

**Ignacio
Alustiza
Bacchella**

ignacioalustiza47@gmail.com

Ikerketak

HEOCeko adierazpide linguistiko batzuek B2 eta C1 azterketetako idazlanetan duten isla. HABE-IXA corpusa aztergai¹

Lan honetan, Helduen Euskalduntzearen Oinarrizko Curriculumeko adierazpide linguistiko zehatz batzuek benetako ekoizpenetan zer nolako agerpena eta erabilera duten ezagutzeko burutu den ikerketa aurkeztuko da. Egiaztatze-prozesuetan aritu izan diren hainbat adituk HEOCeko adierazpideen artean B2 eta C1 mailetan erabiltzeko desiragarri diren 5 fenomeno lehenetsi ondoren, HABE-IXA corpuseko 240 idazlan aztertu dira. Horretarako Markin tresnaren bidez aukeratutako adierazpide linguistikoen agerpen-kopurua, erabilera zuzenak eta erabilera okerrak eskuz etiketatu dira. Honela, teoriak (HEOCak eta adituek) diotena eta ekoizpen errealek islatzen dutenaren arteko aldea enpirikoki aztertuz, bai hizkuntzalaritza konputazionalaren alorrean interes orokorra dutenentzat, bai euskararen irakaskuntza eta ebaluazio esparruetan jarduten duten eragileentzat, ekarpen bat egitea da artikulu honen xede nagusia.

Gako hitzak

adierazpide linguistikoak, corpusak, etiketatzea, idazlanak, irizpideak

1. SARRERA

2020an HABEk eta EHUKo IXA ikerketa-taldeak lankidetza-hitza zuten, besteak beste, IXAk garatutako teknologiak eta hizkuntza naturalaren prozesamendua-
ren alorrean duen esperientzia baliatuz HABEko testuak eta corpusak prozesatu eta
datu-baseratzeko, ondoren ikasleen ekoizpen nahiz azterketen zuzenketetatik eratorri
daitezkeen ikerketak sustatzeko helburu orokorrarekin².

¹ Artikulu honen oinarria den lana Ignacio Alustizak HABE erakundean egindako bekaldian burutu da.

² https://www.habe.euskadi.eus/contenidos/informacion/lankidetza_hitzarmenak/eu_def/adjuntos/HABE-eta-EHUK-onartua-GARBIA-Hitzarmena_eus.pdf

IXA ikerketa-taldeak ibilbide oparoa du euskarazko corpusen garapenaren eremuan. Hala, azken hamarkadetan hainbat motatako corpusak sortu ditu, esaterako, corpus historikoak (Estarrona et al., 2020), sintaktikoki etiketatuak (Aduriz et al., 2002) eta corpus eleaniztunak. Izan ere, hizkuntzalaritza konputazionalen geroz eta garrantzi handiagoa dute corpus linguistikoek, hizkuntzaren erabileraren adibide errealean multzo zabalak eta egituratuak baitira. Adibide horiek idatzizko testuetatik edo ahozko adierazpenetatik lor daitezke; galdera berriak eginez, hizkuntzaren erabileraren eta bariazioaren ikerketa praktikoak egitea ahalbidetzen dute corpusek (Friginal, 2017). Euskarari dagokionez, hizkuntzalaritza konputazionalen diharduten ikertzaileek azpimarratzen dute aro digitalean bizirauteko ezinbestekoa dela euskarazko corpusak ugaritzea eta handiagotzea, eta euskara aztertze baliabideak sortzea, baita helduen hizkuntza i(r)a kaskuntzan lagunduko dutenak ere (Arrieta et al., 2023).

Testuinguru horretan, euskara-ikasleen ekoizpenak biltzen dituen corpusik ez zegoela ikusi zuten HABEk eta IXAk. Horregatik, funtsezkotzat jo zuten urteetan zehar egiaztatze-prozesuetan jasotako idazlanen multzoari corpus baten forma ematea, azterketek balio handia baitute eta hizkuntzaren prozesamendurako oso garrantzitsuak izan baitaitezke. Ikasleen ekoizpenak eta azterketak euskararen ahozko zein idatzizko erabileraren iturri paregabeak diren heinean, hizkuntzaren prozesamendu-tresnak erabiliz posible dugu horiek ustiatzea eta ondorioak ateratzeko hausnarketak egitea. Ikasleen euskararen erabilera erreala ezagutzea irakasleentzat ere garrantzitsua izan daiteke, esate baterako, tarteko hizkuntza eta ikasleek ekoiztutako testuen kalitatea eta idazleen ezaugarriak jakiteko (Arrieta et al., 2023).

