

## ADITZETIKO IZENAK ERREGISTRO BEREIZLE GISA: CORPUS ESPEZIALIZATU BATEN AZTERKETA MORFOPRAGMATIKOA<sup>1</sup>

Itziar San Martin, Joseba Ezeiza eta Igone Zabala  
(Euskal Herriko Unibertsitatea)

[AURKIBIDEA](#)

### 1. Sarrera<sup>2</sup>

Edozein hizkuntzak komunikazio-tresna egokia izateko garatuta izan behar ditu askotariko erregistroak. Erregistro esaten diogu komunikazio-egoera jakin batekin lotuta hiztunek hautatzen dituzten hizkuntza-baliabideen multzoari. Askotan, erregistro baten bereizgarria ez da beste erregistroetan erabiltzen ez den baliabide bat erabiltzea, baizik eta baliabide jakin bat beste erregistroetan ez bezalako maiztasunaz ustiatzea<sup>3</sup>. Adibidez, hizkera orokorrekin alderatuz, esparru espezializatuetakoa erregistroen ezaugarri nabarmen bat da aditzetiko izenen ustiapena. Hau ez da harritzekoa, aditzetiko izenek aise betetzen baitute komunikazio-egoera espezializatuetan sortzen den denominazio beharra (Zabala 2007).

Erregistro espezializatuetera mugatzen bagara, kontuan hartzeko modukoa da gai honen inguruko zenbait ikerketak (Gallegos 2000, 2003) erakutsi dutena, alegia, aditzetiko izenak sortzeko aukeratzen diren atzizkien ustiapena desberdina dela espezialitate-alor desberdinetako testuetan. Honek iradokitzen du bereizgarri tematikoa edo gaia lotuta dagoela komunitate diskurtsibo bakoitzak arloko behar semantiko-pragmatikoei aurre egiteko eskuragarri dituen atzizkien artean egiten dituen hautuekin.

Lan honetan euskaraz aditzetiko izenen ustiapenaren inguruan egin diren ikerketak abiapuntu hartu ditugu (Lersundi et al. 2008) eta izen horiek euskararen hainbat erregistrotan zehar duten banaketaren analisi estatistikoa egin dugu, banaketa-sistema konsistente bat osatzen ote duten argitzen saiatzeko asmoz. Biber-ren (2004) lanari jarraiki, erregistroetan zeharreko aldakortasunaren patroien azterketa funtsezkoa da, bai hizkuntza jakin baten deskripzio xeherako, eta bai hizkuntzaren erabilerari buruzko hizkuntzarteko teoriak garatzeko ere<sup>4</sup>.

<sup>1</sup> Lan hau hiru proiekturen elkargunean dago: RICOTERM-3 (HUM2007-65966-C02-02), EBALUA (EHU 08/53) eta HIZLAN (DIPE08/16).

<sup>2</sup> Eskerrak eman nahi dizkiogu Elhuyar Fundazioko Igor Leturiari ZT corpuseko zenbait datu erauzteko emandako laguntzagatik. Eskerrak eman nahi dizkiegu, halaber, Ametzagaña I+G ikerketa-guneko ikertzaile diren Aitor Maritxalar, Josu Landa eta Sebi Calonjeri, testuak prozesatzeko emandako laguntzagatik.

<sup>3</sup> Erregistroak askotariko orokortasun mailaz defini daitezke. Adibidez, erregistro orokorrez edo erregistro espezializatuhez hitz egin dezakegu oso modu lausoan. Bereizketa zehatzagoa egingo genuke prosa akademikoaren erregistroaz arituko bagina, are zehatzagoa litzateke adituen arteko ikerketa-artikuluek arituko bagina eta are finagoa Fisika Nuklearreko ikerketa-artikuluen metodologia-atalen erregistroaz arituko bagina (Biber 2004:1).

<sup>4</sup> Izan ere, erregistroen aldakortasunean badira unibertsaltzat har daitezkeen ezaugarri batzuk, hizkuntzartekotzat har daitezkeenak eta, beste ezaugarri batzuk aldiz, hizkuntza jakin baten sistemarekin lotuta daudenak.

## 2. Abiapuntua eta helburuak

Hizkuntza-komunitate bakoitzak du bere erregistro-sistema eta komunitateko kideak sistema horrekin baliatzen dira komunikazio-egoera desberdinei aurre egiteko (Ure 1982: 5). Ildo honetatik, esparru jakin batzuetan erabili ez den hizkuntzak ez ditu garatuta izango esparru horretako erregistroak. Soilik hizkuntza-esparru horietan erabiltzearekin batera joango dira sortzen eta finkatzen komunikazio-egoera berrietan beharko diren hizkuntza-baliabideak.

Euskarak esparru espezializatuetan orain arte izan duen ibilbidea laburra da inguruko hizkuntza ez gutxituekin alderatzen badugu, baina badirudi nolabaiteko aurrerapausoak behintzat eman dituela zenbait joera antzematea ahalbidetzen digutenak (Lersundi et al. 2008). Lan honetan aditzetiko izenak aztertu ditugu euskarazko hainbat erregistrotan eta azterketa estatistikoen bidez saiatu gara zehazten ea komunikazio-egoera berrietan hizkuntza-baliabide horien erabilerak patroiz sistematikorik erakusten duen.

Zabala eta Elorduiren (2010) lanaren haritik, abiapuntuko onarpenzat hartu dugu aditzetiko izenak diskurtsoan gertatzen diren nominalizazio-eragiketen emaitza direla. Izen ebentibo hauek testuen ekoizleek beren diskurtsoari abstrakzio mailarik altuena eman nahi diotenean erabiltzen dituzten elementu lexikoak direla onartuko dugu.

Ildo honetatik, Gallegosen (2000, 2003) lanen argitan diseinatu dugu gure lana, zehazkiago, Gallegosek frogatutzat uzten dituen zenbait ideiaren argitan. Batetik, aditzetiko izenak sortzeko atzizkien hautapena (*enganche/enganchado; liga/ligadura*) oso maiz erregistroak baldintzatzen duela dio Gallegosek. Bestalde, morfopragmatikaren bidetik, defenditzen du itxuraz baliokideak izan daitezkeen atzizkien arteko hautapena komunikazio-ekintza markatzeko estrategia diskurtsiboa dela. Horrela, gaztelaniaren erregistro orokorretan atzizki markatugabeak (-Ø, -e, -o) erabiltzeko joera ikus daiteke eta, erregistro espezializatuetan aldiz, aukera markatugabeetatik urruntzeko joera (-ado, -aje, -cion, -dura, -miento). Baina atzizki hauen hautapenean erregistro espezializatu guztiak ez dira berdin urruntzen erregistro orokorretatik. Izan ere, modu polizentrikoan, bakoitza modu batera, urruntzen direla frogatzen du Gallegosek bere lanetan.

