



Lapurdum

Euskal ikerketen aldizkaria | Revue d'études basques |
Revista de estudios vascos | Basque studies review

16 | 2012
Numéro XVI

Hizkeren arteko aldakortasun sintaktikoa aztertzeke metodologiaren nondik norakoak : BASYQUE aplikazioa

Larraitz Uria et Ricardo Etxepare



Édition électronique

URL : <http://journals.openedition.org/lapurdum/2393>

DOI : 10.4000/lapurdum.2393

ISSN : 1965-0655

Éditeur

IKER

Édition imprimée

Date de publication : 1 octobre 2012

Pagination : 117-135

ISSN : 1273-3830

Référence électronique

Larraitz Uria eta Ricardo Etxepare, « Hizkeren arteko aldakortasun sintaktikoa aztertzeke metodologiaren nondik norakoak : BASYQUE aplikazioa », *Lapurdum* [Linean], 16 | 2012, Sarean emana-----an 03 décembre 2014, kontsultatu 21 avril 2019. URL : <http://journals.openedition.org/lapurdum/2393> ; DOI : 10.4000/lapurdum.2393

Hizkeren arteko aldakortasun sintaktikoa aztertzeke metodologiaren nondik norakoak: BASYQUE aplikazioa

Larraitz URIA & Ricardo ETXEPARE
IKER (UMR5478)^{1,2}, IKERBASQUE¹ & CNRS²

Abstracta: Artikulu honetan BASYQUE aplikazioa sortzeko jarraitu den metodologia aurkezten da, metodologia horren inguruko hainbat gogoetarekin batera. Aplikazio hau Iparraldeko hizkeren baitako bariazio sintaktikoa biltzera mugatzen den arren, baliatzen diren metodologia eta bitartekoak erabilgarriak dira beste hizkera batzuen inguruan ere antzeko aplikazioak sortu eta bariazioaren inguruko azterbide berriak irekitzeko.

1. Sarrera

Euskal hizkeren artean, fenomeno linguistiko jakin batzuetan oso nabarmena da bariazio sintaktikoa. Esaterako, Iparraldean, *lanari lotu zaio* esaten dute batzuek eta beste batzuek, aldiz, horrekin batera *lanari lotu da* (datibo-komunztadura egin gabe) esan dezakete; leku batzuetan *Xan artzaina zen* diote eta beste batzuetan *Xan artzain zen* (determinatzailea isilpean gordeaz); eremu dialektal batzuetan *Auzapezak sosa ebatsi omen du* esango dute, eta beste batzuetan berriz, horrekin batera, *Auzapezak sosa ebatsi du omen* esango dute (haustura intonazionalik gabe, *omen* partikularen posizioa aldatuz).

TSABL (*Towards a Syntactic Atlas of the Basque Language*)¹ proiektuaren xedeetako bat, hain zuzen ere, euskaraz hizkeraz hizkera dagoen aldakortasun sintaktikoa aztertzea da. Izan ere, ez da horrelako ikerketa asko burutu orain artean. “It is well known and often mentioned that the study of syntax has only played a very marginal role in traditional dialectology... In fact, in the atlas projects all over the world only a scarce part of the published dialect maps involve syntactic data (Carrilho, 2010; p.57)... the place for syntax was usually very marginal when compared to that of lexicon, phonology or even morphology” (Carrilho, 2010; p.60). Beraz, proiektu honen bidez aurrerapauso bat egin nahi da Iparraldeko hizkeren azterketan, batez ere sintaxiari dagokionean.

1.- <http://www.iker.cnrs.fr/tsabl-towards-a-syntactic-atlas-of-.html?lang=fr>

Horrelako ikerketak egitean, garrantzitsua da jarraituko den metodologiaren nondik norakoak ongi zehaztea (datu-bilketarako baliatuko diren informazio-iturriak, datu-emaielen ezaugarriak, jasoko den informazioa, informazio hori biltzeko eta kontsultatzeko erabiliko diren baliabideak, etab.). Proiektu honetan corpusak eta tresna informatikoak baliatzen dira, oinarriko bitartekoak/euskarriak direnak alor honetan ikertzeko. Corpusak informazio-iturri aberatsak dira eta tresna informatikoei informazioa biltzeko, kudeatzeko eta kontsultatzeko aukera anitz eskaintzen dute.

Hizkeren arteko aldakortasuna aztertzeko, beraz, datu-base sendo bat hornitzeari ekin zaio, hainbat informazio-iturritatik bildutako adibideekin. Datu-base horretan gordetako informazioa izango da ondoren egingo diren azterketen oinarri nagusia. Informazio hori aztertzeko, BASYQUE aplikazioa² garatu da Baionako IKER-UMR5478 zentroaren³ eta EHUko IXA taldearen⁴ arteko lankidetzarekin. BASYQUE aplikazioak euskal hizkeren arteko aldakortasun sintaktikoaren inguruko informazioa gordetzeko, kudeatzeko eta kontsultatzeko askotariko aukerak eskaintzen ditu, baita izaera dialektala duten corpusak (testu-bildumak) eskuratzeko eta aztertzeko ere, besteak beste.

Sarrera txiki honen ondoren, bigarren atalean Iparraldeko hizkeren artean dagoen aldakortasun sintaktikoa aztertzeko jarraitu den metodologia izango dugu aipagai. Jarraian, hirugarren atalean, ikerketa honek eragin dizkigun zenbait gogoeta agertuko ditugu. Bukatzeko, ondorio nagusiak laburbilduko dira laugarren atalean.

2. Metodologia

Sarreran aipatu bezala, hizkeren arteko aldakortasun sintaktikoa aztertzeko BASYQUE aplikazioa garatu da (Uria eta Etxepare, 2011). Bertan, hiru informazio-iturritatik bildutako adibideak gorde eta kontsulta daitezke: galdeketa zehatzen bidez eskuratutako adibideak, literatur corpusetan aurkitutakoak eta bestelako informazio-iturri batzuetatik eskuratutakoak.