1. taula. HABE-IXA corpusaren idazlan-kopuruak maila eta kalifikazio orokorraren arabera

HABE-IXA corpusa	GAI	EZ GAI	GUZTIRA
B1	80	40	120
B2	80	40	120
C1	80	40	120
C2	47	73	120

Egiaztatze-gintza-prozesuetako testu-bildumak aztertu, egituratu eta fitxategien transkripzioak lortzeko lana egin zen HABE eta IXA taldearen arteko hitzarmenaren hasieran eta, lan horren ondorioz, HABE-IXA corpusa CC BY-NC 4.0 lizentziarekin eskuragai da egun (Azpillaga, 2022). Euskarazko idazmen-proben corpusa sortzeko, honako urratsak egin ziren (Arrieta et al., 2023):

- testuak digitalizatzea;
- karaktere-odeketa;
- garbiketa;
- datu-basea egituratzea.

Horrela, HABE-IXA corpusa euskara-maila egiaztatze-deialdietan jasotako 480 idazlanek eta horien ebaluazioek osatzen dute (ikus 1. taula). Horietako 240 aztertu ditugu lan honetan, hau da, B2 eta C1 mailetako testu oro.

2. IKERLANAREN GARAPENA

2.1. PLANTEAMENDUA

Lan honen helburua teoriaren eta benetako ekoizpenek islatzen dutenaren arteko alde-raketa egitea da. Horretarako, bi helburu espezifiko ditugu. Bata, deskriptiboagoa, hainbat adituk aurretik zehaztutako adierazpide linguistiko batzuen agerpenak aztertzea, erabilera zuzenen eta okerre kopuruei erreparatu, eta hala, adierazpide horiekiko espero den kontrol-mailarik dagoen ikustea. Bestea, adierazpide horien erabilerak kalifikazioetan duten eragina aztertzea, *aberastasuna* irizpidean fokua jarrita. Puntu honetan baieztatu nahi da ea, adituek dioten bezala, adierazpide linguistiko horiek ondo erabiltzeak benetan eragiten duen *aberastasuna* irizpidea kalifikazioan.

IXA taldeko kideek aurretik egindako lanetan oinarritu gara hein handi batean (Arrieta et al., 2023; Arriola et al., 2020; Osinalde eta Iruskietak, 2022). Osinaldek eta Iruskietak ikerlan bat burutu dute non 47 etiketako sorta aplikatu dioten HABE-IXA corpuseko 40 idazlaneko lagin bati (B2ko 20 eta C1eko beste 20), eta ondorioztatu dute “interpretazio egokienak azterketa sakonagoa behar duela” (Osinalde eta Iruskietak, 2022, 14. or.). Oso lan interesgarria da; izan ere, testuak Markin erabiliz eskuz etiketatzeaz gain, paragrafoen etiketatze automatiko bat ere burutu dute. Gurean, aldiz, soilik eskuz etiketatu dira testuak, baina lagin handiago batekin. Erabiltzaile aurreratuari (B2) eta gaituari (C1) dagozkien ekoizpenetan jarri da fokua, deialdi jendetsuenak izateaz gain, maila horiek baitira gehien aitortzen direnak DBHko zein unibertsitate-ikasketak amaitutakoan.

2.2. ADIERAZPIDE LINGUISTIKOEN AUKERAKETA

Helduen Euskalduntzearen Oinarritzko Curriculumeko (Gonzalez de Txabarri et al., 2015) adierazpide linguistiko guztiak aztertzea ezinezkoa denez –denbora eta baliabide aldetik muga handiak daudelako–, beharrezkoa izan da horietako batzuei lehentasuna ematea. Aukeraketa egiteko, galdetegi bat bidali zitzaion zenbait adituri, HABEko azterketaldietan parte hartzen duten lau mailetako tutore eta ebaluatzaile ikertzaileei, hain zuzen. Behin betiko galdetegia bidali aurretik, HABEko Ebaluazio eta Egiaztatzerako Arloko teknikarien artean pilotatu zen.

Galdetegiaren behin-behineko bertsioan, HEOCeko adierazpide linguistiko batzuk aukeratu ziren, morfosintaxiarekin eta sintaxiarekin lotutakoak, eta bi galdera planteatu ziren:

1. galdera: Aditu zaren mailako idazlanetako ekintza komunikatiboei erantzuteko, zein adierazpide linguistikok du garrantzi gehien? (“Aberastasuna” parametroan, zein adierazpidek dute pisu handien)

A adierazpide-multzoa: morfosintaxia, oro har

B adierazpide-multzoa: sintaxia
2. galdera: Adierazpide linguistiko hauek zuzen edo oker erabiltzeak zer garrantzi du zure mailan, zuzentasunari dagokionez? (“Zuzentasuna” parametroan, zein adierazpidek dute pisu handien)