Euskararako ere, ez gara hutsetik abiatzen. Izan ere, lan honetan urrats bat eman nahi dugu Lersundi, Zabala eta Elorduik (2008) irekitako bidean. Hizkuntzalari hauek euskaraz ere Gallegosek deskribatutako banaketan antzekoak badirela pentsarazten diguten zenbait joera deskribatu dituzte: hamaika atzizki aztertu dituzte (-dura, -era, -keta, -kunde, -kuntza, -men, -mendu, -pen, -tza, -tze eta -zio). Atzizki hauek dituzten sarrerak aztertu dituzte bi hiztegitan —Sarasolaren Euskal Hiztegia eta Elhuyarren Euskal Hiztegi Modernoan— eta atzizki horien bidez eratorritako lemekin konparatu dituzte sarrera horiek bi corpusetan —Ereduzko Prosa Gaur corpusean (EPG) eta ZT corpus tekniko-zientifikoan (ZT) —. Hasteko, agerian utzi dute euskararen erabilera berriak direla eta, corpusetan era honetako lema ugari daudela hiztegitan jasota ez daudenak. Lema berri horiek, gainera, askoz ere ugariagoak dira ZT corpusean EPG corpusaren liburuaz azpicorpusean baino, eta azpicorpus horretan ugariagoak EPG corpusaren prentsako azpicorpusean baino. Ondorio hau ondo uztartzen da aditzetiko izen ebentiboei diskurtso abstraktua eta bereziki diskurtso espezializatu eraikitzeke egozten zaien eginkizunarekin (Porzig 1930; Iturrioz 1985, 2000; Zabala eta Elordui 2010). Bestalde, Gallegosek aurreikusi bezala, agerian utzi dute atzizki hauek euskaraz ere desberdin ustiatzen direla corpus orokorretan eta corpus espezializatuetan, eta ZT corpusaren azpialal desberdinetan (BZ, LZ, ME, ZZ, Or) ere desberdin ustiatzen direla. Hortaz, Lersundi et al.-en

(2008) azterketek pentsarazten digute aditzetiko izen ebentiboak ematen dituzten atzizkiek euskaraz ere banaketa polizentrikoa izan lezaketela erregistroetan zehar.

Lan honetan, aurrera egin nahi dugu izen ebentiboak eratortzen dituzten atzizkien azterketa morfopragmatikoan; hori dela eta, gure lanak zenbait osagai komun ditu Lersundi, Zabala eta Elorduiren lanarekin, baina zenbait desberdintasun ere baditu. Hasteko, azterketa honetan ez dugu kontuan hartu *-tze* atzizkia eta, beraz, aurreko lanean baino atzizki gutxiago aztertu ditugu. Garbi dago *-tze* atzizkia elementu garrantzitsua dela sistema honetan, eta horren jokaera ere ulertu behar dela sistema bere osotasunean ulertzeko. Nolanahi ere, nahiago izan dugu oraingoan alde batera uztea, aurrerago egin nahi ditugun azterketa xeheagoen zain. Bigarrenik, lan honetan ez dugu aztertu hiztegirik, baina corpus gehiago aztertu ditugu, Lersundi, Zabala eta Elorduik antzemandako joerak bestelako corpusetan ere konsistenteak ote diren egiaztatzeko asmoz. Hirugarrenik, atzizki desberdinak dituzten lema kopuruak ez ezik, atzizki bakoitzak duen agerpen kopurua (maiztasuna) ere hartu dugu kontuan. Azkenik, analisi estatistikoak egin ditugu, antzemandako aldeak esangarriak ote diren egiaztatzeko asmoz.

Hauek dira lan hau aurrera eramateko abiapuntutzat hartu ditugun hipotesiak, analisi estatistikoen bitartez egiaztatu nahi ditugunak:

- a) Aztergai ditugun atzizkiek corpus orokorretan eta corpus espezializatuetan emankortasunari eta jokaerari dagokienez duten aldea (Lersundi et al. 2008) esangarria bada, joera horiek mantendu egingo dira beste corpus batzuetan ere, osaera eta tamaina edozein delarik.
- b) Atzizki hauek ZT corpuseko azpicorpusetan (BZ, LZ, ME, ZZ, TE,...) dituzten joerak alor bakoitzeko testuek jorratzen dituzten gaiekin lotuta daudela dirudi. Hala bada, antzeko joerak antzeman beharko lirateke alor horietako beste corpus batzuk aztertzen ditugunean.
- c) Espero izatekoa da giza eta gizarte-zientzietako corpusen joera urruntzea corpus tekniko-zientifikoetatik, eta gehiago hurbiltzea corpus orokorren joerara.
- d) Alorrek xehekiago mugatzen goazen neurrian, espero izatekoa da zenbait atzizkiren joera besteetatik urruntzen joatea, aztertzen dugun alorraren bereizgarri tematikoen kausaz.

### 3. Metodologia

Hipotesien egiazkotasuna frogatzeko 44.092.669 hitzeko testu masa aztertu dugu. Testu horiek hainbat iturritatik eskuratu ditugu (Ikus 1. taula).

Alde batetik, bi corpus publiko hustu ditugu: Ereduzko Prosa Gaur (EPG)<sup>5</sup> eta Zientzia eta Teknika Corpora (ZTC)<sup>6</sup>. Horrez gain, beste material batzuk ere bildu eta landu ditugu: Pentsamendu Unibertsalaren Klasikoen sailean 2006. urtera arte argitaratutako liburuak (KLS)<sup>7</sup>, UPV/EHUko argitalpen elektronikoen Testu Biltegiko hainbat esku-liburu (TB)<sup>8</sup>, eta

<sup>5</sup> <http://www.ehu.es/euskara-orria/euskara/ereduzkoa/>

<sup>6</sup> <http://www.ztcorpusa.net/cgi-bin/kontsulta.py>

<sup>7</sup> <http://www.klasikoak.com/>

<sup>8</sup> <http://testubiltegia.ehu.es/>

UPV/EHUko zenbait irakaskuntza-material, Euskara eta Eleaniztasuneko Errektoreordetzak eskaintzen dituen Irakasleen Prestakuntza Linguistikorako Tailerretan eskuratuak (IT).

Horrela, bada, ikerketa honetan erabilitako corpusaren testu-masa osoaren % 62 inguru testu orokorrek osatzen dute, prentsako testuek eta literatura-lanek, alegia. Giza zientzietako saio-lanak % 22 inguru dira. Eta beste testu guztiak hainbat espezialitate-alorretakoak dira, alegia, enpresa zientzietakoak, zuzenbideren alorrekoak, zientzietakoak, osasun zientzietakoak, eta teknologiaren alorrekoak.

Ahal izan denean, erregistro edota alor bereko hiru lagin edo gehiago erabili dira, esangarritasun-proba estatistikoak berme handiagoz aplikatu ahal izateko. Horrela bada, lan honetan testu orokorrek osatutako lau lagin aztertu dira (EPG eta ZT corpusetakoak); giza zientzien alorreko beste hiru lagin (KLS corpusekoak), zientzien alorreko lau lagin (ZT corpusekoak), teknologiaren alorreko hiru lagin (ZT, TB eta IT corpusetakoak), eta zuzenbidearen alorreko beste hiru lagin (TB corpusekoak). Alabaina, enpresa zientzien eta osasun zientzien alorren kasuan, bina lagin besterik ez ditugu eskuratu (IT eta TB corpusetakoak). Ondorengo taulan laburbiltzen dira ikerketa-corpusaren ezaugarri nagusiak:

ITURRIA	ALORRAK / ERREGISTROA	HITZ-KOPURUA
EPG	Orokorra_prensa (Pren) Orokorra_literatura (Lib)	25.100.000
ZT	Orokorra_aldizkariak (Orok) Bizitzaren Zientziak (BZ) Lurraren Zientziak (LZ) Materia eta Energia (ME) Zientzia Zehatzak (ZZ) Teknologia (TE) Bestelakoak (BS)	8.529.505
KLS	Giza Zientziak (GZ)	9.656.556
TB	Teknologia (TE) Osasun Zientziak (OZ) Enpresa Zientziak(ENP) Zuzenbidea (ZB)	642.867
IT	Teknologia (TE) Osasun Zientziak (OZ) Enpresa Zientziak(ENP)	163.741

1. taula: Ikerketako corpusaren egitura

Aipatutako corpusetatik *-dura*, *-era*, *-keta*, *-kunde*, *-dura*, *-kuntza*, *-men*, *-mendu*, *-pen*, *-tza* eta *-zio* atzizkiak daramatzaten izen ebentiboak hustu ditugu<sup>9</sup>, eta lema kopuruaren eta lema bakoitzaren maiztasunaren datuak jaso ditugu EXCEL programarekin eratutako datu-base batean. (Ikusi 1. irudia). Beheko irudian ikus daitekeenez, lemarekin batera, hainbat datu jaso dira sarrera bakoitzean: a) lema bakoitza zein aditzetatik eratortzen den; b) zenbat lagin eta corpusetan agertzen den; eta c) zenbat agerpen-kopuru eta zer maiztasun duen lagin