Orain artean, galdeketei eman zaie batik bat lehentasuna, corpus zabaletan baino errazagoa baita guk prestatutako galdeketen bidez aztergai ditugun fenomeno linguistikoei dagozkien adibideak eskuratzea. Baina literatur corpusetan nahiz bestelako iturri batzuetan aurkitutako adibideak jasotzea ere interesgarria deritzogu. Izan ere, bariazio sintaktikoa erakusten duten adibide-kopuru handiagoo biltzeko aukera ematen du horrek, eta hiru informazio-iturrien ezaugarriak desberdinak izan arren, adibide guztiak izan daitezke adierazgarriak aldakortasun sintaktikoaren azterketarako. Hain zuzen ere, ongi hornitutako datu-base bat prestatzea komeni da ondorio linguistikoak atera ahal izateko eta bariazio

2. <http://ixa2.si.ehu.es/atlas2/index.php?lang=eu>

3. <http://www.iker.cnrs.fr/-centre-de-recherche-.html?lang=fr>

4. <http://ixa.si.ehu.es/IXA>

sintaktikoaren inguruko hipotesiak egiaztatu ahal izateko.

Datu gehien galdeketen atalean bildu direnez, atal honetan informazioa jasotzeko egin diren urratsak laburbilduko dira jarraian:

- **Fenomeno linguistikoa(k) hautatzea:** Garrantzitsua da aldakortasun sintaktikoa erakusten duten fenomenoak hautatzea, eta horretarako deskripzio maila egokian kokatzea. Cornips eta Polettok aipatzen duten moduan: "It is of major importance ... to avoid standard variants and/or intermediate variants as much as posible" (Cornips & Poletto 2005; p. 945). Iparraldeko hizkeren arteko aldakortasuna aztertzeo orduan, guk ondoko fenomeno linguistikoak hautatu ditugu, azterketa bibliografiko sakon batek erakutsi duelako fenomeno horietan aldakortasuna gertatzen dela:
 - o datibo-komunztadura
 - o postposizio-egiturak
 - o determinatzailearen erabilera
 - o omen partikularen posizioa
 - o ote partikularen posizioa
 - o zenbatzaile eta aditzaren arteko komunztadura.

- **Galdeketak prestatzea:** Ezinbestekoa da datu-emaleei egingo zaizkien galdeketak ongi diseinatzea emaitza fidagarriak eta baliagarriak eskuratzeko. Proiektu honetan, aldakortasun sintaktikoa erakusten duten fenomeno linguistikoei buruzko 6 galdeketa prestatu dira, eta galdera mota desberdinek osatzen dituzte galdeketa hauek. Zehazki, bost galdera mota baliatu dira:
 - o Itzulpenak, frantsesetik (kasu honetan) euskal hizkerara:
Il s'est appliqué au travail.
Pierre s'est appuyé contre le mur.
Pierre est berger.

 - o Hutsuneak bete:
Ikusi zuelarik ez zuen agurtu. Hori Piarresi ainitz gaitzitu _____.
Egunero Piarresi jarraitzen _____.
Piarresi emaztea eritu _____.

 - o Onarpen testa: posible ote da zure hizkeran? Xuxena da, bai ala ez?
 - a. Oihanari barna joan da.
 - b. Oihanean barna joan da.
 - c. Oihana barna joan da.
 - d. Oihan barna joan da.

 - o Hobespen testa: bi aukera hauetatik zein zaizu xuxenena, zein da hurbilenik zure hizkeratik?
Dena beretzat atxiki du.

- a. Ez dio nehoi deus igorri
- b. Ez du nehoi deus igorri.

Gaur haurrak ongi portatu dira.

- a. Ez dute elgarri deus aurtiki/bota
- b. Ez diote elgarri deus aurtiki/bota

- o Erabiltzen duzu hitz hau? Etsenplu bat eman
Oldartu hitza erabiltzen ote duzu? Etsenplu bat eman.
Gaitzitu hitza erabiltzen duzu? Hala bada, eman etsenplu bat.

Batik bat itzulpen-ariketak egitea eskatu zaie datu-emaileei (% 71,94), baina onarpen zein hobespen testak (% 16,18 eta % 8,63, hurrenez hurren) eta hutsuneak betetzeko ariketak (% 2,15) nahiz “*Erabiltzen duzu hitz hau?*” motakoak (% 1,07) ere baliatu dira (1. taula).

Galdeketa / Galdera mota	Itzulpenak	Hutsuneak bete	Onarpen testa	Hobespen testa	Erabiltzen ?
Datibo-komun.	129	6		7	3
Postposizioak	10		20		
Determinatzailea	53				
Omen	3		13		
Ote	4		12		
Zenbatzaileak	1			17	
	200 (% 71,94)	6 (% 2,15)	45 (%16,18)	24 (%8,63)	3 (%1,07)

1. taula: Galdeketa bakoitzean egindako galdera-mota desberdinen kopuruak.

Sei galdeketen bidez, 278 galdera egin zaizkio guztira datu-emaile bakoitzari.

- **Herriak aukeratzea:** Eskualde desberdinetako herriak hautatu dira Iparraldeko hizkeren ikuspegi zabal eta osatu bat jasotzeko helburuarekin. Guztira, 54 datu-emaile elkarrizketatu dira Iparraldeko 33 herritan:⁵ Lapurdiko 15 herritan⁶, 27 datu-emaile;

5.- Artikulu hau publikatzen den egunean.

6.- Hendaia, Azkaine, Donibane-Lohitzune, Bidarte, Urruña, Ziburu, Sara, Itsasu, Beskoitze, Ahetze, Arbona, Senpere, Jatsu, Mugerre, Luhuso.

Beha Nafarroako 13 herritan⁷, 21 datu-emaile; eta Zuberoako 5 herritan 6 datu-emaile⁸. Aurreikusita dago epe laburrean zuberotar gehiago elkarrizketatzea.

Galdeketak hainbat herritan egin diren arren, herri bereko hainbat datu-emaile ere elkarrizketatu dira, herrian bertan aldakortasunik gertatzen ote den azertu ahal izateko (adibidez, Senperen eta Baigorri 4 biztanleren erantzunak bildu dira; Saran eta Urepelen 3 biztanlerenak; eta beste hainbat herritan 2 biztanlerenak). Hautatutako herriak, Koldo Zuazok (Zuazo, 2008) Iparralderako definitutako sei eremu dialektaletan kokatzen dira (Sartaldeko azpieuskalkia, Sortaldeko azpieuskalkia, Sartalde eta sortaldeko azpieuskalkien tarteko hizkerak, Kostatarra (erdialdekoaren eta nafarlapurtarraren tarteko hizkera), Amikuze (nafar-lapurtarraren eta zuberotarraren tarteko hizkera) eta Zuberotarra).