A adierazpide-multzoa: morfosintaxia, oro har

B adierazpide-multzoa: sintaxia

Lehen bertsioa HABEko teknikarien artean pilotatu ondoren, ondorioztatu zen. adierazpide linguistikoak garrantziaren arabera sailkatzea ez zela egokia. Izan ere, testuen *zuzentasuna* eta *aberastasuna* ebaluatzerakoan ez da halako hierarkiarik hobesten. Ondorioz, garrantziari buruz galdetu beharrean, irizpideen kontrol-mailari buruz egin zen galdera. Horrez gain, kasu-markekin, aditzarekin eta ortografiarekin lotutako adierazpide-multzoak gehitu ziren, argazki osatuagoa izateko. Beraz, galdetegiaren bigarren bertsioan, adierazpide linguistikoak bost multzotan banatu ziren:

- i. kasu-markak;
- ii. aditza;
- iii. aditz-izena;
- iv. sintaxia;

v. ortografia.

Honela zioen galderak:

Zure iritziz, zer neurritan menderatu beharko lituzke adierazpide linguistiko hauek maila bakoitzeko erabiltzaileak bere ekoizpen idatzietan? Adierazi kontrol horren maila, baxuenetik (1) altuenera (5).

Behin kontrol-maila adierazita, uste baduzu adierazpide hori zuzen erabiltzea aberastasun altuaren seinale dela, markatu + (Ab.) zutabeari dagokion laukitxoan

Galdetegiaren bertsio berria HABEko Ebaluazio eta Egiaztatzerako Arloko epaimahai-buruen artean zabaldu zen, eta adierazpideen zerrendan zenbait aldaketa txiki egin ondoren, tutore eta ebaluatzaile ikertzaileei bidali zitzaion behin betiko galdetegia. Guztira 20 lagunek erantzun zuten galdetegia, eta jasotako erantzunak adituen iritzia direnez, horrek bermea ematen digu, batetik, adierazpide jakin batzuen aukeraketarako oinarri bat izateko eta, bestetik, emaitza objektiboen (azterketetan identifikatutako adierazpide linguistikoak) eta subjektiboen (adituen erantzunak) arteko harremana aztertzeko. Hauek dira adituek lehenetsitako adierazpideak.

1. Mugagabea
 - a. Zer, zein eta zenbat galdetzaileekin mugagabea
 - b. Zenbatzaile zehaztugabeekin mugagabea
2. Nominalizazioa
 - a. t(z)ea espero/ erabaki/ pentsatu
 - b. t(z)eak, t(z)eari
3. Konparaziozkoak
 - a. Konparaziozko erromantikoa (gero eta ... - ago)
 - b. Konparaziozkoak ahalik eta...(e)n; zenbat eta ...-ago; hainbat eta ... -ago
4. Berezitasunik ez duten ohiko aditzen erregimena
 - a. (Etorri, joan, kexatu, oldartu)
 - b. (Jan, eskatu, erosi, saldu, jaso, hartu)
 - c. (Ohartu, konturatu, mintzatu, solastatu)
 - d. (Iruditu, kostatu)
5. Aditz problematikoen erregimena
 - a. (Gonbidatu, ezagutu, tratatu, kezkatu, arduratu)
 - b. (Ikusi, ekarri, eraman)
 - c. (Ekin, eraso, heldu, eutsi)

Ikerketa-adierazpideak aukeratzeaz gain, hasieratik pentsatu dugu maila eta kalifikazio orokorra ez ezik, irizpideak ere aintzat hartzea, analisi osagarri bat izate aldera. Horregatik, adituei zabaldutako galdetegietan *aberastasuna* irizpideaz galdetzen da, adierazpideak oker edo zuzen erabiltzeak aberastasunean izan lezakeen eraginari buruzko iritzia jasotzeko. Ariketa hau ezinezkoa zaigu testuaren *koherentzia* eta *kohesioa* irizpideekin; are gutxiago *egokitasuna* parametroarekin, ez baitakigu zein izan zen azterketetan proposatutako kontsigna originala.