<sup>9</sup> KLS, TB eta IT testuak Ametzagaiña I+G Enpresa Gune Teknologikoak prozesatu eta weberatu zituen, beraiek garatutako tresnak erabilia (KAPSULA eta Corpus Arakatzaila). Helbide honetan kontsulta daiteke direlako testuak biltzen dituen corpusa: <http://weblingua.ametza.com/corpus-KL/>

bakoitzean. Guztira, 3650 lehari dagozkien agerpen-datuak bildu ditugu, 22.152 sarreratan jasoak. 1. irudian ikus daitekeenez, EXCEL iragaziek datu-basean hainbat bilaketa-mota egitea errazten dute. Adibide honetan *bereizi* aditzetik eratorritako lehen zerrenda iragazi dugu. Ikusten denez, aditz horretatik eratorritako bost lema daude jasota datu-basean: *bereizketa*, *bereizkuntza*, *bereizmen*, *albo-bereizmen* eta *bereizpen*. Zutabez zutabe erraz ikus daiteke zein corpus eta laginetan detektatu den lemetako bakoitza, zenbat aldiz agertzen den corpus eta lagin horietan, eta agerpen-datuek zer maiztasun ematen duten.

# UGARTEBURU TERMINOLOGIA JARDUNALDIK (2010)

ITZIAR SAN MARTIN ET ALII: "ADITZETIKO IZENAK ERREGISTRO BEREIZLE GISA (...)"

	A	B	C	D	E	F	G
1	Aditza	Atzikia	Lemak	Corpusa	Lagina (L)	Agerpenak	Maiztasuna (M) 10000 hitzek
2424	Bereizi	KETA	Bereizketa	EHuko Irakasleen Tailerretako testuen corpusa	IT_TE	1	0,1689
2425	Bereizi	KETA	Bereizketa	EHuko Irakasleen Tailerretako testuen corpusa	IT_ENP	1	0,1659
2426	Bereizi	KETA	Bereizketa	EHuko Irakasleen Tailerretako testuen corpusa	IT_OZ	2	0,4517
2633	Bereizi	KETA	Bereizketa	Ereduzko Prosa Gaur	EPG_Lib	505	0,3855
2634	Bereizi	KETA	Bereizketa	Ereduzko Prosa Gaur	EPG_Pren	244	0,2033
3126	Bereizi	KETA	Bereizketa	Euskarazko argitalpen elektronikoen Testu Biltegia (EHU)	TB_ENP	13	1,1963
3127	Bereizi	KETA	Bereizketa	Euskarazko argitalpen elektronikoen Testu Biltegia (EHU)	TB_OZ	5	0,3533
3128	Bereizi	KETA	Bereizketa	Euskarazko argitalpen elektronikoen Testu Biltegia (EHU)	TB_Zb_2	21	3,4302
3129	Bereizi	KETA	Bereizketa	Euskarazko argitalpen elektronikoen Testu Biltegia (EHU)	TB_Zb_3	14	0,9114
3130	Bereizi	KETA	Bereizketa	Euskarazko argitalpen elektronikoen Testu Biltegia (EHU)	TB_Zb_1	34	3,6554
3447	Bereizi	KETA	Bereizketa	Pentsamenduaren Klasikoen Corpusa	KLS_1991/1994	108	0,6582
3448	Bereizi	KETA	Bereizketa	Pentsamenduaren Klasikoen Corpusa	KLS_1995/1998	489	2,0289
3449	Bereizi	KETA	Bereizketa	Pentsamenduaren Klasikoen Corpusa	KLS_1999/2006	895	1,5966
4207	Bereizi	KETA	Bereizketa	Zientzia eta Teknika Corpusa	ZTC_ME	15	0,2407
4208	Bereizi	KETA	Bereizketa	Zientzia eta Teknika Corpusa	ZTC_LZ	9	0,2693
4209	Bereizi	KETA	Bereizketa	Zientzia eta Teknika Corpusa	ZTC_TE	108	0,5185
4210	Bereizi	KETA	Bereizketa	Zientzia eta Teknika Corpusa	ZTC_BZ	114	0,4715
4211	Bereizi	KETA	Bereizketa	Zientzia eta Teknika Corpusa	ZTC_OR	168	1,1967
4212	Bereizi	KETA	Bereizketa	Zientzia eta Teknika Corpusa	ZTC_BS	147	1,4086
4213	Bereizi	KETA	Bereizketa	Zientzia eta Teknika Corpusa	ZTC_BS	1	0,0096
4214	Bereizi	KETA	Bereizketa	Zientzia eta Teknika Corpusa	ZTC_ZZ	9	0,1443
5813	Bereizi	KUNTZA	Bereizkuntza	Ereduzko Prosa Gaur	EPG_Pren	7	0,0583
5814	Bereizi	KUNTZA	Bereizkuntza	Ereduzko Prosa Gaur	EPG_Lib	222	1,6947
5919	Bereizi	KUNTZA	Bereizkuntza	Euskarazko argitalpen elektronikoen Testu Biltegia (EHU)	TB_Zb_2	2	0,3267
6007	Bereizi	KUNTZA	Bereizkuntza	Pentsamenduaren Klasikoen Corpusa	KLS_1991/1994	107	0,6521
6008	Bereizi	KUNTZA	Bereizkuntza	Pentsamenduaren Klasikoen Corpusa	KLS_1995/1998	304	1,2613
6009	Bereizi	KUNTZA	Bereizkuntza	Pentsamenduaren Klasikoen Corpusa	KLS_1999/2006	256	0,4567
6237	Bereizi	KUNTZA	Bereizkuntza	Zientzia eta Teknika Corpusa	ZTC_OR	106	0,7551
6238	Bereizi	KUNTZA	Bereizkuntza	Zientzia eta Teknika Corpusa	ZTC_BZ	4	0,0165
6239	Bereizi	KUNTZA	Bereizkuntza	Zientzia eta Teknika Corpusa	ZTC_ME	1	0,0160
6240	Bereizi	KUNTZA	Bereizkuntza	Zientzia eta Teknika Corpusa	ZTC_TE	3	0,0144
6241	Bereizi	KUNTZA	Bereizkuntza	Zientzia eta Teknika Corpusa	ZTC_BS	31	0,2970
6537	Bereizi	MEN	Bereizmen	EHuko Irakasleen Tailerretako testuen corpusa	IT_OZ	1	0,2259
6588	Bereizi	MEN	Bereizmen	Ereduzko Prosa Gaur	EPG_Lib	24	0,0183
6589	Bereizi	MEN	Bereizmen	Ereduzko Prosa Gaur	EPG_Pren	7	0,0058
6721	Bereizi	MEN	Bereizmen	Euskarazko argitalpen elektronikoen Testu Biltegia (EHU)	TB_TE	2	0,2357
6722	Bereizi	MEN	Bereizmen	Euskarazko argitalpen elektronikoen Testu Biltegia (EHU)	TB_OZ	5	0,3533
6799	Bereizi	MEN	Bereizmen	Pentsamenduaren Klasikoen Corpusa	KLS_1991/1994	13	0,0792
6800	Bereizi	MEN	Bereizmen	Pentsamenduaren Klasikoen Corpusa	KLS_1995/1998	49	0,2033
6931	Bereizi	MEN	Albo-bereizmen	Zientzia eta Teknika Corpusa	ZTC_ME	2	0,0321
6968	Bereizi	MEN	Bereizmen	Zientzia eta Teknika Corpusa	ZTC_ME	69	1,1070
6969	Bereizi	MEN	Bereizmen	Zientzia eta Teknika Corpusa	ZTC_BZ	31	0,1282
6970	Bereizi	MEN	Bereizmen	Zientzia eta Teknika Corpusa	ZTC_OR	17	0,1211
6971	Bereizi	MEN	Bereizmen	Zientzia eta Teknika Corpusa	ZTC_TE	186	0,8929
9685	Bereizi	PEN	Bereizpen	Zientzia eta Teknika Corpusa	ZTC_OR	39	0,277811812
9686	Bereizi	PEN	Bereizpen	Zientzia eta Teknika Corpusa	ZTC_BZ	13	0,053764819
9687	Bereizi	PEN	Bereizpen	Zientzia eta Teknika Corpusa	ZTC_TE	1	0,004800759
9688	Bereizi	PEN	Bereizpen	Zientzia eta Teknika Corpusa	ZTC_ZZ	1	0,016035972
11349							
11350							
11351							
11352							
11353							

