldi gutxiago ditu *pausu* lemak eta, horietako asko *pauso* edo *urrats* izan beharko balira ere, erdia edo Urruñako Pausu auzoari egiten diote erreferentzia. Hartara, *urrats* lemak *log-likelihood* negatiboa (-1,6289) badu ere, lexiko akademikoan kontuan hartzeko modukoa dela pentsatu dugu. Oro har, hiru lema horiekin adierazi nahi den kontzeptuak askoz ere agerraldi kopuru handiagoa du *HARTAeus* corpusean *Dabilena* corpusean baino, eta testu akademikoan elementu bereizgarria dela esan daiteke, beraz.

1. taula

HARTAeus eta *Dabilena* corpusetan *pauso*, *pausu* eta *urrats* lemen agerraldi kopuruen alderaketa

	HARTAeus (agerraldi kopurua / M)	Dabilena (agerraldi kopurua / M)
<i>pauso</i>	182,64	107,42
<i>pausu</i>	176,55	28,82
<i>urrats</i>	156,46	178,74
Guztira	515,65	314,98

Lehen hurbilketatzat hartu behar den prozesuan, 846 elementuko zerrenda osatu dugu: 365 izen, 246 aditz, 182 adjektibo eta 53 adverbio. *Log-likelihood* altuena duten lau kategorietako lehen 20 elementuak jaso dira 2. taulan, alegia, *Dabilena* corpus orokorreko erabileratik gehien urruntzen diren elementu lexikoak. Elementu horietako batzuk egun erabat ezagun eta ohikoak egiten bazaizkigu ere, 70eko hamarkadaren bukaera aldera euskara unibertsitatean sartu berria zenean hasi ziren erabiltzen eta promozionatzen testu akademikoetarako. Esate baterako, zenbait zientzialariren arteko hainbat gogoeta eta eztabaidaren ondoren, *garatu*, *garapen* eta *lagin* erabiltzea sustatu zuen Jose Ramon Etxebarriak 1979 urtean *Anaitasuna* aldizkarian idazten zuen zutabea (Txurruka 2020: 136-137).

2. taula*Log-likelihood* balio handieneko 20 lemak kategoriaren arabera

Izenak			Aditzak		
	HARTAeus F (M)	Dabilena F (M)		HARTAeus F (M)	Dabilena F (M)
irudi	3.570,37	484,32	erabili	1.582,30	353,06
taula	1.579,56	92,50	aztertu	769,54	115,65
lagin	862,99	56,05	lortu	1.154,91	293,47
kasu	2.552,44	769,94	kalkulatu	443,52	59,25
analisi	933,61	87,09	definitu	397,86	50,53
ekuzazio	517,79	17,13	burutu	330,89	40,67
emaitza	1.786,25	450,50	aipatu	739,70	241,41
sistema	1.981,68	591,21	ikusi	220,39	17,97
atal	1.552,16	402,40	garatu	521,45	138,97
erlazio	665,73	62,82	oinarritu	380,51	79,67
metodo	737,88	85,73	azaldu	601,50	190,01
parametro	418,86	25,47	identifikatu	195,12	20,62
ezaugarri	1.174,70	303,87	eragin	770,45	331,14
balio	1.384,13	415,57	agertu	300,75	66,11
prozesu	1.555,51	518,40	neurtu	192,99	27,16
mota	1.209,10	332,20	erlazionatu	245,05	47,13
modu	2.611,79	1.201,18	inplementatu	119,02	8,04
aldagai	465,74	42,09	konparatu	150,38	17,94
egitura	987,79	240,70	deskribatu	830,11	421,11
sekuentzia	336,37	19,57	sortu	136,37	15,32
Adjektiboak			Adberbioak		
	HARTAeus F (M)	Dabilena F (M)		HARTAeus F (M)	Dabilena F (M)
ezberdin	1.582,30	353,06	hala	1.538,16	548,08
altu	769,54	115,65	bezala	2.301,00	1.126,92
desberdin	1.154,91	293,47	zuzen	440,17	64,96
ageri	443,52	59,25	azkenik	622,51	163,80
baxu	397,86	50,53	jarraian	390,86	91,99
teoriko	330,89	40,67	egun	568,63	189,26
beharrezko	739,70	241,41	lehenik	297,10	59,46
maximo	220,39	17,97	behin	748,53	309,73
digital	521,45	138,97	orokorrean	292,53	58,87
posible	380,51	79,67	soilik	606,07	265,90
berdin	601,50	190,01	horrela	904,39	524,07
erlatibo	195,12	20,62	hain	591,76	331,26
egoki	770,45	331,14	honela	301,67	125,40
esanguratsu	300,75	66,11	guztiz	378,07	193,96
total	192,99	27,16	nahiko	311,71	146,84
automatiko	245,05	47,13	lehendabizi	94,67	20,26
konstante	119,02	8,04	ondoren	1.242,89	937,10
esperimental	150,38	17,94	automatikoki	105,32	33,23
hurongo	830,11	421,11	aurrerago	79,75	21,71
minimo	136,37	15,32	zehazki	1.75,95	85,50

Testu akademikoetan oso maiz ez dagokien esanahiarekin erabilia ikusten ditugun zenbait elementu ageri dira zerrendan: *pauso*-#*pauso*-*urrats*; *proba* / *froga*; *hazkunde* /#*hazkuntza*, *balio* / *balore*. Zerrendako beste elementu batzuk dira erabiltzaileei *a* itsatsiaren inguruko zalantzak sortzen dizkietenak: *garrantzi*, *emaitza*, *ordena*. Hortaz, lortutako zerrenda baliagarria da komunikazio akademikoari buruzko ikastaroetan eta material didaktikoetan erabiltzeko.