1. irudia. HABE-IXA corpusaren fitxategiak metadatuekin

HABE_IXA-224-B2-2.txt,B2,GAI_nahiko,B,C,C,C,C
HABE_IXA-225-B2-2.txt,B2,GAI_nahiko,B,B,C,C,C
HABE_IXA-226-B2-1.txt,B2,EZ_GAI,C,C,C,C,D
HABE_IXA-227-B2-2.txt,B2,GAI_nahiko,B,B,C,C,C
HABE_IXA-228-B2-2.txt,B2,GAI_nahiko,C,C,C,C,C
HABE_IXA-229-B2-2.txt,B2,GAI_nahiko,C,C,C,C,C
HABE_IXA-230-B2-3.txt,B2,GAI_ondo,A,A,A,A,A
HABE_IXA-231-B2-1.txt,B2,EZ_GAI,C,C,C,C,C
HABE_IXA-232-B2-1.txt,B2,EZ_GAI,C,C,D,D,D
HABE_IXA-233-B2-1.txt,B2,EZ_GAI,C,C,C,C,D
HABE_IXA-234-B2-1.txt,B2,EZ_GAI,C,C,C,D,D
HABE_IXA-235-B2-2.txt,B2,GAI_nahiko,C,C,C,C,C
HABE_IXA-236-B2-2.txt,B2,GAI_nahiko,B,C,C,C,C
HABE_IXA-237-B2-1.txt,B2,EZ_GAI,C,D,D,C,D
HABE_IXA-238-B2-1.txt,B2,EZ_GAI,C,C,D,C,D
HABE_IXA-239-B2-3.txt,B2,GAI_ondo,B,B,B,B,B
HABE_IXA-240-B2-3.txt,B2,GAI_ondo,B,B,B,B,B
HABE_IXA-241-C1-1.txt,C1,EZ_GAI,B,B,C,C,C
HABE_IXA-242-C1-1.txt,C1,EZ_GAI,C,C,D,D,D
HABE_IXA-243-C1-1.txt,C1,EZ_GAI,D,C,D,C,D

HABE-IXA corpuseko idazlan bakoitza gutxienez bina aztertzailek kalifikatu dute, itsu bikoitzeko sistemaren bidez. Desadostasuna egonez gero, tutore batek erabaki du azken kalifikazioa, bi aztertzaileen proposamenak kontuan izanda (Arrieta et al., 2023, 6. or.). 1. irudian ikusten denez, HABE-IXA corpusean idazlan bakoitzari fitxategi bat dagokio, honako antolaera errespetatuz:

- fitxategiaren zenbakia (225, 226, etab.);
- maila (B1, B2, C1, eta C2);
- kalifikazio orokorra (“ez_gai”, “gai_nahiko”, eta “gai_ondo”);
- HEOCeko irizpideak (egokitasuna, koherentzia, kohesioa, aberastasuna eta zuzentasuna, hurrenez hurren) beren puntuazioarekin (A,B,C,D eta E).

Hortaz, corpusak ematen du kalifikazio jakin batzuk dituzten testuak aztertu eta alderatzeko aukera eta hemen abantaila hori aprobetxatu nahi izan dugu.

2.3. IDAZLANAK ETIKETATZEA

Fase enpirikoa aurrera eramateko 2024ko uztailetik irailera bitartean B2 eta C1eko ehun eta hogeina idazlan aztertu eta eskuz etiketatu ziren Markin tresna baliatuta. Idazlan bakoitza etiketatu ondoren, bi aldiz errepasatu zen, gure intereseko adierazpide bakar bat ere markatu gabe gera ez zedin. Adierazpide bakoitzaren balizko erabilera oker eta zuzen asko dituzten zerrendak prestatu ziren landa-lanari ekin baino lehen, zalan-tza-egoeratan euskarri gisa balio zezaten.

2. irudia. Markin tresnarekin aztertutako idazlan bat

The screenshot displays a web-based interface for analyzing linguistic data. At the top, there are tabs for 'Normal view', 'Statistics', 'All annotations', and 'Marking tree'. The main area shows a text document with various linguistic annotations in different colors (green, yellow, red). Below the text, there is a 'Statistics' section with a table summarizing the data.

Annotation	Explanation	Help link	Categories	Value	Points lost	Points gained
1 NORMALIZADIA			Grammar	-1	-1	
2 ADITZ PROBLEMATIKOEN ERREGIMENA_2				1		2
1 MUGAGABEA_2			Content	1		1
			Totals	1	-1	3

Etiketatzeko-prozesua eragozpenik gabe burutu zen. Markin tresnak digitalizatutako fitxategiak banaka aztertzea eta etiketatzea ahalbidetzen du, eta etiketak komenigarriagoa den moduan moldatu ahal dira. 2. irudian ikusten denez, forma zuzenak zein okerrak identifikatzen eta markatzen dira –berdez eta gorritz, hurrenez hurren–, eta haien gaineko oharra ere egin daitezke; esaterako, erabilera okerra bada, adierazpide horrek izan beharko lukeen forma zuzena jar daiteke oharretan. Etiketatzeko amaitzean, programak estatistikak ateratzeko aukera ematen du; datuen ustiaketa-lana nabarmen errazten da horrela.