1. IRUDIA: CORPUSETATIK ERAUZITAKO LEMA BAKOITZERAKO BILDUTAKO INFORMAZIOA

Lan honen abiapuntuko hipotesiak egiaztatzeko hainbat proba estatistiko erabili ditugu:

- Atzizkien banaketan corpus orokorren eta espezializatuen joeren arteko aldeak esangarriak ote diren egiaztatzeko, Mann-Whitneyren proba estatistikoa erabili dugu.
- Bigarren hipotesia egiaztatzeko, hau da, egiaztatzeko ea atzizki hauen banaketak eta joerak alorreko testuek jorratzen dituzten gaiekin zerikusirik ote duten, ANOVA probaren bidez aztertu dugu lema ezberdinen kopuruaren (N) eta lema horien agerpenen maiztasunaren (M) datuen banaketaren esangarritasuna .
- Hirugarren hipotesia frogatze aldera, alegia, giza eta gizarte-zientzietako corpusen joera corpus tekniko-zientifikoenetik modu esangarrian urruntzen ote den, eta corpus orokorren joerara gehiago hurbiltzen ote den egiaztatzeko, datuen analisi faktoriala egin dugu.
- Laugarren hipotesia egiaztatzeko, hau da, ea alorrak xehekiago mugatzen goazen neurrian zenbait atzizkiren joera besteetatik urruntzen ote doan ikusteko, giza eta gizarte zientzietako testuak eta zientzien alorrekoak isolatu ditugu, eta analisi faktorialaren emaitzak grafikoki irudikatu ditugu.

Proba estatistiko horiek guztiak egiteko SPSS programa erabili dugu. Zenbait irudi grafiko egiteko, eta datuen banaketaren erregresio-indizeak kalkulatzeko EXCEL programa erabili dugu.

#### 4. Emaitzak

Azterketaren lehen fasean lortutako emaitzen laburpen orokorra 2. taulan dago jasota. Datu horiek testu-masa osoari dagozkio. Ikusten denez, *-zio* atzizkia da aztertutako corpusetan gehien ustiatzen dena alde handiz; *-kunde* atzizkia berriz, lema gutxien ematen dituen eta denetan maiztasun txikiena duena. Edonola ere, atzizkien banaketa (hurrenkera) ez da berdina bi zutabeetan. Adibidez, *-tza* atzizkia ez da emankortasun handikoa (101 lema ezberdin), baina lema horiek maiz samar agertzen dira corpusean (136.315 agerpen). *-dura* atzizkiak, aldiz, emankortasun handiagoa du (286 lema ezberdin), baina askoz ere maiztasun txikiagoa (39.276 agerpen). Datu eta hurrenkera horiek, bestalde, ezberdinak dira corpusaren eta laginaren arabera.

Nren eta Mren balioen banaketen artean dagoen aldeak pentsarazi digu komeni litekeela bi datuak kontuan hartzea atzizki hauen ustiapena aztertzeko orduan, alderaketen emaitzak ezberdinak izango baitira lemen kopuruak edota lema horien maiztasunak kontuan hartuta.

Lema ezberdinen kopuruak (N)		Agerpen kopuruak (M)	
ZIO	1.453	ZIO	306.655
KETA	597	ERA	229.955
PEN	551	KETA	178.980
DURA	286	PEN	154.637
MENDU	192	TZA	136.315
MEN	157	MEN	101.508
KUNTZA	141	KUNTZA	69.217
ERA	138	MENDU	54.507
TZA	101	DURA	39.276
KUNDE	33	KUNDE	27.715
GUZTIRA	3.650	GUZTIRA	1.298.765

2. taula: corpusetarik erauzitako lema ezberdinen kopuruak eta lema horien agerpen kopuruak

Hirugarren taulan ikus daiteke hustuketa-fasean bildutako lema nola banatzen diren corpus desberdinetan zehar. Jasotako 3.650 lemetatik ia erdiak (1.762 lema, % 48,27) corpus bakar batean agertzen dira. Lema gehienak (1.285) ZT corpusean bakarrik agertzen dira, baina corpus txikietan ere aurkitu ditugu lema eskusiboak. Bestalde, azken zutabeetan ikus daitekeenez, datu hauen esangarritasuna areagotu egiten da corpusaren tamaina hartzen badugu kontuan. Izan ere, ZT corpusa, adibidez, EPG baino hiru aldiz txikiagoa da. Datu horren arabera, lema eskusiboen kopuruak ZT corpusean du pisurik handiena; atzetik, TB eta IT corpusak datoz, eta EPG zein KLS corpusetan datu hau ez da oso adierazgarria. Datu hauek erakusten dute gaia zenbat eta espezifikagoa izan, eta testuak zenbat eta espezializatuagoak eta abstraktuagoak izan, orduan eta premia handiagoa dagoela aditzetiko izenak sortzeko eta ustiatzeko. Emaizta honek berresten du Lersundi, Zabala eta Elorduiren (2008) lanaren ondorioetako bat.

Corpusa	Lema eskusiboen kopurua	%	L <sup>226</sup>
ZTC	1.285	35,21	150,65
EPG	397	10,87	15,82
KSL	54	1,48	5,59
TB	22	0,60	34,22
IT	4	0,11	22,43

3. taula: lemen banaketa corpus desberdinetan zehar

Aurreko atalean esan dugunez, corpus orokorren eta corpus espezializatuen arteko erkaketa egiteko Mann-Whitneyren proba estatistikoa erabili dugu. Proba horretan hirugarren taulan islatu ditugun emaitzak lortu ditugu. Emaizta horiek agerian uzten digute lema ezberdinen kopuruei dagokienez corpus orokorren eta corpus espezializatuen arteko aldeak

<sup>226</sup> L: [(lema kopurua / corpusaren tamaina) x 1000000]



esangarriak direla atzizki guztien kasuan. Maiztasunaren datuei dagokienez, aldiz, zenbait atzizkiren kasuan bakarrik antzematen da alde esangarria corpus orokorren eta corpus espezializatuen artean, alegia, *-zio*, *-keta*, *-men* eta *-dura* atzizkien kasuan. Hala ere, kontuan hartzekoa da corpus espezializatuen artean askotariko jakintza-alorretakoak daudela, eta espero izatekoa dela gaien heterogeneotasunak datuen nolabaiteko dispersioa ere eragitea. Ildo horretatik, nabarmentzekoa da corpus espezializatuak hain heterogeneoak izanik ere, lema-kopuruei dagozkien alde horietako batzuk (aipatutako atzizkiei dagozkienak, hain zuzen) esangarriak izatea.