- **Datu-emaileak bilatzea:** Lau adin-tarteko datu-emaileak elkarrizketatu dira: 35 urtez beherakoak, 36-50 urte bitartekoak, 50-70 urte bitartekoak eta 70 urtez gorakoak. Datu-emaileen adin-tartea zehaztuz, hizkeren bilakaera ere aztertzeo aukera izango da. Gazteen artean, euskaraz ikasi duten eta ez duten hiztunak bilatzen dira, eta gazteei egindako elkarrizketek aukera ematen dute ikusteko i) tokian tokiko nahiz familia giroko hizkerek zenbateraino dirauten bizirik gazteen artean, ii) eskolatzeak hizkeren erabileran edo erabileran ezean izan dezakeen eragina eta iii) frantsesa bezalako hizkuntza menderatzaile batekin harreman zuzenean bizitzeak zein eragin izan dezakeen gazteen hizkeran. Galdekaturiko datu emaile gehienak 50 urtetik gorakoak dira. Nolanahi ere, egungo egunean, galdekaturiko datu emaile gehienak 50 urtetik gorakoak dira. Kasu guztietan, datu-emaileen anonimotasuna gordetzen da.
- **Galdeketak egin eta elkarrizketak grabatzea:** Iparraldeko hizkerak ezagutzen dituen bertako pertsona bat ibili da inkesta lana egiten eta elkarrizketak grabatzen. Garrantzitsua da elkarrizketagileak euskalkien ezagutza izatea, baita galdeketak egiten esperientzia izatea ere, momentu jakin batzuetan datu-emaileei behar duten laguntza eskaini ahal izateko.
- **Erantzunak transkribatzea:** Grabatutako elkarrizketak transkribatu eta datu-base batean gorde dira BASYQUE aplikazioaren bidez. Erantzun bakoitzak hiru transkripzio ditu: erantzuna bere horretan, erantzunari dagokion idazkera formala eta idazkera formalari dagokion testu normalizatua:

Transkripzio fonologikoa: Deneri harriak botatu diozkate.

Idazkera formala: Deneri harriak botatu diozkate.

Testu normalizatua Denei harriak botatu dizkiete.

7.- Urepele, Baigorri, Bidarra, Uharte-Garazi, Behorlegi, Suhukune, Gabadi, Garruze, Amenduze-Unaso, Jutsi, Larzabale, Iholdi, Heleta.

8.- Mitikile, Maule-Lextarre, Ezpeize-Ündüreine, Barkoxe eta Pagola.

Horrek aukera ematen du, besteak beste, corpus paraleloak sortzeko, ezinbestekoak direnak landutako hizkeren tratamendu automatikoa bideratu ahal izateko⁹.

Erantzunak etiketatzea: Transkribatutako erantzunak etiketatu ere egiten dira, hau da, bildutako adibide bakoitzari propietate linguistikoki jakin batzuk esleitzen zaizkio. Propietate linguistikoki horiek informazio sintaktikoari buruzkoak dira, eta horiei esker jakin daiteke adibide bakoitza zein motako bariazio sintaktikori dagokion. Esaterako, datiboaren eta aditz-laguntzailearen arteko komunztadura aztertzeko, ondoko propietate hauek zehazten dira: datibo-argumentua badagoen edo ez, argumentu mota zein den (izen berezia, izen ordaina, ezezko polaritatedun elementua, datibo bihurkaria, datibo elkarkaria...) eta datibo-argumentuaren eta aditzaren artean komunztadurarik badagoen ala ez:

Lanari lotu **zaio**.

Propietate linguistikokiak: Dative BAI

Dative argument: DP definite

Agreement BAI

Arropa **horri** botoi bat josten ahal **zenuke**?

Propietate linguistikokiak: Dative BAI

Dative argument: demonstrative

Agreement EZ

Postposizioen adibideei dagokionean, berriz, esleitzen diren propietateek postposizioa simplea edo konplexua den adierazten dute. Egitura konplexuen kasuan, postposizioaren lehen osagaiaren deklinabide atzizkia eta postposizio elementua zehazten dira; egitura simpletan, berriz, deklinabide atzizkia:

Eni **buruz** itzuli da.

Propietate linguistikokiak: POSTP_complex

POSTPdative

postp-BURUZ

Oihanean ibili dira.

Propietate linguistikokiak: POSTP_simple

Inessive

9.- Proiektu honen beste helburuetako bat da, hain zuzen ere, biltzen den informazioa oinarri hartuta, Iparraldeko hizkeren azterketa eta prozesamendua bideratzea. Lan honi 2013. urtetik aurrera ekingo zaio, Larraitz Uriaren doktorego ondoko ikerketa egitasmoaren barruan.

Determinatzailearen galdeketaren bidez, datu-emaileek erantzunetan determinatzaileerik erabili duten ala ez aztertu nahi da, eta erabiltzen duten kasuetan, determinatzaile zehaztua, zehaztugabea edota partitiboa erabili duten:

Pierre **artzain** zen.

Propietate linguistikoa: DET_in_PRED EZ (artikulerik ez predikatuan)

Pierre **artzaina** zen.

Propietate linguistikoa: DETsing_in_PRED BAI (artikulu zehaztu singularra predikatuan)

Pierre **artzain bat** zen.

Propietate linguistikoa: DETindef-sing_in_PRED BAI (artikulu zehaztugabe singularra predikatuan)

Ebidentzialitateari dagokionean, *omen* eta *ote* partikulak esaldi barruan zein posiziotan agertzen diren zehazten da propietate linguistikoen bidez:

Bere osaba preso **omen** da.

Propietate linguistikoa: PRT-omen AUX

Omen, egun horretan etorri ziren.

Propietate linguistikoa: Parenthetical omen

Gaitzitu **ote** dugu?

Propietate linguistikoak: V PRT-ote AUX

Zer **ote** zaio gertatu?