3. DATUEN ANALISIA

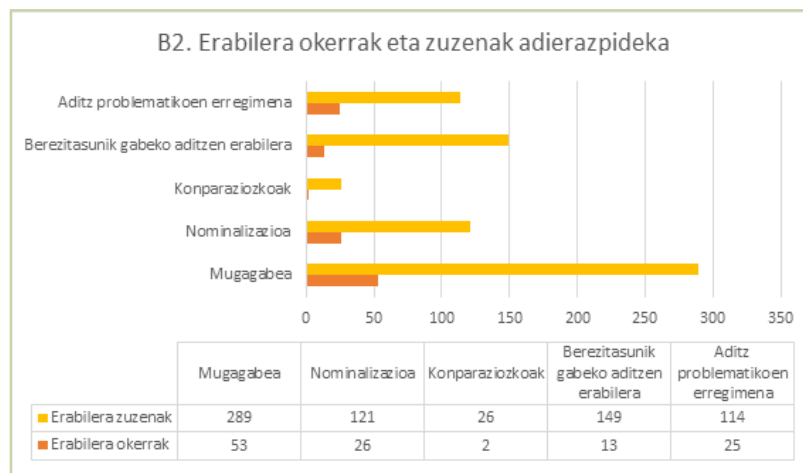
Lehenago azaldu denez, aztertu diren adierazpideak hautatzeko adituek adierazi dute B2 eta C1eko ekoizpenetan zer kontrol-maila beharko luketen ikasleek adierazpide horiekiko, 1etik 5erako eskala bat erabilita. Gero adierazpide hori ondo erabiltzea aberastasun altuaren seinalea den edo ez adierazi behar izan dute. Bi faktoreak aintzat hartuta, etiketatze-prozesuak utzitako datuen azterketari heldu diogu.

Orotara 1830 etiketa jarri dira, 818 B2an eta 1012 C1ean. Hau espero izatekoa zen, C1eko idazlanak B2koak baino luzeagoak baitira.

3.1. B2-KO IDAZLANAK

B2an 818 etiketa jarri dira, horietako 119 okerrak izanik (% 15).

3. irudia. B2. Erabilera okerrak eta zuzenak adierazpideka



Mugagabea

Bi bariedade bereizten dira: “Zer, zein eta zenbat galdetzaileekin mugagabea” eta “Zenbatzaile zehaztugabeekin mugagabea”. Dena den, bien kontrol-maila zein aberastasun altuaren seinale bezala berdintsu baloratu direnez, ez ditugu bereizi datuen B2ko testuen analisirako. Agerpen gehien duen adierazpidea da, eta espero den kontrol-maila islatzen dela esan daiteke.

Aberastasunari dagokionez, adituek esandakoa ikus daiteke idazlanetan: nahiz eta mugagabearen erabilera zuzen asko izan, testu batek ez du asko irabaziko *aberastasuna* irizpidearen kalifikazioan.

Nominalizazioa

Hemen ere bi bariedade daude: “-t(zea) espero, erabaki, eta pentsatu”, eta “-t(z)eak, -t(z)eari”. Kontrol-maila altuagoa espero dute adituek lehenengo kasuan. Eta hala da idazlanetan: erabilera okerrak % 5 dira. “-t(z)eak, -t(z)eari” moldeak, aldiz, askoz gutxiago erabiltzen da eta erabilera okerraren kopurua % 25 da.

Aberastasunari dagokionez, “-t(z)eak, -t(z)eari” bariedadea zuzen erabiltzea aberastasun altuaren seinaleztat jotzen da. Argi ikus daiteke “gai_ondo” testuetatik “ez_gai” testuetara pasatzean bai agerpen-kopurua bai erabilera zuzenen ehunekoak hazi egiten direla, eta horrek eragina duela *aberastasuna* kalifikatzean: lehenengoak “A” edo “B” izaten dute, eta bigarrenak, berriz, “C”.

Konparaziozkoak

Agerpen gutxieneko adierazpidea da 28 erabilera eskasekin, horietako 2 oker izanik. Bi bariedade daude: “konparaziozko erromanikoa (gero eta ...-ago)” eta “ahalik eta ...-(e)n; zenbat eta ...-ago; hainbat eta ...-ago”. Nahiz eta bi formak ia osorik zuzen erabiltzen diren, lehenengoarekiko kontrol-maila altuagoa espero da.

Aberastasunari dagokionez, adituek balorazio altuena esleitu diote “ahalik eta ...-(e)n; zenbat eta ...-ago; hainbat eta ...-ago” formari. Hau nolabait *aberastasuna* irizpidearen kalifikazioan islatzen da, nahiz eta kopuru txikiak izan: zuzen erabiltzen den bakoitzean irizpideak “C”-tik gorako kalifikazioa izaten du. Oker erabili denean, aldiz, “D” kalifikazioa izan du.