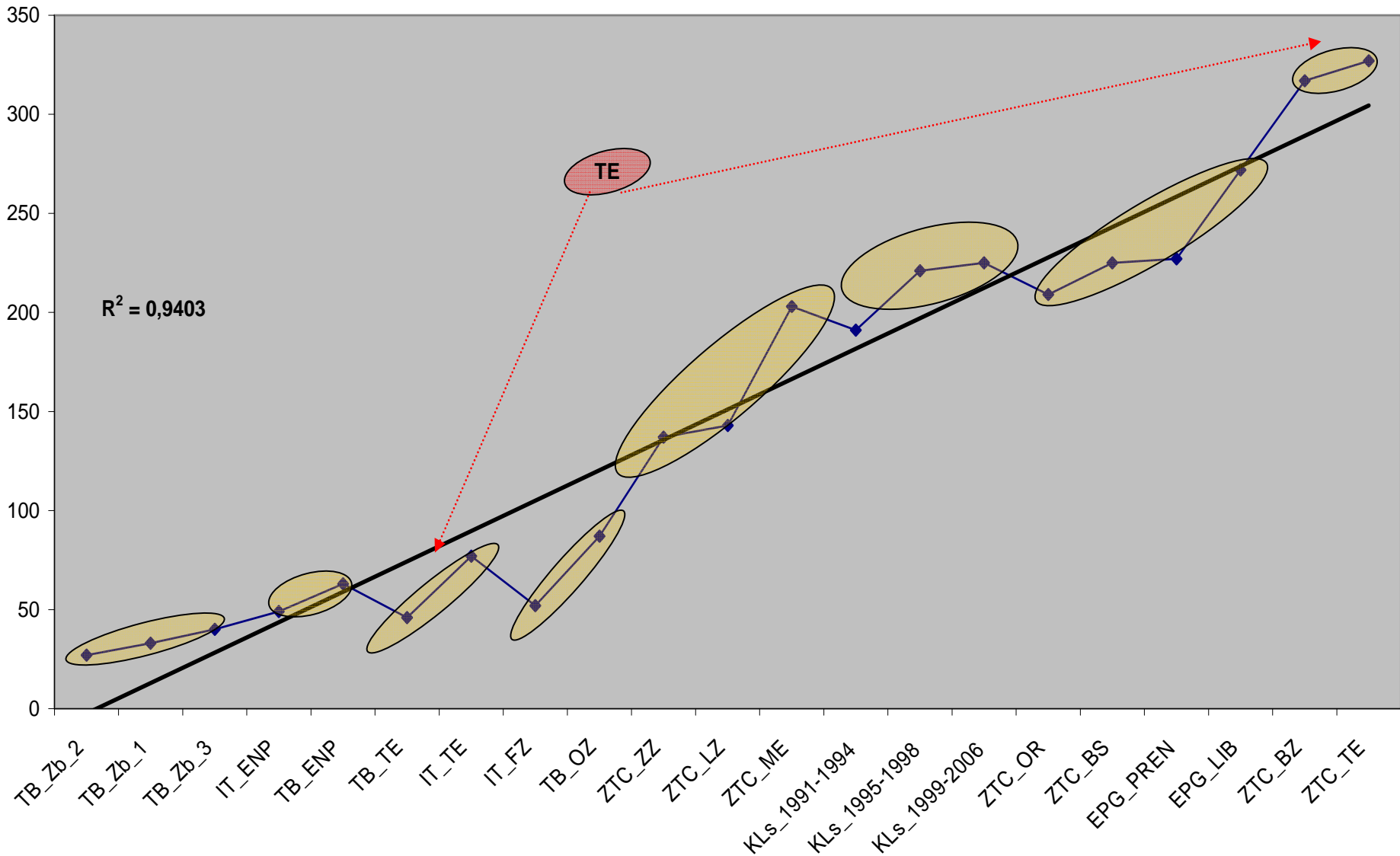
	<b>ERA</b>	<b>KUNTZA</b>	<b>KETA</b>	<b>MENDU</b>	<b>KUNDE</b>	<b>MEN</b>	<b>PEN</b>	<b>DURA</b>	<b>TZA</b>	<b>ZIO</b>
N (Sig.)	0.001	0.004	0.010**	0.002**	0.000**	0.000**	0.006**	0.004**	0.001**	0.007**
M (Sig.)	0.110	0.535	0.000**	0.535	0.971	0.038*	0.17a	0.000**	0.743	0.000**

4. taula: corpus orokorren eta corpus espezializatuen arteko aldean esangarritasun-indizeak Mann-Whitneyren probaren arabera.

Azterketa honen bigarren fasean egiaztatu nahi izan dugu ea izen ebentiboak eratortzen dituzten atzizki hauen banaketak eta joerak alorreko testuek jorratzen dituzten gaiekin zerikusia ote duten. Bigarren eta hirugarren irudietan ikus daiteke *-keta* atzizkirako kalkulaturako N eta Mren datuen erkaketa.

Bigarren irudian *-keta* atzizkidun lemen kopuruak jasotzen dira. Ikus daitekeenez, jakintza-arlo bereko laginak ardatzaren segmentu berdintsuan kokatzen dira, eta datuak hein handi batean hurbiltzen dira banaketa lineal batera. Izan ere, erregresio linealaren indizearen datua ( $R^2$ ) oso esangarria da, 0,90etik gorakoa baita.