Gainera, lexiko akademikoaren zerrendatik abiatuta, kolokazio akademikoak lortu ditugu Gurrutxaga & Alegria-ren (2011) eta Gurrutxaga *et al.*-en (2018) neurriak integratuta dituen *Erauztermen* modulua erabiliz. Fraseologiaren tradizioaren arabera, kolokazioak dira harreman sintaktikoa duten hitz-bikoteen konbinazioak, modu bereizgarrian asimetrikoak direnak. Konbinazio asimetrikoak direla esan daiteke bikotekideetako baten (oinarriaren *base*) hautapena haren esanahiaren arabera era askean egiten bada ere bestea (kolokatuaren *collocate*) hautapena oinarriak murrizten duelako (Mel'čuk 1995, 1998). Konbinazioaren finkapenaren atzean dago erabilera-maiztasun handia, azkenean instituzionalizazioa eramaten duena, alegia, hiztun-komunitate baten norma bihurtzera.

Lexiko akademikoaren zerrendako elementu batzuek ez dute ematen kolokazio askorik, baina beste batzuk oso emankorrak dira. Asimetria horiek beste hizkuntza batzuetan ere gertatzen dira, eta esan dezakegu euskararako lortu dugun kolokazioen zerrenda konparagarria dela espainolerako lortu denarekiko (Aranzabe *et al.* 2022). *Emaitza* izenaren kolokazioak ikus daitezke 3. taulan.

Nolanahi ere, zenbaitetan diskurtso-komunitate akademikoetan erabile-ra-maiztasun handia dela eta instituzionalizatzen diren konbinazioak ez dira hiztegieta edo estilo-liburuetan kodifikatzen direnak. Adibidez, zenbait estilo-libururen eta hiztegiaren arabera, *urrats* izena *egin* aditzarekin kolokatzen da (*urratsak egin*). *Dabilena* corpus orokorrean ere, *urratsak egin* kolokazioa (8.735 agerraldi) *urratsak eman* kolokazioari (6.780 agerraldi) gailentzen zaio. *Pauso* izenaren kasuan, aldiz, *pausoak eman* (8.769 agerraldi) da gailentzen dena, eta *pausoak egin* konbinazioak ez du kolokaziotzat hartzeko langa gainditzen. *HARTAeus* corpusean, aldiz, *urratsak eman* eta *pausoak eman* konbinazioek baino ez dute gainditzen kolokaziotzat hartzeko langa. Aipatu beharrekoa da *pauso* izenak *jarraitu* eta *izan* aditzarekin ere ematen dituela kolokazioak *HARTAeus* corpusean eta, *urrats* izenak, aldiz, *eman* aditzarekin baino ez duela ematen kolokaziorik. *Dabilena* corpus orokorra askoz ere handiagoa da, eta askotariko kolokazioak ematen ditu *urrats* izenak corpus horretan, baita *jarraitu* / *jarraiki* aditzarekin ere. Erabileraren erabileraz finkatzen doazen joera idiosinkratikotzat har daitezke horiek, eta ez dirudi justifikatuta dagoenik testu akademikoetan *urratsak egin* kolokazioaren aldeko borroka sutsua egitea.

3. taula

*Emaitz*a izenaren kolokazioak *HART*Aeus corpusean,
maiztasunaren arabera antolatuta

Oinarria	Kolokatuak	
	Aditzak	
emaitza	ABS	izan, lortu, eman, aztertu, ikusi, hobetu, aurkeztu, erakutsi, azaldu, alderatu, jaso, konparatu, etorri, adierazi, analizatu, interpretatu, gainditu, ebaluatu, eskuratu, agertu, eskaini, hartu, baieztatu, behatu, erabili, kalkulatu, ulertu, kalkulatu, bildu, osatu, jaso
	ERG	erakutsi
	DAT	erreparatu, egokitu
	INES	ikusi, oinarritu, eragin, azaldu, islatu, baloratu, agertu
	SOZ	alderatu, konparatu
	ABL	abiatu, ondorioztatu
	Adjektiboak	
emaitza	on, hobe, orokor, positibo, egoki, zehatz, esperimental, negatibo, esanguratsu, adierazgarri, oker, analitiko, altu, global, interesgarri, nagusi, teoriko, ekonomiko, onargarri, hurbildu, baxu, kaxkar, argi, hobeago, txar, zuzen, fidagarri, garrantzitsu, kuantitatibo, erabilgarri, hoberen, fisiko	
	Izenak	
emaitza	analisi, interpretazio, azterketa, kartografia, fidagarritasun, konparaketa, zehaztasun, kalitate, deskribapen, lorpen, ondorio, estaldura	

Nolanahi ere, corpuseko datuak aztertuta, ikus daiteke ikasleen testuetan ohikoak diren zenbait kolokazio-desbideratze, corpusean kuantitatiboki gailentzen ez badira ere. Adibidez, *emaitzak*(ABS) *adierazi* (19 agerraldi) eta *emaitzek*(ERG) *erakutsi* (37 agerraldi) konbinazio zuzenek kolokaziotzat hartzeko neurriak gainditzen dituzte, baina ikasleen lanetan maiz ageri den eta okertzat hartu behar den *emaitzek*(ERG) *adierazi* (7 agerraldi) konbinazioak ez du gainditzen kolokaziotzat hartzeko langa. Nolanahi ere, euskara akademikoari buruzko ikastaroetan eta idazketarako laguntza-tresnetan kontuan hartu baharko lirateke halako desbideratzeak ere, erabilera okerrak zuzentzen laguntzeko. Izan ere, aditzen semantika lexikoaren desbideratzeak ere badira, eta agramatikaltzat ere har daitezke (Zabala 2003), baina hizkuntzalariak ez diren erabiltzaileentzat errazagoa da zuzenketaz jabetzea kolokazio-desbideratze modura azalduta desbideratze sintaktiko-semantiko modura azalduta baino.