Berezitasunik ez duten ohiko aditzen erregimena

Adierazpide honek 4 bariedade ditu, baina baloratzeko orduan adituek ez dute bereizketarik egin. Beraz, datuen analisirako guk ere denak batera hartu ditugu. Esan daiteke badagoela harremana espero den kontrol-mailaren eta idazlanek islatzen dutenaren artean: 162 agerpenetik 149 zuzenak dira (% 92). Esleitutako aberastasun seinalea baxua da beste adierazpideekin konparatuta, eta testuek erakusten dutenaren arabera, baieztatu daiteke: adierazpide hau zuzen erabiltzeak ez duela zertan *aberastasuna* irizpidearen kalifikazioan eragin.

Aditz problematikoen erregimena

Aurrekoarekin gertatzen den antzera, adierazpide honek 3 bariedade eduki arren, denak batera baloratu eta aztertu dira. Adierazpide hau guztien artean kontrol-maila baxuenarekin baloratua da. Hala ere, erabilera zuzenen kopurua ez da batere mespretxagarria: 139tik 114 zuzenak dira (% 82). Erabilera zuzenen ehunekoak jaisteko joera du “gai_ondo” diren testuetatik “ez_gai” direnietara pasa ahala, kontrol-mailak kalifikazio orokorrean eragina duela erakutsiz.

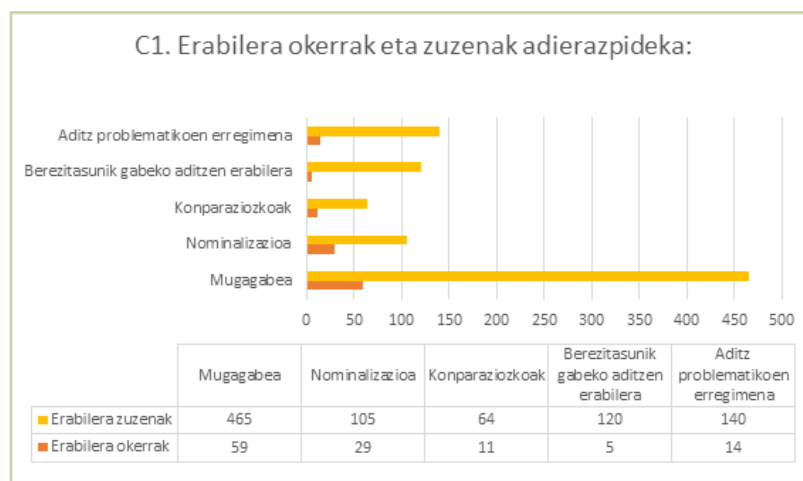
Aberastasunari erreparatuz gero, ez da ikusten korrelaziorik adierazpide hau zuzen erabiltzearen eta *aberastasuna* irizpidea ondo kalifikatzearen artean. Izan ere, aberastasun

altuaren seinaleztat jo da adierazpidea, baina badaude testu batzuk 4 agerpen zuzenekin “D” kalifikazioa dutenak eta “ez_gai” direnak; eta alderantziz: 4 erabilera oker izan arren *aberastasuna* irizpidean “C” kalifikazioa lortu duten testuak topa daitezke.

3.2. C1-EKO IDAZLANAK

C1ean 1012 etiketa jarri dira, horietako 118 okerrak izanik (% 12), B2an baino bat gutxiago. Gertatzen dena da erabilera zuzenen kopurua handitzen dela, testu luzeagoak direlako eta adierazpideekiko kontrol-maila orokorra hobea izaten delako maila honetan. Beraz, erabilera okerrean ehunekoa jaitea espero izatekoa zen.

4. irudia. C1. Erabilera okerrak eta zuzenak adierazpideka



Mugagabea

B2ko testuetan gertatzen den antzera, *mugagabea* da agerpen gehien dituen adierazpidea. C1eko idazlan batean espero den kontrol-mailak eta ustiatutako datuek bat egiten dutela esan daiteke: forma zuzenak % 88 dira. C1ean mugagabea ez da baloratzen aberastasun altuaren seinale bezala, eta hau nabaria da kalifikazioetan. Esaterako, 6 edo 7 erabilera zuzen izan arren, *aberastasuna* irizpidean “D” kalifikazioa duten hainbat testu topa daitezke.

Nominalizazioa

Adituek espero duten kontrol-maila altua ez da datuetan islatzen. Bi barietateak batera hartuta, erabilera okerrean ehuneko handiena dute (% 21). Ia portzentaje bera errepikatzen da bi barietateak banaka kontsideratuz gero: % 22.