Propietate linguistikoak: PRT-ote AUX V

Azkenik, zenbatzaileei dagokien galdeketaren bidez bildutako erantzunetan, propietate linguistiko hauek esleitzen dira: zenbatzaile mota (*anitz* ala *guti*), zenbatzailea zein argumentutan agertzen den (subjektuan, objektuan ala zehar-objektuan) eta argumentu horren eta aditzaren arteko komuntadura singularra ala plurala den:

Ikasle **anitz** ikusi **ditut** aurten eskolan.

Propietate linguistikoak: QUANTmany
QUANTmany_in_ARGobj
Agreement-PLUR

Rugbyko jokalaria **guti da** euskalduna.

Propietate linguistikoak: QUANTfew
QUANTfew_in_ARGsubj
Agreement-SING

Beraz, propietate linguistikoak oso baliagarriak dira bariazio sintaktikoaren inguruko bilaketak edota azterketak egiteko garaian. Izan ere, propietate linguistikoen bidez, modu errazean eskura daitezke ikerlari bakoitzari interesatzen zaizkion datuak: datibo-argumentuarekin komunztadurarik ez duten adibideak, ezezko polaritadedun elementua duten eta aditzarekin komunztatzen ez diren datibo-argumentuak, *guti* zenbatzailaren eta aditzaren artean komunztadura plurala erakusten duten adibideak, eta abar luze bat.

Erantzunak glosatzea: Erantzunak etiketatzeaz gain, glosatu ere egiten dira, hau da, esaldiko elementu bakoitzari dagokion informazio linguistikoa zehazten da:

[Lanari]-N-DET-sing-DAT [lotu]-V [da]-AUX3S(abs)
[Atean]-N-DET-sing-INE [jo]-V [du]-AUX3S(abs)3S(erg)

Glosakautomatikokiesleitzen dira EHUko IXA taldeangaratutako baliabide-informatikoak baliatuz. Horrelako atlas linguistikoetan HAPen alorreko baliabideak integratzea berrikuntza ez ezik, aurrerapauso garrantzitsua ere bada, horrek erraztu egiten baitu batetik, aldakortasun sintaktikoaren inguruko datuak eskuratzeko eta aztertzeo aukera, eta bestetik, bildutako adibideen tratamendu automatikoa bideratzeko lana, horixe delarik, hain zuzen ere, proiektu honen beste helburuetako bat.

Erantzunak mapa gainean ikustea: BASYQUE aplikazioaren bidez egiten diren bilaketen erantzunak mapa gainean islatzen dira. Eta askotan atlas sintaktikoetan gertatzen ez den moduan, BASYQUEk eskaintzen dituen mapak dinamikoak dira (estatikoak izan beharrean).

Horrela, bada, BASYQUE aplikazioaren bidez, *Galdeketak* atalean Iparraldeko 33 herritan 54 datu-maileri egindako 6 galdeketen bidez 13.270 adibide jaso eta landu dira orain artean.

Literatur corpusen atalari dagokionez, bost idazleren lanetan (9 lan, guztira) aztergai ditugun fenomeno linguistikoei dagozkien adibideak aurkitu eta bildu dira. Denetara, 2.932 adibide. Adibide hauek ere galdeketen atalean egin den modu berean gorde eta etiketatu dira.

Fenomeno linguistikoa	Adibide-kop.	%
Datibo-komunztadura	2.431	75,7 %
Postposizioak	169	5,26 %
Determinatzailea	316	9,84 %
Omen	139	4,32 %
Ote	148	4,60 %
Zenbatzaileak	8	0,24 %
	3.211	

2. taula: Literatur corpusetan fenomeno linguistiko bakoitzeko bildu den adibide-kopurua.

Azkenik, bestelako iturrietako datuak eskuratzeko atalak ere sortu ditugu. Bi eratako datu iturburuak ditugu gogoan, *www.ahotsak.com* eta *www.norantz.org* datu baseetan erakusgai direnak. Bi proiektu hauek helburu desberdinak dituzte: *ahotsak.com*-en xedea da herriko ahozko ondarea eta herri hizkerak bildu eta hedatzea. Horretarako, gai konkretu baten inguruko testigantzak grabatzen dituzte; jardun libreak dira. *NORANTZ* proiektuan, aldiz, galdeketa egiten dira, gaur egungo Iparraldeko euskaldunen gramatika oinarriak, joera handienetan, zer bidetarik doazen ikusteko helburuarekin. Datuak biltzeko eta lantzeko jarraitu duten metodologia, beraz, ezberdina da kasu bietan: batean gai baten inguruko jardun libreak biltzen dira eta bestean, galdera zehatzei dagozkien erantzunak. Ondorioz, BASYQUE aplikazioan ere adibide horiek ez lirerateke modu berean jasoko.

Ahotsak.com proiektuko adibiderik ez dago oraindik BASYQUE aplikazioan kontsultagarri, Iparraldeko hiztunen testigantza ugari bildu dituzten arren, elkarrizketa horiek ez daudelako oraindik transkribatuta edo landuta. Baina etorkizun batean iturri honetako adibideak jasotzeko atala prest dago: *Ahotsak.com* proiektuan grabatutako bideo-fitxategiak BASYQUE aplikazioan kargatuko dira; testingantza horietan aldakortasun sintaktikoari lotutako adibideak badaude, esaldi horiek transkribatu egingo dira eta bakoitzari dagozkion propietate linguistikoeekin eta glosekin batera gordeko dira. Ez da ahaztu behar, baina, hauek ez direla galdeketa zehatzei lotutako adibideak izango, gai jakin baten inguruko jardun libreak baizik.

NORANTZ proiektuan adibideak galdeketa bidez eskuratzen diren arren, BASYQUE aplikazioan adibide horiek ez dira galdeketen atalean gordetzen, bestelako iturrien atalean baizik, gai bati lotutako adibide solteak bailiran (datibo-komunztadurari dagozkion adibideak, determinatzailearen ingurukoak, etab.). *NORANTZ* proiektuan guztira bildutako 46.138 erantzunetatik, 1.581 adibide jaso dira orain artean BASYQUE aplikazioan (3. taula).