3. FORMULA AKADEMIKOAK

Kolokazioek eta formulek zenbait ezaugarri komun dituzte, besteak beste, erabilera-maiztasuna eta elementuen hautapenaren murrizketak, baina desberdinak dira beste ezaugarri batzuei dagokienez. Hasteko, kolokazioak bi elementuko unitate fraseologikoak izaten dira, eta elementu horiek maiz ez dira jarraian agertzen diskurtsoan. Adibidez, *emaitza* izenak kolokazioa osatzen du *lortu* aditzarekin, nahiz eta batzuetan bi elementuak elkarrengandik urrun egon: ***emaitzak*** modu «*automatizatuago*» batean ***lortu*** nahi izan *ditugu*. Formulak, aldiz, diskurtsoaren zati jarraituak dira, eta askotan bi osagai baino gehiago dituzte. Formulak identifikatzeko, n-gramak (2-5 osagaikoak) erauzten dira, eta maiztasun-atalase bat gainditzen duten eta diskurtso-funtzio jakin batekin lot daitezkeen hitz-multzoak hautatzen dira. Corpus-hizkuntzalaritzaren tradizioan *lexical bundle* deritze hitz-multzo horiei, eta ohikoena izaten da atalasea izatea 10 agerraldi milioi bat hitzeko (Biber *et al.* 1999). HARTAVas proiektuan García-Salido *et al.*-en (2019) diskurtso-funtzioen sailkapena erabili da, Biber *et al.*-en (2004) eta Hyland-en (2008) sailkapenean oinarrituta dagoena, baina xeheagoa eta kontsulta-tresnaren erabiltzaileentzat ulerterrazagoa dena. Guztira 39 diskurtso-funtzio erabili dira hiru bloketan banatuta. Ikus 4. taula. Azpimarratu beharrekoa da diskurtso-funtzio guztietarako formulak erauzi direla ondo garatutako hizkuntzetarako erabilitako mugak erabiliz: 10 agerraldi milioi bat hitzeko. Beste hizkuntza batzuetan ohikoa den bezala, bi hitzeko formulak dira ugariak.

Erauketa-modu horrekin bat datozen formula batzuen adibideak eta esleitu zaien diskurso-funtzioa dira hurrengo hauek: *aipatu bezala*, *aurretik aipatu den bezala* (Diskursoaren beste atal batera bideratzea), *aipatu beharrekoa da*, *azpimarratzekoa da* (Zerbait azpimarratzea edo nabarmentzea), *beste modu batera esanda* (Birformulatzea), *horrek ez du esan nahi* (aurkaritza adieraztea), *zenbat eta* (Korrelazioa adieraztea), *lan honen helburu nagusia da* (Helburuak aurkeztea), *besteak beste* (Adibideak ematea).

Nolanahi ere, zenbait formula monolexiko plurimorfemiko ere identifikatu dira (42 formula). Hona hemen adibide batzuk: *laburbilduz* (Laburbiltzea), *ondorioz* (Ondorioa adieraztea), *-i dagokionez* (Gai bat sartzea), *bestalde* (Informazioa gehitzea), *askotan* (Maiztasuna adieraztea).

Kontsulta-tresnari eta irakaskuntza-materialei begira, bi atal ez-jarraitu uztartzea eskatzen duten formula batzuen kasuan, atal biak aipatzea komeni dela uste dugu, corpusean maiz desagokiro erabilita ikusten baitira: *alde batetik ...bestetik* /*#bestalde* (Ordenatzea/Informazioa gehitzea); *zenbat eta* /*#gero eta...* /*orduan eta* (Datuen korrelazioa / progresioa adieraztea). Ohikoak diren beste akats batzuk ere identifikatu dira, adibidez, ortografia-akatsak (*kontutan hartu behar da*, *gutxi gora behera*, *pausuak eman* -> *kontuan hartu behar da*, *gutxi gorabehera*, *pausoak eman*), izen-kohesioaren akatsak

(*honek ez du esan nahi* -> *horrek ez du esan nahi*), hitz-hurrekera desegokiak (*lan honen helburua [...] da* -> *lan honen helburua da ...*), desegokitakun semantikoak (*hau da* formula kausa adierazteko erabiltzea, birformula-ziorako erabili beharrean).