**KETA atzizkidun lema ezberdinen kopurua**

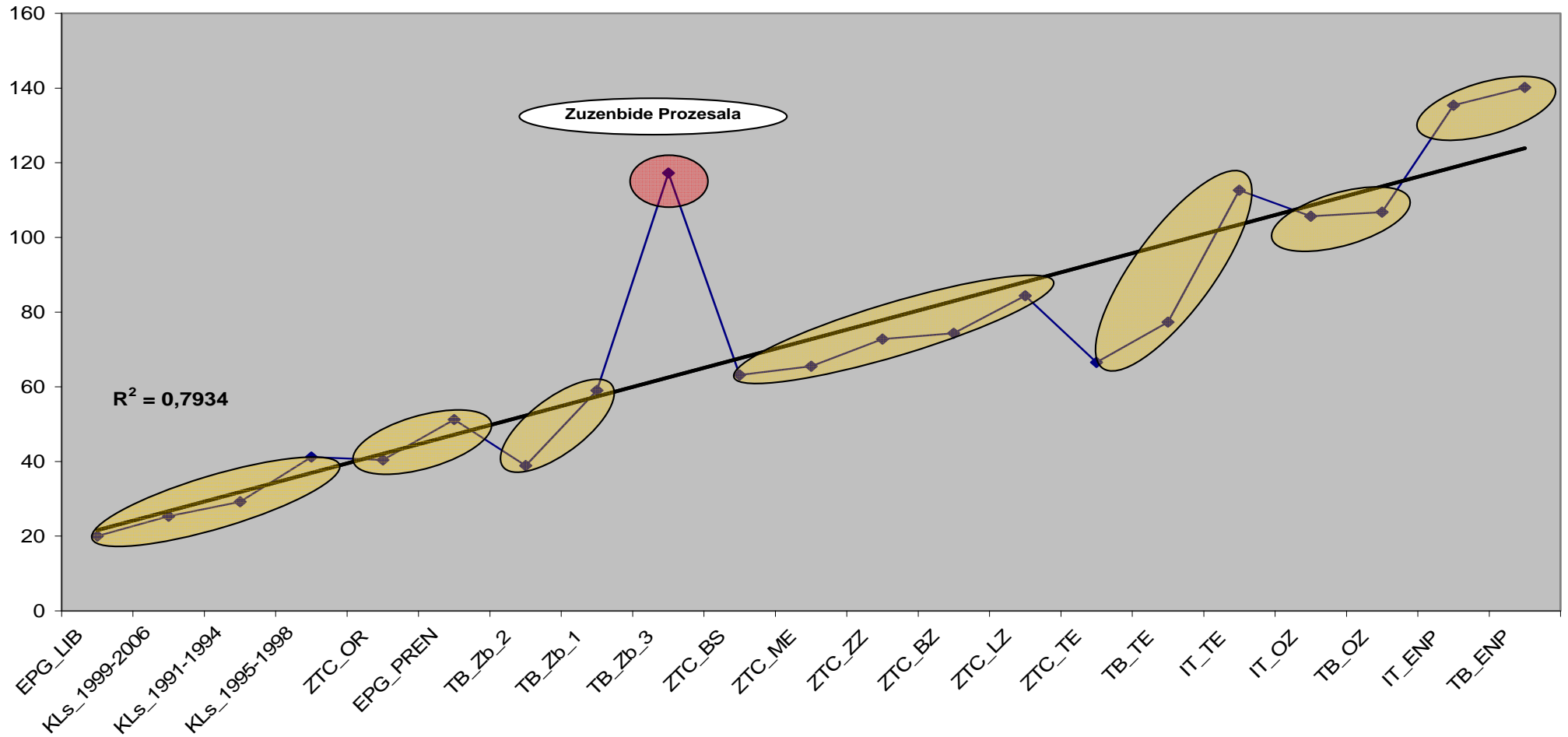


2. IRUDIA: -KETA ATZIZKIARI DAGOZKION N BALIOEN BANAKETA CORPUS DESBERDINETAN ZEHAR

Atzizki guztien kasuan lortu ditugu antzeko banaketak Edonola ere, espero genuen moduan, ia kasu guztietan ikusi dugu laginen bat edo beste bere multzoko besteetatik nolabait aldentzen dela. *-keta* atzizkiaren Nren datuen banaketan, adibidez, arlo teknikoko lagin bat da beste bietatik aldentzen dena. Horren arrazoiak ugariak izan daitezke: laginen tamaina, lagina osatzen duten testuen gaiak, testuen iturriak e.a. Batzuetan azalpena begien bistakoa izan daiteke, baina beste batzuetan jokabide berezi hori argitzeko ikerketa kualitatibo sakonagoak behar direla uste dugu.

Adibidez, 3. irudian ikusten dugu zuzenbidearen arloko lagin batean beste bietan baino nabarmen maizago erabiltzen direla *-keta* atzizkidun lemak. Aurreko irudian ikusi dugunez, Nren datuen aldetik berdintsuak dira hiru laginak, baina, horietako batean lema horiek maizago erabiltzen dira. Bereizi egiten den lagin honek zuzenbide prozesaleko testuak biltzen ditu. Hortaz, ez da harrigarria lagin honetan *epaiketa*, *ikerketa*, *atxiloketa*, *miaketa*, *galdeketa*, *salaketa*, *giza hilketa*, *arakaketa* eta antzeko lemak zuzenbidearen beste alor batzuetako testuak biltzen dituzten lagin batzuetan baino maizago azaltzea.

**KETA atzizkidun lemen maiztasuna (M: agerpen-kopurua 10000 hitzeko laginean)**



3. IRUDIA: -KETA ATZIZKIARI DAGOZKION M BALIOEN BANAKETA CORPUS DESBERDINETAN ZEHAR

Atzizkien corpusetan zeharreko banaketa horien tinkotasuna eta esangarritasuna finkatzeko, erregresio linealaren indizea kalkulatu eta aldagai bakarreko ANOVA proba egin dugunean, 5. taulan bildutako datuak lortu ditugu. Taulan ikus daitekeenez, emaitzak erabatekoak izan dira. Izan ere, bai N eta bai M indizeen banaketak estatistikoki esangarriak dira. Guztiz tinkoa ez dirudien banaketa bakarra *-kunde* atzizkiaren maiztasun-datuena da. Edozein modutan ere, lehen ikusi dugunez, *-kunde* da sistema honetan pisu txikiena duen atzizkia: lema gutxi ematen ditu eta lema horiek maiztasun txikia dute oro har corpus guztietan. Ondorioz, corpusaren tamainaren arabera, *-kunde* atzizkiaren maiztasun-datuen arteko aldea handi samarra izan daiteke lagin batzuen eta beste batzuen artean, lagin horiek guztiak jakintza-arlo berekoak izan arren.

Edozein modutan, *-kunde* atzizkiaren agerpen-maiztasunari dagokion Rren balioa esangarritasunaren mugan dago. Horrela bada, badirudi lortutako emaitzek berresten dutela gure lanaren bigarren hipotesia, hau da, atzizki hauek dituzten joerak alor bakoitzeko testuek jorratzen dituzten gaiekin lotuta daudela.

	Lema ezberdinen kopurua		Agerpen-maiztasuna	
	R	Sig.	R	Sig.
ERA	0.708	0.00**	0.584	0.05*
KUNTZA	0.722	0.00**	0.652	0.01*
KETA	0.610	0.03*	0.518	0.02*
MENDU	0.767	0.00**	0.682	0.01**
KUNDE	0.817	0.00**	0.407	0.06
MEN	0.789	0.00**	0.676	0.01**
PEN	0.714	0.00**	0.635	0.02*
DURA	0.691	0.01**	0.407	0.07
TZA	0.752	0.00**	0.644	0.02*
ZIO	0.650	0.01**	0.580	0.05*

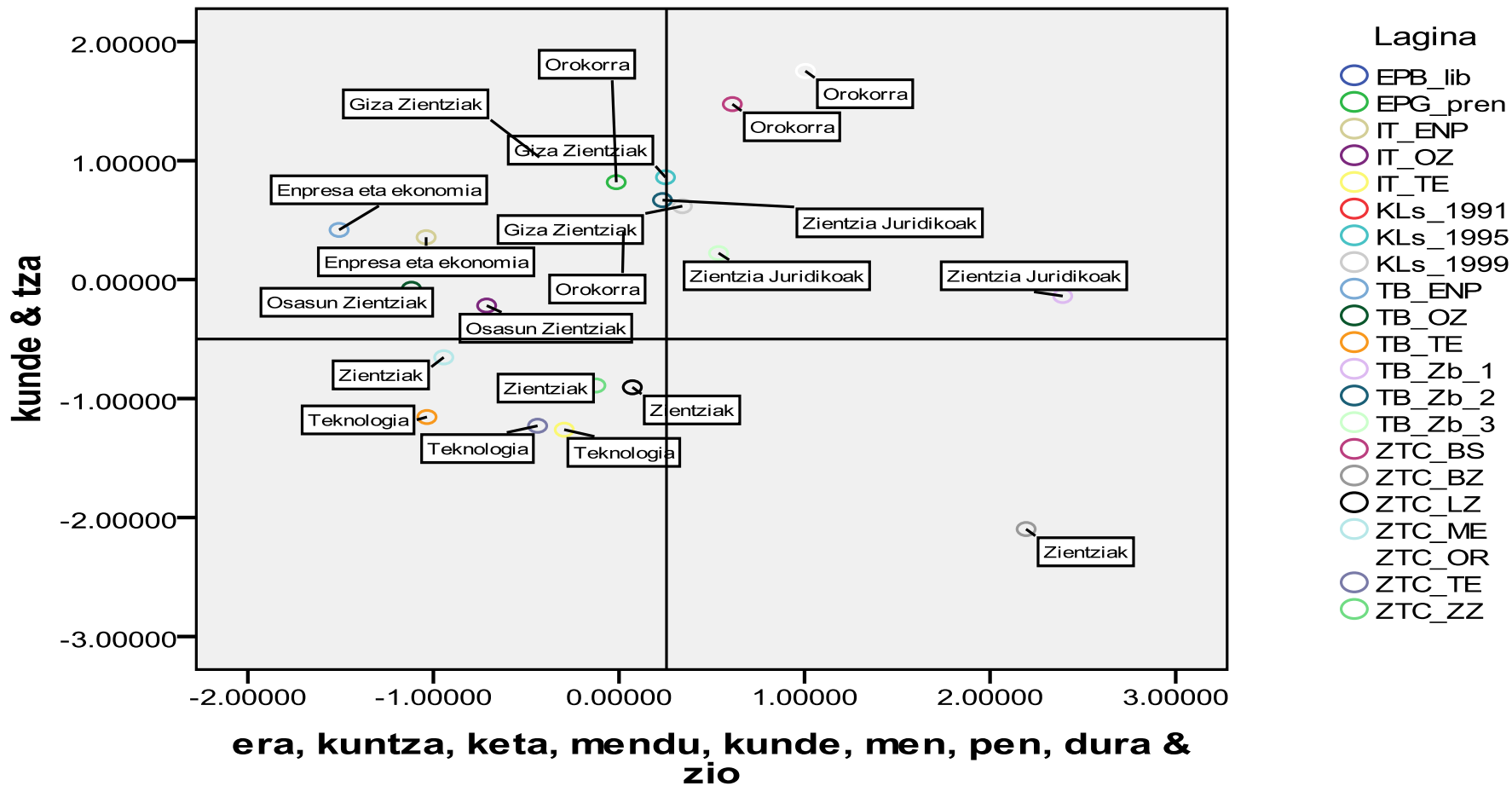
5. taula: Nren eta Mren banaketen tinkotasunaren neurria (erregresio-indizea)

Hirugarren hipotesia egiaztatze aldera, giza eta gizarte-zientzietako corpusak corpus tekniko-zientifikoekin eta corpus orokorrekin alderatu ditugu analisi faktorialaren bitartez. Lortu ditugun emaitzek, 4. irudian ikus daitezkeen moduko banaketak erakutsi dizkigute.