Fenomeno linguistikoa	Lapurdi	Behe-Nafarroa	Guztira
Datibo-komunztadura	304	317	621
Determinatzailea	379	313	692
Omen	3	4	7
Ote	20	25	45
Zenbatzaileak	121	95	216
	827	754	1.581

3. taula: *NORANTZ* proiektuan bildu eta geureganatu ditugun adibideen kopurua.

Beraz, hiru informazio-iturri baliatuz, aldakortasun sintaktikoa erakusten duten hainbat adibide jaso dira, bakoitzari dagokion informazio osagarriarekin batera.

3. Gogoeta, jarraitutako metodologiaren inguruan

Hizkeren arteko aldakortasun sintaktikoa aztertzea hasiera batean dirudiena baino ataza konplexuagoa da: “Although we usually conceive language as a single object of scientific study, microvariation clearly shows that each speaker has his/her own I-language, which

minimally differs from the I-language of everybody else, including people educated in the same environment at the same time. We could conceive the study of language therefore as similar to the study of families of bacteria in biology: the object of our investigation are families of very closely related languages, which, at first sight, appear to be identical, though they are not when subtler tools of inquiry are used” (Cornips & Poletto 2005; p. 940). Horregatik, aldakortasun sintaktikoaren alorra ikertzerakoan, garrantzitsua da ikerketaren helburuak argi izatea eta helburu horiek lortzeko jarraituko den metodologia ahal den neurrian finkatzea.

Egingo den azterketa xehetasunez hasieratik planifikatzea, ordea, zaila izan ohi da, aurreikusi ezin diren hainbat fenomeno agertu ohi baitira ikerketan aurrera egin ahala (elkarrizketak egiten direnean, datuak transkribatzen direnean, edo erantzunak lantzeko garaian, esate baterako). Garrantzitsua da, beraz, ezusteko horiei erantzuteko baliabideak eskura izatea, bai sor daitezkeen zalantzen aurrean erabakiak hartzeko edota irizpideak finkatzeko lan-taldea eta bai erabiltzen diren baliabide informatikoak kudea ditzakeen informatikaria.

Proiektu honetan, esan bezala, hiru informazio-iturri edo atal bereizten dira: galdeketak, literatur corpusak eta bestelako iturriak. Informazio-iturri bakoitzak ditu bere alde onak eta txarrak, abantailak eta ahuleziak.

Galdeketei dagokienean, hauen abantailak bat da ikerlari bakoitzak prestatzen duela bere helburuak lortzeko komeni zaion galdeketa mota, finkatzen dituela elkarrizketatuko dituen datu-emaeleen ezaugarriak, hautatzen dituela zein herritan egingo diren galdeketak, etab. Alegia, ikerlari bakoitzak diseinatzen duela komeni zaion metodologia. Hori ez da posible, noski, literatur corpusetatik edota bestelako iturrietatik datuak biltzeko garaian.

Galdeketek baliagarriak eta fidagarriak diren adibideak erraz eta azkar eskuratzeko aukera ematen dute, aztergai diren fenomeno linguistikoei dagozkien adibideak eskuratzeko prestatuak daudelako, hain zuzen. Kontuan izan behar da, ordea, galdeketak erantzuten dituenean, datu-emaele egoera artifizial batean hitz egiten ari dela, edo ari daitekeela, egunerokotasunean aurkitzen ez duen egoera batean. Hau da, datu-emaele nolabait behartu egiten da egitura konkretu bat erabiltzera edo hautu bat egitera, ez da libreki jarduten. Ondorioz, hizkeraren naturaltasuna apur bat galtzeko arriskua egon daiteke (Cornips & Poletto 2005; p. 947).

Galdeketak galdera mota desberdinez osatu ahal dira. Proiektu honetan, esaterako, itzulpenak baliatu dira gehienbat, baina orotara 5 galdera-mota egin zaizkie datu-emaeleei. Galdera mota batzuek eta besteek, guztiak dituzte alde onak eta txarrak, ez baitago egokiena den galdera mota bakar bat. Galdera motaren egokitasuna landu nahi den fenomeno linguistikoaren arabera ere bada, hau da, fenomeno batzuk lantzeko egokiagoak dira galdera mota batzuk, baina beste hainbat fenomeno lantzeko, aldiz, bestelako galderak baliatzea komeni izaten da.

Itzulpen-ariketek bakoitzaren hizkera modu nahiko naturalean erabiltzeko aukera eskaintzen dute. Horrez gain, corpus homogeen bat eskuratzeko aukera eskaintzen dute, hau da, itzulpen-ariketen bidez segurtatzen da datu-emaele guztiak jatorrizko egitura jakin bati dagokion adibidea emango dutela, galdera berdinak egiten baizaizkie denei. Baina bi sistemen (jatorrizko eta xede hizkuntzen) egiturak edota ezaugarriak desberdinak izan daitezke eta

horrek kasu batzuetan ‘arazoak’ eragin ditzake. Azterketa honetan, oro har, galderak modu naturalean eta zalantza handirik gabe itzuli dituzte datu-emaileek. Baina hainbat kasutan, hiztun batzuek adierazi izan dute itzultzeko eskatu zaion egitura ez dela beraien hizkeran erabiltzen: *Ez da errexa bilatzea hitza parez parekoa* (1944an jaio eta Mugerrekoa den datu-emailea); “konparatu” *da hitz bat ez dakiguna edo ez duguna erabiltzen* (1944, Mugerre); *Iruditzen zaut arrunt frantses egitura kopiatzen dela, ez dut sentitzen euskarazko egitura bat bezela* (1961, Behorlegi); “allons” euskaraz ez diagu erraiten hemen (1943, Baigorri).

Itzulpenak egiteko garaian, zalantzaren bat edo beste ere adierazi izan dute datu-emaileek: *Beti preseski badut arazo bat forma horrekin, beti ez dakit “amari hurbildu natzai” edo “amarenganat hurbildu niz” edo beste bat, “amaz hurbildu niz”, ez dakit sekulan itzulpena nola egin* (1955, Bidarra).

Beste kasuren batean, egin zaion galdera itzuli arren, berak ez lukeela horrela esango adierazi izan du datu-emaileak: *posible litzateke “joan” aditzarekin (“sosa joanen da gure alabari”) baina nik erran beharko banu ez nuke hori enplegatuko* (1964, Itsasu).