4. taula

HARTAEus corpusetik erauzitako diskurtso-formula akademikoei esleitutako diskurtso-funtzioen sailkapena

Testua egituratzea	Ikerketaren edukiari erreferentzia egitea	Iritzia ematea eta irakurleari zuzentzea
Informazioa gehitzea	Definitzea eta deskribatzea	Baieztapenak moteltzea
Konparatzea	Denominatzea	Beharra adieraztea
Mugatzea	Konparatzea	Ebaluazio bat adieraztea
Adibideak ematea	Taldeak ezartzea	Zerbait azpimarratzea edo nabarmentzea
Kausa adieraztea	Kantitatea adieraztea	Ziurtasuna adieraztea
Baldintza adieraztea	Datuen arteko korrelazioa adieraztea	Iturria aipatzea
Ondorioa adieraztea	Maiztasuna adieraztea	Posibilitatea adieraztea
Aurkaritza adieraztea	Datuen balioen progresioa adieraztea	
Kontzesioa adieraztea		
Lanari erreferentzia egitea	Denbora adieraztea	
Gai bat sartzea	Datuak aurkeztea	
Alternatiba bat aurkeztea	Ikergeraia aurkeztea	
Salbuespen bat sartzea	Hipotesiak aurkeztea	
Ordenatzea	Metodologia aurkeztea	
Diskurtsoaren beste atal batera bideratzea	Ondorioak aurkeztea	
Birformulatzea	Helburuak aurkeztea	
Laburbiltzea		

Unitate fraseologikoen ezaugarri bat da elementuen hautapen-askatasunaren murrizketa. Adibidez, *helburu* eta *xede* izenak erdisinonimoak badira ere, *helburu* izena daramaten 9 formula eta 15 kolokazio detektatu ditugu, baina ez dugu detektatu *xede* izena daraman inolako konbinazio lexiko akademikorik, *xede* lema corpusen 404 agerraldi baditu ere. Hortaz, lanaren helburuak azaldu behar direnan, *lan honen helburua da* bezalako formulak

askoz ere hedatuago daude erabileran, #*lan honen xedea da* segida baino eta, beraz, egokiagotzat har daitezke testu akademikoetarako.

4. NOMINALIZAZIOA ETA GERTAERA-IZENAK

Erregistro akademikoez ari garelarik, ezin dugu aipatu gabe utzi nominalizazioaren gaia. Izan ere, erregistro horien garapenean eragiteko helburua duten zenbait estilo-liburutan, nominalizazioa arazotzat hartzen da, erdara trakeski imitatuz «berezko euskal testuen» sistema bortxatzen duena (Alberdi & Sarasola 2001: 223-224; Garziak 2005: 86). Aholku horiek normalean ez dute kontuan hartzen nominalizaziorako joera gramatika-sistemearekin baino areago estiloarekin lotutako ezaugarria dela eta edozein hizkuntzatan erregistroetan zeharreko aldakortasun handia erakusten duela (Azpiazu 2004).

Nominalizazioa funtsezko eragiketa da jakintza eraikitzea eta transmititzea helburu duten diskurtso akademikoetan. Izan ere, nominalizazioaren bitartez lortzen den trinkotze eta abstrakzioari esker, kontzeptu konplexuak objektuak bailiran trata daitezke diskurtsoan (Iturrioz 2000). Ildo horretatik, oro har onartutako ideia da diskurtso akademikoei diskurtso orokorrek baino gehiago ustiatzen dituztela izenak, eta idatzizko erregistroek ahozkoek baino areago ustiatzen dutela nominalizazioa (besteak beste, Kocourek 1991, Biber *et al.* 1999, Biber 2006, Azpiazu 2004, Parodi & Venegas 2004, Castelló 2007, Zabala & Elordui 2010). Ideia hori frogatzen duten datu kuantitatiboak ere eskaini dira. Adibidez, Biber-ek (2006) *Longman Spoken and Written English Corpusean* 300.000 izen aurkitu ditu milioi bat hitzeko azpicorpus akademikoan eta, ahozko corpusean, aldiz, 150.000 izen milioi bat hitzeko. Ehunekoei emanda, kalkulatzen du azpicorpus akademikoaren esanahidun hitzen % 60 inguruan izenak direla. Areago, diskurtso akademiko idatziek diskurtso orokorrek eta ahozkoek baino izen gehiago ez ezik, izen-sintagma luzeagoak ere erabiltzen dituzte (besteak beste, Kocourek 1991, Biber 2006, Quiroz 2008).

Nominalizazio deritzen fenomenoak oso konplexuak dira eta hizkuntzaren azterketa-maila guztietan (morfologia, sintaxia eta diskurtsoa, semantika eta pragmatika) azaleratzen dira horien aztarnak (Iturrioz 2000: 33). Maila horiek guztiak konektatuta daude eta, azken batean nominalizazio-estrategietara jotzeko motibazioa diskurtso-beharretatik sortzen bada ere, aditzetiko izenak (*azterketa*, *joera*, *antolakuntza*, *gainezarmen*, *zabalkunde*, *hartzaile*, *hondakin*, *neurgailu*...) eta adjektiboak (*desiragarri*, *disolbagaitz*, *emankor*, *eginkizun*...) behar semantiko-pragmatiko horien azken kristalizazioa dira: diskurtso-komunitateen barruko interakzio diskurtsiboen emaitza.