Bistan denez, banaketa faktorialaren goiko sektorean corpus orokorrak eta giza eta gizarte-zientzietakoak biltzen dira eta, beheko sektorean, zientzia eta teknologiakoak. Osasun zientzietakoak bien artean kokatzen dira. Ezkerreko koadrantean, bestalde, giza zientzietako, enpresa zientzietako eta EPG corpuseko testu orokorren laginak daude; eskuineko aldean, berriz, zientzia juridikoetakoak eta ZT corpuseko testu orokorrak. Esan daiteke, beraz, alorraren arabera, modu koherentean banatzen direla aztertutako lagin guztiak.

Grafiko hau Nren datuei dagokie, baina maiztasunaren datuen analisi faktorialak ere berresten du hipotesi hau, bertan corpusen banaketak beste egitura bat baldin badu ere.

### Corpusen banaketa faktoriala lema-kopuruen arabera

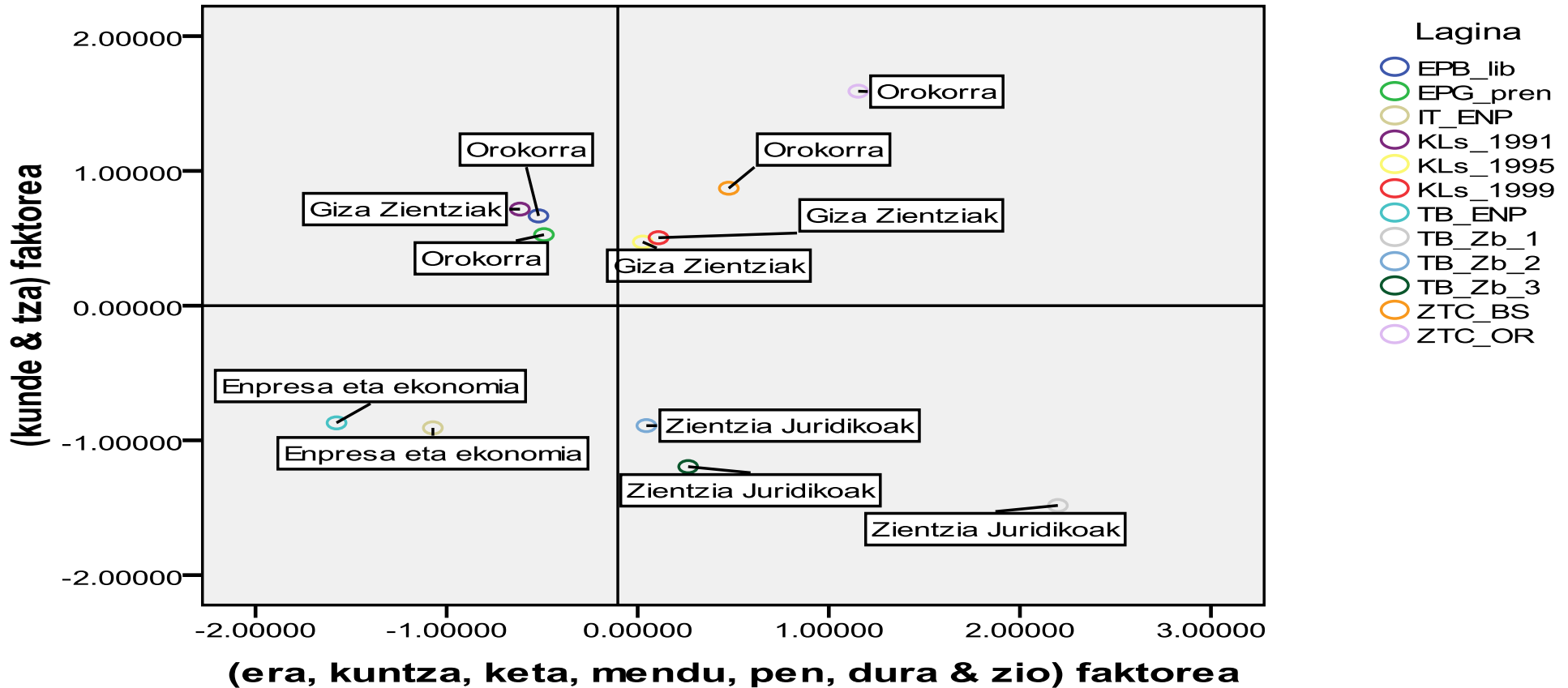


4. IRUDIA: NREN BANAKETA FAKTORIALA CORPUS GUZTIETAN

.Bestalde, giza eta gizarte zientzietako testuak eta zientzien alorrekoak isolatu ditugunean eta analisi faktorialaren emaitzak grafikoki irudikatu ditugunean, ikusi dugu banaketaren bereizmena areagotu egiten dela.