Halaber, bada galdera bat non datu-emaileari galdetzen zaion ea beste moduren batean esango lukeen erantzun berdina. Zehazki, *La maison est orientée sud eta La fenêtre donne sur la mer* esaldiak itzultzeko eskatzen zaio. Hitztunak “etxea *hegoari buruz* itzulia da” edo “leihoak *itsasoari buruz* ematen du” erantzuten badu, galde egin behar zaio ea perpauz bera datibo hutsez ere eman dezakeen, hau da, “hegoari” edo “itsasoari” bakarrik erabiliz. Izan ere, corpusetan horrelako adibideak aurkitzen baitira. Galdera horien erantzunak, baina, iruzkinen atalean gordetzen dira: “*hegoari itzulia da etxea*” *ere xuxena dela, eta hegoari vis-a-vis ere* (1930, Luhuso); “*buruz*” *kentzen baginu (“itsasora ematen du leihoak”) erraten ahal da ere, ulertzen da* (1961, Heleta); *hobeki da “buruz”ekin, “itsasoari emaitte du” baino* (1922, Baigorri); *Berdin erran daiteke “buruz” kenduta, “Leihoak emaitte du itsasoari”* (1930, Uharte Garazi); *Nik ez nezake erran “buruz” gabe, *etxea hegoari da* (1944, Arbona). Baina iruzkin gisa gordetzen diren erantzun hauek ez dira mapan islatzen, azterketarako adierazgarriak izan arren. Beraz, konturatu gara galderak horrela egitea ez dela komeni, egokiagoa dela onarpen nahiz hobespen test moduan egitea.

Hutsuneak bete motako galderek ere hiztunei nahi duten erantzuna emateko aukera eskaintzen diete, hau da, hitzunik bere hizkeran baliatzen duen forma edo egitura emango du hutsunean falta dena betetzeko garaian. Horrelako galderetan gerta liteke hitzunik erantzun posible bat baino gehiago izatea (kasu gutxi batzuetan gertatu da hori). Aukera dago iruzkinen atalean oharrak egiteko (“*zako*” *eta “zitzai” guretzako gauza bera da, eta orainean “zitzako*” (1935, Jutsi); Datu-emaileak dio ‘diozki’ eta ‘diozka’ berdin, biak erabiltzen direla (1930, Luhuso)), baina erantzun bakar bat hartzen da ontzat. Komeni da horrelako galderak prestatzean erantzun bat eskatzen duten galderak asmatzea, iruzkinetako oharretan jasotzen diren aukera posible horiek ez direlako ikusten erantzun gisa bilaketa bat egiten denean (nahiz eta oharretan aipatzen diren aukera posible horiek guztiak berdin baliagarriak diren hiztunaren hizkeran).

Test motako ariketak itxiagoak izan ohi dira, datu-emaileari jarritako hainbat egituretatik

zuzenena zaiona aukeratzea eskatzen baitzaio. Hala ere, emandako aukeren artean datu-emaileak erantzun egokirik aurkitzen ez badu, bere hizkeran erabiliko lukeen forma emateko eskatzen zaio. Zenbatzaileen galdetegian, adibidez, badago hobespen test bat, zeinaren bidez datu-emaileari bi adibidetatik egokiena zaiona aukeratzeko eskatzen zaion: “*Bi bertsolariri aditu diot*” edo “*Bi bertsolariri aditu diet*”. Baina hainbat datu-emailek beste forma batzuk hobetsi dituzte: “*Bi bertsolariri aditu dut*” (4 datu-emailek) eta “*Bi bertsolariri aditu ditut*” (2 datu-emailek). Erantzun horiek test-galdeketa txertatu dira aukera gisa, beste datu-emaile batzuentzat ere baliagarriak izan daitezkeelakoan.

Test motako ariketek duten desabantaila bat da batzuetan datu-emaileek erantzun bat hobetsi arren, beste aukeraren bat ere posible dela adierazten dutela. Eta horrek nahasmena sor dezake, ez baita islatzen zenbateraino hobesten duen hiztunak erantzun bat bestearen aldean. Horrelakoetan aukera bakar bat baino ez litzateke ontzat hartu beharko, baina hori ere ez da justua hiztun batek forma bat baino gehiago onartzen baditu. Onarpen-testak, hala ere, ebaluzio-testak (“baloratu erantzun egokiena Otik 10era” motakoak) baino egokiagoak dira. Izan ere, puntuazioa oso subjektiboa izan liteke eta ondorioz, horrelako galderekin lortzen diren erantzunak ez dira hain fidagarriak izaten. Proiektu honetan ez da ebaluzio-testik baliatu horregatik.

“Erabiltzen al duzu ... hitza?” motako galderetan, ziurrak diren erantzunetan ez dago arazorik. Baina hiztun batzuek esan izan dute galdetutako hitza erabili ez arren ezagutzen dutela edota galdetutako hitzaren ordezkari beste bat erabiltzen dutela. *Nik ez dut erabiltzen baina entzuten da* (1959ean jaio eta Ahetzekoa den datu-emailea); *egia erran mintzatzean gehiago erraten dugu “kexatu” edo “samurtu”, “oldartu” gero ikasi nian literatura pixkat irakurtzean-eta* (1949, Amendüze-Unaso); “*jarraiki” beti problemak pausatzen ditu aditz horrek, zeren eta “jarraiki zaizkio” erabili behar da eta konplikatu da, eta beraz “segitu” erabiltzen dugu* (1955, Bidarra); “*Gaitzitu*”? “*Gaitzesi*”? *Guk erraten dugu “txokatu”* (1949, Urruña).

Beraz, zalantzen aurrean datu-emaileek egin izan dituzten oharrak gorde egiten dira iruzkinak jasotzeko prestatu den eremuan. Horrela, erantzunak aztertzean ikerlariak datu-emaileak izan duen zalantzaren azalpena ikusi eta kontuan har dezake. Datu-emaileen zalantza edota ohar horiek, ordea, ez dira mapa gainean islatzen, adierazgarriak izan daitezkeen arren. Bestalde, datu-biltzailearen lana da zalantzen aurrean hiztunari bidea erraztea, laguntza ematea. Baina horrek ere izan dezake eragozpen bat. Izan ere, datu-emailea erantzuna asko pentsatzera eramane dezake eta, ondorioz, gerta liteke ongi pentsatutako erantzuna ez bat etortzea naturalki erabiliko lukeen egiturarekin.