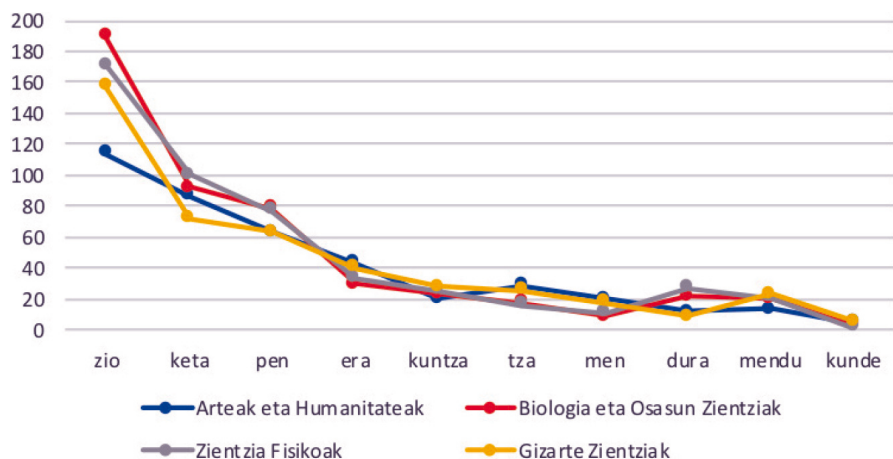
Euskarak, hizkuntza eranskarietan ohikoa den bezala, atzizki eratorle ugari ditu eta, erregistro akademikoaren garapenarekin lotuta, arreta berezia erakarri dute gertaera-izenek, atzizki ugari baititu euskarak horrelako izenak

sortzeko: *-era*, *-keta*, *-kuntza*, *-men/-pen/-mendu*, *-dura*, *-tza*, *-kunde*, *-zio* edo *-tze*. Estilo-liburuetan bi kezka aipatu ohi dira horien inguruan: a) Aditzean edo aditz-izenean (*apurtu*, *apurtze*) oinarritutako egiturak «egokiagoak» dirateke, aditzetiko izenetan (*apurketa*) oinarritutakoak baino. b) Aditzetiko izenetarako erabiltzen diren atzizkietan aldakortasun asistematikoa handia eta eraketa oker ugari bide daude. Hizkuntzalariek morfologia-eraketaren eta semantikaren ikuspuntutik aztertu dituzte, batez ere, gertaera-izenak (Azkarate 1991, 2008; Euskaltzaindiaren Gramatika Batzordea 2021). Azkaratek (2008), adibidez, atzizki emankorrenak hartzen duten oinaren eta eratorriaren semantikaren karakterizazioa egin du: *-era* atzizkiak *da* laguntzailea hartzen duten aditzekin konbinatzeko joera du (*etorrera*, *egoera*, *bilakaera*), baina zenbait aditz iragankorrekin ere elkartzen da (*idazkera*, *erabilera*...), eta kopapena, mugimendua, egoera, egoera-aldaketa edo modua adierazten duten izenak sortzen ditu. *-keta* atzizkia, aldiz, aditz iragankorrekin elkartzen da batez ere, baina ez predikatu psikologikoekin edo ez-ergatiboekin (*erosketa*, *erakusketa*, *prestaketa*). Azkenik, *-men/-pen/-mendu* atzizkia batez ere aditz psikologikoekin elkartzen da (*miresmen*, *gaitzezpen*, *adierazpen*).

Erregistroetan zeharreko aldakortasunaren ikuspuntutik egindako beste ikerketa batzuek, aldiz, agerian utzi dute faktore morfopragmatikoen eragin nabarmena dutela atzizki eratorleen emankortasunean, eta erregistroen bereizgarriak ere izan daitezkeela (Gallegos 2000; Lersundi *et al.* 2008; Guz 2009; San Martín *et al.* 2010). Ildo horretatik, aztertu dugu *HARTAEUS* corpusen gertaera-izenak sortzen dituzten atzizkien emankortasuna lau hiperreremuetan. Diskurtso-komunitate akademikoetan partekatzen eta finkatzen doazen baliabideak identifikatu nahi izan ditugu eta, beraz, 10 agerraldi milioi bat hitzeko baino agerraldi kopuru txikiagoa duten izenak baztertu ditugu, joera orokorretatik urruntzen diren erabilera indibidualak baztertzearren. Joera orokortzat hartu ohi denarekin bat etorrira (Azkarate 2008), *-zio*, *-keta* eta *-pen* atzizkiak dira emankorrenak corpus akademikoaren lau hiperreremuetan ere. Lemen banaketari dagokionez, aldea oso handia ez bada ere, badirudi Zientzia Fisikoetan eta Biologia eta Osasun Zientzietan oso antzera banatzen direla atzizkiak, eta Arteak eta Humanitateak eta Gizarte Zientziak, aldiz, apur bat urruntzen direla horien joeratik: *-zio*, *-keta*, *-pen* eta *-dura* emankorrak dira Zientzia Fisikoetan eta Biologia eta Osasun Zientzietan eta, *-era*, *-tza* eta *-men* atzizkiak, aldiz, apur bat emankorrak Arte eta Humanitateetan eta Gizarte Zientzietan.