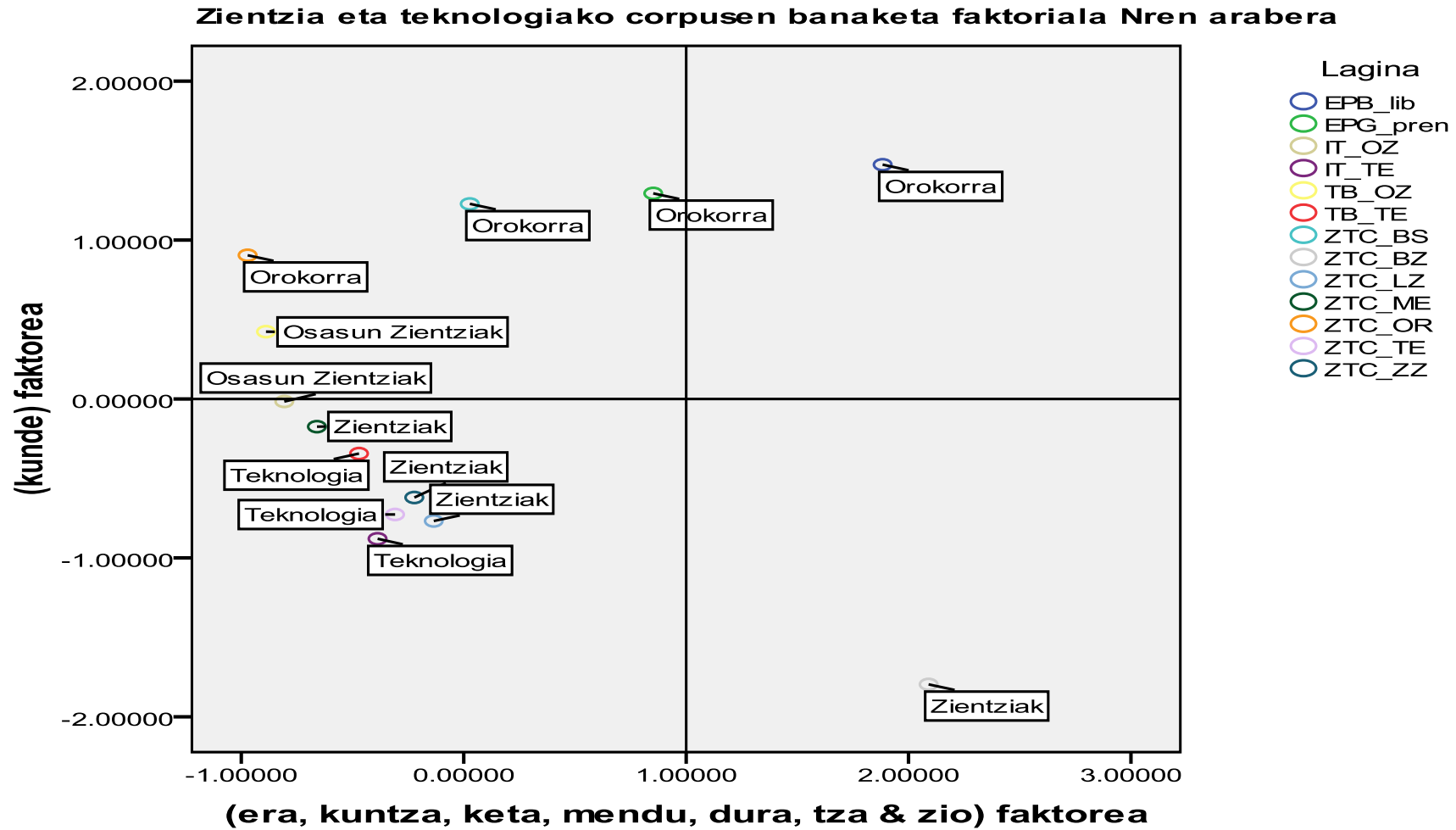
5. irudian ikus daitekeenez, giza zientzietako testuak eta testu orokorrak goiko sektoreetan agertzen dira eta, espezializatuak diren beste bi alorrak, aldiz, beheko sektorean agertzen zaizkigu. Corpus orokorrak ere, beste irudikapen guztietan bezala, banatuta agertzen zaizkigu: EPG corpusari dagozkionak maparen gune batean eta ZT corpusari dagozkionak beste batean. Zientziaren alorreko testuak isolatu ditugunean ere bereizmen handiagoko banaketa agertzen zaigu (ikus 6. irudia). Hortaz, zantzu guztiek berresten dute laugarren hipotesia ere.

### Giza eta Gizarte Zientzietako Corpusen banaketa faktoriala Nren arabera



5. IRUDIA: NREN BANAKETA FAKTORIALA: GIZA ZIENTZIEN ALORREKO CORPUSAK VS CORPUS OROKORRAK





6. IRUDIA: NREN BANAKETA FAKTORIALA: ZIENTZIA ETA TEKNIKAREN ALORREKO CORPUSAK VS CORPUS OROKORRAK

Laburbilduz, datuei aplikatu dizkiegun proba estatistikoen emaitzek egiaztatzen dituzte bigarren atalean lanaren abiapuntutzat eman ditugun hipotesiak.

## 5. ONDORIOAK

Bi ondorio nagusi nabarmendu nahi ditugu lan honi bukaera emateko:

1. Aditzetiko izenen erabilera-maiztasuna eta aditzetiko izenak sortzeko orduan egiten den atzizkien hautapena erregistroekin lotuta dago. Zehazkiago, esan daiteke hiztunek egiten dituzten aukerak lotuta daudela testuen espezializazio mailarekin, eta testuek jorratzen duten espezialitate-alorrekin eta gaiarekin. Honetan euskarak bat egiten du beste hizkuntza batzuetan deskribatu diren joerekin (Cademártori et al., 2006 eta Guz, 2009, besteak beste).
2. Izen ebentiboak eratortzen dituzten atzizkien banaketa ez da azarezkoa: atzizki horiek sistema bat osatzen dute.

Bi ondorio horiek kontuan hartuta, tentu handiz jokatu behar da sistema honen gaineko edozein interbentzio egin aurretik. Funtsezkoa deritzogu banaketaren motibazio semantiko-pragmatikoak aztertzeari. sistema desorekatu dezakeen interbentziorik egin nahi ez badugu.

## Bibliografia

- BIBER, D. (2004): *Dimensions of Register Variation. A Cross-linguistic Comparison*, Cambridge: Cambridge University Press.
- BIBER, D. (2006): *University language: A corpus-based study of spoken and written registers*, Amsterdam: John Benjamins.
- CADEMÁRTORI, Y., PARODI, G. & VENEGAS, R. (2006): “El discurso escrito y especializado: caracterización y funciones de las nominalizaciones en los manuales técnicos”, *Literatura y Lingüística*, 17: 243-265.
- FERGUSON, Ch. a. (1983): “Sport announcer talk: syntactic aspects of register variation”, *Language in Society*, 12, 15-30 or.
- GALLEGOS, A. (2000): “Morfología y registro. Algunas relaciones entre tradiciones discursivas y morfología derivativa nominal en español”, *Función* 21-24:1, 142-215.
- GALLEGOS, A. (2003): *Nominalización y registro técnico. Algunas relaciones entre morfopragmática, tradiciones discursivas y desarrollo de la lengua en español*, Doktoretza-tesia, Albert-Ludwigs- Universitat.
- GUZ, W. (2009): “English affixal nominalizations across language registers” *Poznań Studies in Contemporary Linguistics* 45(4): 461–485.
- ITURRIOZ, J.L. (1985): “Abstracción substantiva: reificación de contenidos proposicionales”, in MELENA J.L. (arg.) *Symbolae Ludovico Mitxelena septuagenario oblatae*, Gasteiz: EHU, 396-414.
- ITURRIOZ, J.L. (2000): “Diversas aproximaciones a la nominalización. De las abstracciones a las macrooperaciones textuales”, *Función* 21-24:1, 32-140.

- LERSUNDI, M., ZABALA, I. eta ELORDUI, A. (2008): “Aditzetiko izenen emankortasunaren azterketa morfopragmatikoa euskarazko corpus orokor eta berezitueta”, in ARTIAGOITIA, X. eta LAKARRA, J.A. (argit.): *Gramatika jaietan Patxi Goenagaren omenez*, ASJUren gehigarriak, LI, Bilbo: EHUKo Argitalpen Zerbitzua, 491-506 or.
- PORZIG, W. (1930): “Die Leistung der Abstrakta in der Sprache” *Blätter für deutsche Philosophie*. 4: 66-67
- URE, J. Ch. a. (1982): “Introduction: approaches to the study of register range”, *International Journal of the Sociology of Language*, 35, 5-23 or.
- ZABALA, I. eta ELORDUI, A. (2006): “Specialized Discourse and the Linguistic System of Basque”, in B. Fernández eta I. Laka (argit.): *Andolin gogoan. Essays in honour of Professor Eguzkitza*, Bilbo: EHUKo Argitalpen Zerbitzua, 233-249 or.
- ZABALA, I. (2007): “Izen-sintagmaren egituratik nominalizazio-estrategietara”, Donostiako ikastaroak: *Izen-sintagmaren egitura euskaraz: sintaxia eta interpretazioa* Nazioarteko Elkargunea. 2007ko ekainaren 18 eta 19an (eskuizkribu argitaratu gabea).
- ZABALA, I. eta ELORDUI, A. (2010): “Nominalizazioa estrategia diskurtsibo gisa: corpus akademikoaren azterketaren beharra”, in ALBERDI, X. eta SALABURU, P. (argit.) (2010): *Ugarteburu terminologia jardunaldiak: euskararen garapena esparru akademikoetan. Espezialitate hizkerak eta terminologia IV*, Euskal Herriko Unibertsitatea.

[AURKIBIDEA](#)