Aberastasunari dagokionez ere ez da ikusten harremanik espero dena eta benetan gertatzen denaren artean: “-t(z)eak, -t(z)eari” forma zuzen erabiltzeak eragin handiagoa izan beharko luke “-t(zea) espero, erabaki, eta pentsatu” forma egoki erabiltzeak baino. Baina badirudi “ez_gai” zein “gai_nahiko” diren testu askotan “-t(z)eak, -t(z)eari” formaren erabilera zuzenek ez dutela eragiten irizpidearen kalifikazioan.

Konparaziozkoak

B2koetan bezala, agerpen gutxien duen adierazpidea da. Kontrol-maila altua espero da eta orokorrean egiaztatzen da: forma zuzenak % 87 dira. Barietateak banaka hartuz gero, “konparaziozko erromanikoa (gero eta ... -ago)” formak kontrol-maila altuagoa erakusten du (% 93 dira zuzenak), eta “ahalik eta ...-(e)n; zenbat eta ...-ago; hainbat eta ...-ago” barietateak islatzen duena baxuagoa da (% 71).

Aberastasunari dagokionez, ikus daiteke badagoela korrelazioa “ahalik eta ...-(e)n; zenbat eta ...-ago; hainbat eta ...-ago” nominalizazio forma zuzen erabiltzearen eta *aberastasuna* irizpidearen artean.

Berezitasunik ez duten ohiko aditzen erregimena

Espero den kontrol-maila altuena islatzen da datuetan: 125 agerpenetik, 5 erabilera oker bakarrik (% 4). Izan ere, B2 eta C1eko idazlanak batera hartuta ere erabilera okerrren ehuneko txikiena duen adierazpidea da hau. Ez da baloratu aberastasun altuaren seinale bezala eta hau ere ikusten da datuetan, ez baitago korrelaziorik zuzen erabiltzea eta irizpidearen kalifikazioaren artean.

Aditz problematikoaren erregimena

C1eko testuetan adierazpide honen erabilera okerrren portzentajea txikia da (% 9), espero den kontrol-maila egoki baieztatuz. Aberastasunaren seinaleari dagokionez, balorazio ertaina esleitu diote adituek. Horixe da datuetatik ondoriozta daitekeena. Adibidez, bai “ez_gai” bai “gai_nahiko” diren hainbat idazlanek “D” kalifikazioa lortu dute irizpidean, 4 edo 5 agerpen zuzen eta 0 oker dituzten arren. Adierazpide hau egoki erabiltzeak ez du zertan kalifikazio onik ekarri *aberastasuna* irizpidean.

4. ONDORIO GISA

Lan honen xedea izan da corpusgintza eta euskararen irakaskuntza esparruei ekarpena egitea, teoria eta praktikaren arteko harremana jorratuz. Alde batetik, HEOCean zehaztuta dauden adierazpide linguistikoaren artean batzuk lehenesteko eskaera luzatu genien HABEko egiaztatze-prozesuetan ibili izandako 20 aztertzaile eta tutoreei. Ondoren, aukeratutako 5 adierazpideak HABE-IXA corpuseko B2 eta C1eko idazlanetan duten isla aztertu dugu. Lehenetsitako adierazpideen agerpena eta erabilera aztertzeaz gain, idazlanak kalifikatzeko neurtu ohi den *aberastasuna* irizpideari ere erreparatu diogu.

240 testuko lagina aztertu eta gero, espero zitezkeen alderdi batzuk egiaztatu daitezke, hala nola, C1ean etiketa gehiago jarri izatea (idazlanak luzeagoak direlako) eta maila horretan bai nominaletan bai ehunekoetan erabilera zuzen gehiago egotea. Hala, egiaztatze-prozesuetan aritutako eragileek aukeratutako adierazpideekiko desiragarri den kontrol-maila neurri handi batean benetako ekoizpenetan islatzen da, bai B2an bai C1ean. Honek koherentzia erakusten du HABEren curriculumaren eta irakasle, tutore eta ebaluatzaileen prestaketaren artean, are gehiago kontuan hartzen bada adituen iritzia aurtan jaso dela eta corpuseko azterketa etiketatutak, aldiz, aurreko urteetakoak direla.

Hala eta guztiz ere, adierazpide bakoitza bakarka hartuta ikusten da datuak ez direla unibokoak. B2ko “aditz problematikoaren erregimena” eta C1eko “nominalizazioa” kasuek garbi erakusten dute batzuetan ez dagoela harremanik teoria eta praktikaren artean. Kasu hauetan espero zen kontrol-maila eta aberastasun altuaren seinalea ez dira ekoizpenetan islatzen. Gainera, ñabardura gehiago egoten dira adierazpideen barietateak aztertzen ditugunean: “nominalizazioa” eta “konparaziozkoak” fenomenoak dira egoera horren adibidea. Datuen analisisan salbuespenak topatzeak zuzenketa-prozesuetan irizpide objektiboak zehaztuta edukitzeaz gain alderdi subjektibo bat ere badagoela gogorarazten digu.