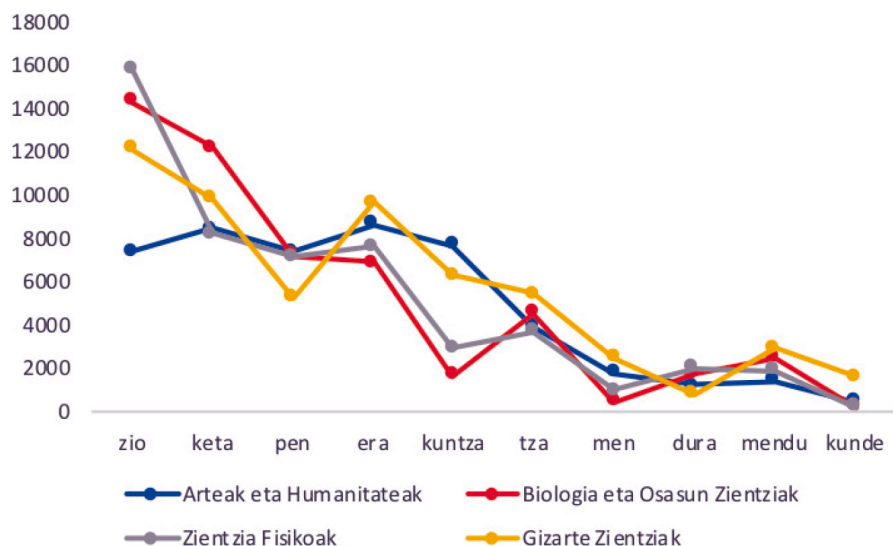
Agerraldi kopuruari ereparatuta, aldiz, ikus daiteke Zientzia Fisikoak gailentzen direla *-zio* atzikiaren erabileran, eta *-keta* atzizkiaren erabileran, aldiz, behera egiten dutela, eta Arte eta Humanitateekin partekatzen direla. Biologia eta Osasun Zientziak gailentzen dira *-keta* atzizkiaren erabileran, eta Arte eta Humanitateak eta Gizarte Zientziak nabariki gailentzen dira *-era*, *-kuntza* eta *-men* atzizkien erabileran. Joera batzuk ikus badaitezke ere, kata txiki hau lehen hurbilketa bat baino ez da: espezialitate-alor xeheagoe-

tan zeharreko banaketa aztertu behar da, atzizki batzuk edo beste batzuk us-tiatzeko motibazio semantiko-pragmatikoak detektatzeko.



1. irudia

Gertaera-izenak sortzen dituzten atzizkien emankortasuna
HARTAEus corpusaren lau hipereremuetan, ematen duten lema kopuruen arabera



2. irudia

Gertaera-izenak sortzen dituzten atzizkien emankortasuna
HARTAEus corpusaren lau hipereremuetan,
 agerraldi kopurua/ M, 10 agerraldi/M baino gutxiago dituztenak baztertuta

5. ONDORIOAK

Artikulu honetan, GrAL eta MAL testu-generoez osatutako *HARTAeus* corpora (3.285.098 hitz) aurkeztu, eta erregistro akademikoen garapena aztertzeke baliagarria dela defendatu dugu. Lexiko akademikoa, kolokazio akademikoak eta diskurtso-formula akademikoak erauzi ditugu corpusetik, ondo garatutako hizkuntzetarako erabili ohi diren neurriak eta atalaseak erabiliz. Hortaz, esan daiteke testu akademikoen idazketarako gako diren eraikuntza-bloketzat hartzen diren konbinazio lexiko akademikoak neurri batean behintzat garatuak direla euskaraz.

Gertaera-izenak sortzen dituzten atzizkiek *HARTAeus* corpusaren lau hipereremuetan (Arteak eta Humanitateak, Biologia eta Osasun Zientziak, Zientzia Fisikoak eta Gizarte Zientziak) duten emankortasuna ere aztertu dugu, eta zenbait joera antzeman daitezkeela erakutsi dugu. Garbi dago, ordea, joera morfopragmatiko sendoak antzemateko askoz ere azterketa xeheagoak egin behar direla.

Egin ditugun azterketak euskararen erregistro akademikoen ikerketarako lehen urratsak baino ez dira, etorkizunerako ikerketa-ildoak zabaldu nahi dituztenak.

6. BIBLIOGRAFIA

- Amorrortu, Estibaliz, Ane Ortega & Jone Goirigolzarri. 2017. Euskararen hiztun aktibo bihurtzeko gakoak. *BAT Soziolinguistika Aldizkaria* 104(3). 11-49.
- Alberdi, Xabier & Ibon Sarasola. 2001. *Euskal estilo libururantz. Gramatika, estiloa eta hiztegia*. Bilbo: UPV/EHUko Argitalpen Zerbitzua.
- Alonso-Ramos, Margarita, Marcos García-Salido & Marcos García. 2017. Exploiting a Corpus to Compile a Lexical Resource for Academic Writing: Spanish Lexical combinations. In Iztok Kosem, Jelena Kallas, Carole Tiberius, Simon Krek, Miloš Jakubíček & Vit Baisa (arg.). *Electronic lexicography in the 21st century. Proceedings of eLex 2017 conference*, 571-586. Leiden (the Netherlands).
- Aranzabe, Maria Jesús, Antton Gurrutxaga & Igone Zabala. 2022. Compilación del corpus académico de noveles en euskera HARTAeus y su explotación para el estudio de la fraseología académica. *Procesamiento del Lenguaje Natural* 69. 95-103.
- Azkarate, Miren. 1991. *Irakaskuntza Proiektua*. Euskal Filologia Saila. Filosofia eta Hezkuntza Zientzien Fakultatea. UPV/EHU. (Argitaragabea).
- Azkarate, Miren. 2008. Gertaera- eta emaitza-izenak. In Xabier Artiagoitia & Joseba A. Lakarra (arg.). *Gramatika jaietan. Patxi Goneagaren omenez*, 93-106. Bilbo: ASJU gehigarriak. UPV/EHUko Argitalpen Zerbitzua.
- Azpiaz, Susana. 2004. *Las estrategias de nominalización*. Frankfurt: Peter Lang.
- Biber, Douglas, Stig Johanson, Geoffrey Leech, Susan Conrad, & Fenegan Edward. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Pearson.