Galdeketekin gerta daitezkeen beste arazo bat da datu-emaile batzuen erantzunetan inkoherentziak aurkitzea, hau da, galdera batzuetan txertatutako emandako egituraren bat erabili edo ontzat eman ahal izan dutela beste momenturen batean. Egitura desberdinak behin edo bitan bakarrik erabili baditu, nahastu egin dela pentsa daiteke; galdeketa guztian askotariko egiturak erabili baditu, modu jakin bat nabarmendu gabe, erabili dituen forma guztiak bereak dituela pentsa daiteke. Eta kasu horretan, arrazoia aztertu beharko litzateke: datu-emailea Iparraldean barrena asko ibili izan dela edo ibiltzen dela, hizkera desberdinak hitz egiten dituzten lagunartea duela, etxeko eta laneko hizkerak/erregistroa nahasten dituela edota

besterik gabe, leku horretan forma bat baino gehiago elkarrekin bizi direla. Izan ere, hainbat faktorek izan dezakete eragina datu-emaileen erantzunetan (Cornips & Poletto 2005; p. 942)

Literatur corpusen kasuan, ez dago audiorik, testu idatziak dira. Corpus idatzietan maiz gertatzen diren fenomeno linguistikoak edo egiturak erraz bildu ohi dira, eta horrelako kasuetarako informazio-iturri aberatsa da corpusa. Hain ohikoak ez diren fenomenoak aurkitzeko, aldiz, testu masa handia aztertu beharra dago eta adibide gutxi batzuk eskuratzeko denbora asko behar izaten da. Proiektu honetan, adibidez, datibo-komunztadurari dagozkion adibide ugari bildu dira testu idatzietan (2.431, guztira), fenomeno oso ohikoa delako, baina zenbatzaileen eta aditzaren komunztadurari dagozkion 8 adibide baino ez dira aurkitu orain artean (*anitz* zenbatzailearen eta aditz-laguntzailearen arteko komunztadura singularra duten 5 adibide eta komunztadura plurala duten 3 adibide, hain zuzen).

Bestelako iturriak atalean bildutako datuei dagokienean, beste proiektu batzuetan egin den lana baliatzeak duen abantaila nagusia da grabaketak eginda daudela eta, ondorioz, ez dagoela datu-emaileak bilatu eta elkarriketatzerara atera beharrik. Denbora asko eta lan handia aurrezten da horrela. Desabantaila da, ordea, proiektu bakoitzak bere helburuak dituen, ezin direla datu horiek guztiak maila, atal edota modu berean gorde. Beste proiektuetan bildutako adibideek eta adibide horien inguruko informazioak ez dute zertan bat etorri gure proiektuan biltzen den informazioaren ezaugarriekin (datu-emaileen perfilekin, datu-bilketarako aukeratzen den gune geografikoarekin, adibideak transkribatzeko edo lantzeko finkatzen diren irizpideekin, etab.). Horregatik, hain zuzen ere, bereizi dira hiru atal BASYQUE aplikazioan.

Ahotsak.com proiektuaren kasuan, esaterako, grabatu dituzten testigantzek duten alde on bat zera da: datu-emaileek edo hiztunek baliatzen duten hizkuntza galdeketa erantzuten dituztenean erabil dezaketena baino naturalagoa eta espontaneoagoa dela, hiztunak pentsatu gabe mintzaren baitira. Ondorioz, zalantza gutxi izan ohi dituzte eta hizkeraren adierazgarri ezinhobeak dira. Aldiz, kasu honetan dugun eragozpen bat da, Iparraldean grabatu dituzten testigantza gehienak transkribatu gabe daudela oraindik, eta transkripzio horiek asko erraztuko lukete lana.

NORANTZ proiektuaren kasuan, berriz, galdeketa zehatzak egiten zaizkie datu-emaileei. Alde honetatik, hurbilago daude bi proiektuen helburuak. Baina horrek ez du esan nahi *NORANTZ* proiektuko galdera-erantzun guztiak baliagarriak direnik BASYQUE aplikaziora ekartzeko. Ondorioz, proiektu honetarako interesgarriak diren datuak besteengandik bereizteko eta eskuratzeko lana egin behar da.

Galdeketen bidez bildutako adibideez gain, literatur corpusetan eta bestelako iturri batzuetan aurkitzen diren adibideak geureganatzea interesgarria eta aberasgarria zela pentsatu genuen, horrek aldakortasun sintaktikoaren inguruko adibideen kopuru handiagoa biltzeko aukera ematen duelako. Adibide horiek, ordea, hiru ataletan bereizi dira. Horren desabantaila bat da ezin dela propietate linguistiko berdinak dituzten adibide guztien bilaketa bakar bat egin, hau da, ezin direla hiru ataletan dauden datibo-komunztadurarik gabeko adibide guztiak, esaterako, bilaketa bakar baten bidez eskuratu. Bilaketak atalka egin behar dira. Baina horrela egitea egokiagoa zela aurreikusi zen, iturri desberdinetan sortu diren eta ezaugarri desberdinak

dituzten datu-emaileen adibideak direlako. Litekeena da, gainera, galdeketen bidez jasotzen diren erantzunen eta beste informazio-iturri batzuetan bildu direnen artean desberdintasunak egotea “the results obtained through elicitation data may differ from those appearing in spontaneous speech ... dialect corpora became also important heuristic tools for the study of non-standard syntax” (Carrilho, 2010; p.59).

Bestalde, ikerlari bakoitzak prestatuko galdeketa zehatzen bidez proiektu honetan bilduko diren erantzunen kopurua aurreikus daitekeen arren, ez da gauza bera gertatzen literatur corpusetan edota bestelako iturrietan dauden adibideekin, hau da, ez da ziurra beste iturri horietan aztergai ditugun fenomenoek dagokien adibiderik aurkituko denik. Galdeketa bidez egiten diren galdera-erantzunen datuak kontrolatzea, beraz, erraza izan ohi da, literatur corpusetan nahiz bestelako iturrietan zer aurkituko den aurreikustea ezinezkoa den bitartean.