Dena den, kontziente gara ikerketa honek dituen mugaz. Esate baterako, estatistikaren alorrean “korrelazio” hitzak duen esanahia hemen zentzu malguagoan hartzen dugu. Ondorio sendagoetara iristeko, azterketa sakonagoak egin beharko lirake, ikerketa definitzeko aditu gehiagoren iritzia eta ikuspuntuak kontuan hartuz.

Hemen *aberastasuna* irizpideari begirada bat eman diogu. *Zuzentasuna* ere nolabait behatu da adierazpideekiko kontrol-mailaz aritu garenean. Begi onez ikusiko litzateke gai nontzeko irizpideak objektiboki aztertuko dituzten tresnak sortzea. Nola aztertu testu baten *kohesioa* edo *egokitasuna* tresna informatiko baten bidez?

Aurrera begira uste dugu ondo legokeela beste mota bateko ikuspegia aplikatzea HABE-IXA corpusari. Esaterako, alde zurreko galderekin ikerketa bat hasi eta gero erantzunen bila corpusera jo beharrean, hemen egiten saiatu garen bezala, corpusetik bertatik ikerketa-galderak sortuko dituen ariketa bat proposatzea. Hau da, corpuseko testuetan erregulartasunak aurkitzea eta hortik abiatuta galdera teorikoak pentsatzea. Horretarako denbora eta baliabideak behar dira, jakina. Agian, epe laburrean, adimen artifizialaren aurrerapenak corpusetan patroiak erraz topatzea ahalbidetuko dute. Ikusteke dago, beraz, teknologia berriek corpusgintzari ekarriko dioten onura.

Edozein modutan ere, euskarazko corpusak areagotu eta horien eskutik eztabaida zein ikerketa berriak loratuko direlakoan gaude, eta, hala, ikasleen eta azterketarien prestakuntzara bideratzea da erronka.

Bibliografia

- Aduriz, I. Aldezabal, I., Aranzabe, M., Arrieta, B., Arriola, J.M., Atutxa, A., Díaz de Ilaraza, A., Gojenola, K., Oronoz, M., eta Sarasola, K. (2002). Construcción de un corpus etiquetado sintácticamente para el euskera. *Procesamiento del Lenguaje Natural*, 29, 5-11.
<http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/3230/1721>
- Arrieta, E., Odriozola, I., Arregi, X. eta Iruskieta, M. (2023). HABE-IXA euskarazko idazmen-proben corpuseko idazlanen mailakatze automatikoa. *e-Hizpide*, 101.
<https://doi.org/10.54512/AGR9063>
- Arriola, J. M., Iruskieta, M. eta Goikoetxea, J. (2020). Hizkuntza teknologiak hizkuntzen ikas-irakaskuntzan: zenbat aukera hainbat erronka. *eHizpide*, 95.
<https://doi.org/10.54512/LOUF1436>
- Azpillaga, X. (2022). Euskarazko testuen komunikagaitasun-maila automatikoki sailkatzeko lehendabiziko urratsak. *e-Hizpide*, 99. <https://doi.org/10.54512/YZBR4203>
- Estarrona, A., Etxeberria, I., Etxepare, R., Padilla, M. eta Soraluze, A. (2020). Sintaktikoki etiketatutako euskarazko corpus historikoa eraikitzen. *Fontes Linguae Vasconum 50 urte. Ekarpen berriak euskararen ikerketari. Nuevas aportaciones al estudio de la lengua vasca.*, 237-251.
<https://doi.org/10.35462/fontes50urte.15>
- Friginal, E. (Ed.). (2017). *Studies in Corpus-Based Sociolinguistics* (1. argitalpena). Routledge.
<https://doi.org/10.4324/9781315527819>
- Gonzalez de Txabarri, J., Elozegi, L., Agirre, H., Larramendi, A., Maiz, A. eta Gonzalez de Txabarri, L. M. (2015). *Helduen Euskalduntzearen Oinarriko Curriculum (HEOC)*. HABE, Helduen Alfabetatze eta Berreuskalduntzerako Erakundea. <https://doi.org/10.54512/HEOC-2021>
- Osinalde, M. eta Iruskieta, M. (2022). Hizkuntza-ikasleen testu-corpus etiketatuaren analisisa eta interpretazioa B2 eta C1 mailetan. *e-Hizpide*, 100.
<https://doi.org/10.54512/HLFA9295>