- Biber, Douglas, Susan Conrad & Viviana Cortes. 2004. *If you look at...: lexical bundles in university teaching and textbooks*. *Applied Linguistics* 25 (3). 371-405.
- Biber, Douglas. 2006. *University Language. A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.
- Biber, Douglas & Susan Conrad. 2009. *Register, Genre, and Style*. Cambridge: Cambridge University Press.
- Cabré, Maria Teresa. 2003. Terminología y normalización lingüística. In Xabier Alberdi, Iñaki Ugarteburu & Pello Salaburu (arg.). *Espezialitate hizkerak eta terminologia jardunaldiak*, 11-25. Bibo: UPV/EHuko Argitalpen Zerbitzua.
- Castelló, Montserrat. 2007. Escribir y comunicarse en contextos científicos y académicos. Conocimientos y estrategias. Barcelona: Graó.
- Coxhead, Averil. 2000. A New Academic Word List. *TESOL Quarterly* 34 (2). 213-238.
- De Cock, Silvie, Sylviane Granger, Geoffrey Leech & Tony Mc Enery. 1998. An automated approach to the phrasicon of EFL learners. In Sylviane Granger (arg.). *Learner English on Computer*, 67-79. Londres: Addison, Wesley, Long.
- Drouin, Patrick. 2007. Identification automatique du lexique scientifique transdisciplinaire. *Revue Française de Linguistique Appliquée* 12 (2). 45-64.
- Durrant, Philip. 2013. Formulaicity in an agglutinating language: the case of Turkish. *Corpus Linguistics and Linguistic Theory* 9(1). 1-38.
- Erman, Britt & Beatrice Warren. 2000. The idiom principle and the open choice principle. *Text-Interdisciplinary Journal for the Study of Discourse* 20. 29-62.
- Euskaltzaindiaren Gramatika Batzordea. 2021. *Euskararen Gramatika*. Bilbo: Euskaltzaindia.
- Frankenberg-Garcia, Ana, Robert Lew, Jonathan C. Roberts, Geranti Paul Rees & Nirwan Sharma. 2018. Developing a writing assistant to help EAP writers with collocations in real time. *ReCALL* 31 (1). 23-39.
- Gallegos, Alfonso. 2000. Morfología y registro. Algunas relaciones entre tradiciones discursivas y morfología derivativa nominal en español. *Función* 21-24. 142-215.
- Granger, Sylviane & Magali Paquot. 2015. Electronic lexicography goes local: Design and structures of a need-driven online academic writing aid. *Lexicographica: International Annual for Lexicography* 31 (1). 118-141.
- García Salido, Marcos, Marcos García, Milka Villayandre & Margarita Alonso-Ramos. 2018. A Lexical Tool for Academic Writing in Spanish Based on Expert and Novice Corpora. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*, 260-265.
- García Salido. 2021. Compiling an Academic Vocabulary List of Spanish. DOI:10.13140/RG.2.2.27681.33123
- Gardner, Dee & Mark Davies. 2004. A new academic vocabulary list. *Applied Linguistics* 35 (3). 305-327.
- Garzia, Juan. 2005. *Kalko okerrak*. Ikasmaterialen Aholku Batzordea. Estilo-liburua-aren bigarren atala. Vitoria-Gasteiz: Eusko Jaurlaritza.
- Gurrutxaga, Antton & Iñaki Alegria. 2011. Automatic extraction of NV expressions in Basque: basic issues on cooccurrence techniques. In *Proceedings of the Workshop on Multiword Expressions: from parsing and generation to the real world*, 2-7.
- Gurrutxaga, Antton, Iñaki Alegria & Xabier Artola. 2018. Caracterización computacional de la idiomatidad: aplicación a la combinación nombre+verbo en eus-

- kera. In Ruiz Miyares, L. (arg.). *Estudios de Lexicología y Lexicografía. Homenaje a Eloína Miyares Bermúdez*. Santiago de Cuba (Cuba).
- Guz, Wojciech. 2009. English Affixal Nominalizations Across Language Registers. *Poznań Studies in Contemporary Linguistics* 45(4). 461-485.
- Hyland, Ken. 2008. As can be seen: lexical bundles and disciplinary variation. *English for Specific Purposes* 27 (1). 4-21.
- Hymes, Dell. 1974. *Foundations in Sociolinguistics: An Ethnographic Approach*. Philadelphia: University of Pennsylvania Press.
- Iturrioz, José Luis. 2000. Diversas aproximaciones a la nominalización. De las abstracciones a las macrooperaciones textuales. *Función* 21-24. 32-140.
- Kocourek, Rostislav. 1991. Textes et termes. *Meta* 36 (1). 71-76.
- Labov, William. 1974. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
- Laurén, Christer, Johan Myking, & Heribert Picht. 2002. Language and domains: a proposal for a domain dynamics taxonomy. *LSP & Professional Communication*. Vol 2 (2).
- Lersundi, Mikel, Igone Zabala & Agurtzane Elordui. 2008. Aditzetiko izenen emankortasunaren azterketa morfopragmatikoa euskarazko corpus orokor eta berezitueta. In Xabier Artiagoitia & Joseba A. Lakarra (arg.). *Gramatika jaietan. Patxi Goneagaren omenez*, 491-506. Bilbo: ASJU gehigarriak. UPV/EHUko Argitalpen Zerbitzua.
- Mel'čuk, Igor. 1995. Phrasemes in Language and Phraseology in Linguistics. In Martin Everaert, Erik-Jan van der Linden, André Schenk & Rob Schreu (arg.). *Idioms: Structural and Psychological Perspectives*, 167-232. Hillsdale: Lawrence Erlbaum Associates.
- Mel'čuk, Igor. 1998. Collocations and Lexical Functions. In Anthony Paul Cowie (arg.) *Phraseology. Theory, Analysis and Applications*, 23-53. Oxford: Clarendon Press.
- Mel'čuk, Igor. 2015. Clichés, an Understudied Subclass of Phrasemes. *Yearbook of Phraseology* 5. 35-50.
- Nam, Kilin, Hyun-ju Song, & Jun Choi. 2016. A morpheme-based analysis of lexical bundles in Korean: an interface between corpus-driven approach and lexicography. *Lexicography. Journal of ASIALEX* 3. 39-62.
- Öksüz, Doğuş Can. 2019. Collocational processing in typologically different languages, English and Turkish: Evidence from corpora and psycholinguistic experimentation. PhD. Lancaster University.
- Paquot, Magali. 2018. Phraseological Competence: A Missing Component in University Entrance Language Tests? Insights from A Study of EFL Learners' Use of Statistical Collocations. *Language Assessment Quarterly* 15 (1). 29-43.
- Parodi, Giovanni & René Venegas. 2004. Bucólico: Aplicación computacional para el análisis de textos: Hacia un análisis de rasgos de la informatividad. *Literatura y Lingüística* 15. 223-257.
- Quiroz, Gabriel Ángel. 2008. Los sintagmas nominales extensos especializados en inglés y en español: descripción y clasificación en un corpus de genoma. Dokto-retza-tesia. Universitat Pompeu Fabra.
- San Martin, Itziar, Joseba Ezeiza & Igone Zabala. 2010. Aditzetiko izenak erregistro bereizle gisa: corpus espezializatu baten azterketa morfopragmatikoa. In Xabier