Bukatzeko, BASYQUE aplikazioak hizkeren arteko aldakortasun sintaktikoaren inguruan askotariko bilaketak egiteko aukera eskaintzen digu: informazio-iturri desberdinen, galdeketen, galdeketetako atalen, datu-emaileen, datu-emaileen jaiotze-urtearen edota adin-tartean, herrien, eremu dialektalen edota propietate linguistikoen arabera, besteak beste. Egindako bilaketen erantzunak mapa gainean islatzen dira. Hain zuzen ere, atlas edo mapa dialektalei egotzi izan zaie informazio sintaktiko eskasa eskaintzen dutela: “in traditional dialect maps globally, no more than 5% of those that are published involve syntactic data and this is purely chance. Moreover, the data are not organised in a way that permits immediate comparison and analysis of the syntactic information contained in the maps. Similarly, traditional dialect monographs usually do not provide information about the syntax, which can only be indirectly recovered in some cases through the description of morphological patterns” (Cornips & Poletto 2005; p. 940). Baina aipaturiko azterketekiko gure atlasak duen berrikuntza bat da informazio sintaktikoa esplizituki kontsultagarri dagoela, adibide bakoitzari aldakortasun sintaktikoarekin lotura duten propietate linguistikoak esleitzen zaizkiolako. Eta atlas tradizionaletan gertatu ohi den ez bezala, mapak guztiz dinamikoak dira, bilaketa bakoitzeko mapa bat eskuratzen delarik. Adibidez, postposizioen galdeketa dagokion mapa, edo postposizio egitura jakin batzuei dagokiena, edo postposizio egitura jakin batzuen eta jaiotze-urte jakin batzuetako datu-emailei dagokien mapa, beste hainbat aukerarekin batera.

4. Ondorioak

Artikulu honetan, Iparraldeko hizkeren artean dagoen aldakortasun sintaktikoa aztertzeko jarraitu den metodologia aurkeztu da, Basyque aplikazioa eratzeko orduan aurkitu diren hainbat arazo eta galdera agertuz.

Argi dago horrelako azterketetan datu fidagarriak eta baliagarriak eskuratu ahal izateko, garrantzitsua dela baliatuko diren metodoak ongi aukeratzea. “Syntactic dialect atlas projects, in particular, have to deal with valid methods for obtaining reliable and sufficient syntactic data from dialects across a large geographical area” (Cornips & Poletto 2005; p. 940). Hala ere, jakina da ez dagoela metodo egoki edo baliagarri bakar bat, hau da, guztiek dituztela abantailak, baina baita hutsuneak ere. “The advantages and disadvantages of both naturalistic methods behind corpus-data and elicitation techniques have been recurrently discussed.

It may easily be acknowledged that none of them is exempt of problems” (Carrilho, 2010; p.58). Hau horrela izanik, gure ondorio nagusietako bat da aberasgarria dela datuak edo adibideak hainbat informazio-iturritatik biltzea, eta datu-bilketarako teknika edo metodo bat baino gehiago erabiltzea ere egokia dela, bakoitzak dituen alde onak eta txarrak kontuan hartuz. Baina informazioa jasotzeko iturria nahiz metodoa edozein izanik ere, garrantzitsua da datuak irizpide berdinak jarraituz tratatzea, hau da, koherentzia bat mantentzea adibideak transkribatzeko garaian, etiketatzerakoan, glosatzerakoan, ... Irizpideak finkatzea ezinbestekoa da, horrela baliaagarritasun, osotasun edota fidagarritasun handiagoa erakutsiko baitu bildutako informazioak.

Proiektu honetan, bariazio sintaktikoa aztertzeke BASYQUE aplikazioa garatu da, aldakortasunaren inguruko adibideak eta adibideei dagokien informazioa biltzeko, kudeatzeko eta kontsultatzeko aukera anitz eskaintzen dituena. BASYQUE aplikazioaren bidez eta hiru informazio-iturri baliatuz jaso diren adibideek nahiz adibide horiei erantsitako datuek euskarri sendo bat osatzen dute Iparraldeko hizkeren arteko bariazio sintaktikoaren azterketarako. Hain zuzen ere, bildutako datuetan oinarrituta, haien azterketa xehea egiteari ekingo zaio epe laburrean, aldakortasunaren inguruko hipotesiak betetzen diren egiaztatzeke edota ondorio linguistiko eta dialektologikoak ateratzeko asmoz.

Halaber, proiektu honetan bildutako datuek ezinbesteko oinarria osatzen dute corpus paraleloak sortu eta Hizkuntzaren Azterketa eta Prozesamenduaren (HAP) alorrean euskal hizkeren tratamendu automatikoa bideratu ahal izateke. Horrela, corpus paraleloak oinarri hartuta, hainbat tresna edo aplikazio garatu ahalko dira etorkizunean, hala nola, estandarra-euskalki hiztegiak, testuak euskara estandarretik euskalkira (eta alderantziz) itzultzeke erramintak edota euskalkiak irakasteko aplikazioak, esaterako.

Ikerketa hau Iparraldeko hizkeretara mugatzen den arren, baliatzen diren metodologia eta bitartekoak erabilgarriak dira beste hizkera batzuen ikerketarako ere.

Informazio hau guztia kontsultagarri jartzea ekarpen interesgarria dela uste dugu bai dialektologiaren eta bai sintaxi teoriaren alorrean aurreratzen jarraitzeke.

Erreferentziak

- Cornips, L. & C. Poletto 2005. “*On standardising syntactic elicitation techniques. PART I*”. *Lingua* 115 (7), 939-957.
- Carrilho, Ernestina. 2010. “*Tools for dialect syntax: the case of CORDIAL-SIN (an annotated corpus of Portuguese dialects)*”. In Gotzon Aurrekoetxea and Jose Luis Ormaetxea (eds.) *Tools for Linguistic Variation*. Bilbao: Universidad del Pais Vasco. 57-70.
- Uria L., Etxepare R. 2011. *BASYQUE: Aplicación para el estudio de la variación sintáctica*. *Revista Linguamática* (ISSN: 1647-0818), V.3, nº 1, 35-44.
- Zuazo, K. 2008. *Euskalkiak: euskararen dialektok*. Elkar (ISBN: 978-84-9783-626-5).