- Alberdi & Pello Salaburu (arg.). *Euskararen garapena esparru akademikoetan*, 224-242. Bilbo: EHUKo Argitalpen Zerbitzua.
- Seoane, Elena & Douglas Biber. 2021. A corpus-based approach to register variation. In Elena Seoane & Douglas Biber (arg.). *Corpus-based Approaches to Register Variation*, 2-16. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Siyanova-Chanturia, Anna. 2015. On the «holistic» nature of formulaic language. *Corpus Linguistics and Linguistic Theory* 11. 285-301.
- Tutin, Agnès & Marie-Paule Jacques. 2018. Le lexique scientifique transdisciplinaire: une introduction. In Agnès Tutin & Marie-Paule Jacques (arg.). *Lexique transversal et formules discursives des sciences humaines*, 1-26. ISTE Editions.
- Txurruka, Jesus Mari. 2020. Nola sartu zen euskara Zientzi Fakultatean. In Jacinto Iturbe, Jesus Mari Txurruka & Aitor Castañeda (arg.). *Amets bikoitza. Euskara zientifikotzea eta zientzia euskaratzea*, 57-190. Bilbo: UPV/EHUKo Argitalpen Zerbitzua.
- Villayandre, Milka. 2018. «HARTA» de noveles. Un corpus de español académico. *CHIMERA. Romance Corpora and Linguistic Studies* 5(1).131-140.
- Zabala, Igone. 2003. Causation and Semantic Control. Diagnosis of Incorrect Uses in Minorized Languages. In Bernard Oiharçabal (arg.), 255-284. *Inquiries into the Lexicon-Syntax Relations in Basque*. Bilbo: ASJU.
- Zabala, Igone. 2019. The Elaboration of Basque in Academic and Professional Domains. In Grenoble, Leonore, Pia Lane & Unn Røyneland (arg.) & Ivan Igarua & Lourdes Oñederra (euskal arg.). *Linguistic Minorities in Europe Online*. De Gruyter Mouton. <https://doi.org/10.1515/lme.9612443>
- Zabala, Igone. 2021. Euskaltzaindiaren Hiztegiaren ekarpena lexiko espezializatua- ren eta ez-espezializatuaren harmonizazioan. In Andres Urrutia (arg.). *Arantzazutik mundu zabalera. Euskararen normatibizazioa: 1968-2018. IKER 40*, 285-299. Bilbo-Madrid-Frankfurt: Euskaltzaindia-Iberoamericana Vervuert.
- Zabala, Igone & Agurtzane Elordui. 2010. Nominalizazioa estrategia diskurtsibogisa: corpus akademikoaren azterketaren beharra. In Xabier Alberdi & Pello Salaburu (arg.). *Euskararen garapena esparru akademikoetan*, 224-242. Bilbo: UPV/EHUKo Argitalpen Zerbitzua.
- Zabala, Igone, Mikel Lersundi, Igor Leturia, Iker Manterola & Gotzon Santander. 2013. GARATERM: euskararen erregistro akademikoaren garapenaren ikerketarako lan-ingurunea. Xabier Alberdi & Pello Salaburu (arg.). *Terminologia naturala eta terminologia planifikatua euskararen normalizazioari begira*, 98-114. Bilbo: UPV/EHUKo Argitalpen Zerbitzua.
- Zabala, Igone, Itziar San Martin & Mikel Lersundi. 2014. Linguistic and sociolinguistic factors that influence the detection, implantation and circulation of natural terminology in academic uses of Basque. In Pascaline Dury, Jose Carlos de Hoyos, Julie Makri-Morel, François Maniez, Vincent Renner & Maria Belén Villar Díaz (arg.). *La néologie en langue de spécialité. Neology in specialized languages. La neología en lengua de especialidad*, 141-164. Lyon: Publications du CRTT.

