

Fontes

Linguae Vasconum

50 urte. Ekarpen berriak
euskararen ikerketari

Nuevas aportaciones al
estudio de la lengua vasca

Testu-corpusen informazio
morfosintaktikoaren etike-
tatze automatikoa hizkuntz
ezagutzan oinarrituz: zenbait
arazo, hainbat erronka

ITZIAR ADURIZ, JOSE MARI ARRIOLA

Argitaratzaileak / Editores

Ekaitz Santazilia

Dorota Krajewska

Eneko Zuloaga

Borja Ariztimuño



Gobierno de Navarra
Nafarroako Gobernua

Testu-corpusen informazio morfosintaktikoaren etiketatze automatikoa hizkuntz ezagutzan oinarrituz: zenbait arazo, hainbat erronka

Rule-based automatic morphosyntactic tagging of corpus: some problems and many challenges

ITZIAR ADURIZ

Ixa taldea, Euskal Herriko Unibertsitatea UPV/EHU & Bartzelonako Unibertsitatea
itziar.aduriz@ub.edu

JOSE MARI ARRIOLA

Ixa taldea, Euskal Herriko Unibertsitatea UPV/EHU
josemaria.arriola@ehu.eus

JASOA: 2019/08/01 | BEHIN-BEHINEAN ONARTUA: 2019/10/03 | BEHIN BETIKO ONARTUA: 2019/12/11

LABURPENA

Maila morfosintaktikoan etiketatutako euskarazko corpusen desanbiguatze-lanetan urtetan aritu ondoren, bide horretan topatutako hainbat zailtasunen berri eman-go dugu artikulu honetan eta, horrekin batera, hainbat irizpide birplanteatzeko beharra ere azalduko dugu. Testuingurua hizkuntzalaritza konputazionala izanik, guk erabilitako metodologia erregeletan oinarritutako gramatikena da, hau da, informazio linguistikoa baliatuz aurrera eramaten dena. Zehazki, kategoria gramatikalaren eta azpikategoriaren etiketatzeari begira, hurbilpen linguistiko batetik desanbiguazio-emaitzak ulertzeko, analisi horien iturriak azalduko ditugu; nola erdiesten ditugun analisi horiek, zer zailtasun dituzten desanbiguatzeko, erregelak diseinatzerakoan kontuan hartu beharreko zenbait ezaugarri linguistiko eta etorkizunerako lanak.

Gako-hitzak: hizkuntzaren prozesamendua; kategoria gramatikalaren etiketatzea; erregeletan oinarritutako desanbiguatzea; corpus hizkuntzalaritza.

ABSTRACT

In this article, we will explain some disambiguation problems encountered after being involved for many years in disambiguation tasks of Basque corpora tagged at morphosyntactic level and the need of rethinking some criteria. Having in mind that our context is computational linguistics, our methodology is rule-based grammar, that is to say, the methodology is carried out based on linguistic information. Specifically, regarding the tagging of part of speech categories and subcategories, in order to understand the disambiguation results from a linguistic point of view, we show how these analyses are obtained, the difficulties which arise in disambiguating between linguistic (sub)categories, which linguistic issues arise in certain linguistic features to be considered when new rules are designed, and future perspectives.

Keywords: language processing; POS tagging; rule based disambiguation; corpus linguistics.

1. Sarrera

Egungo hizkuntzalaritzan, testu-corpusen erabilera eta ustiaketa gero eta zabalagoa izanik, corpusetako datu linguistikoak ustiatzeko ezinbestekoa dugu hitzen kategoria gramatikala etiketatzea (hemendik aurrera ingelesezko terminoa erabiliko dugu: *Part Of Speech (POS) tagging; word-class tagging* gisa ere ezagutzen da) (Giesbrech & Evert, 2009). Hizkuntza Naturalaren Prozesamenduak (HNP) joera horretan izan duen garrantzia ezin da ukatu. Zeregin hori testuko hitz bakoitzari kategoria zuzena esleitzean datza, testuingurua kontuan izanik. Haren izaera oinarritzakoarengatik giltzarritzat hartu izan da HNP aurrera eramaten den hizkuntza guztietan. Halaber, etiketatzeak corpusa aberastu eta hizkuntzalaritzako azterketa orokorrak egiteko aukerak ere zabaldu egiten ditu (Aduriz *et al.*, 2006; Ixa taldea & Elhuyar Fundazioa, 2007).

Hala ere, garrantzia aitortu arren, ikuspegi konputazionalako lan askotan, arrazoi praktikoengatik batik bat, ez dira morfologiaren prozesamenduaz arduratu. Zentzu batean, esan daiteke analisi morfologikoa eta *POS tagging* morfologia aberatseko hizkuntzetara mugatu direla. Horrela, HNPko sistemak morfologia aberatseko hizkuntzetara hedatzean sistema horiek sendotasuna eta doitasuna lortzeko erronka aitortu dute hainbat autorek (Tsarfaty, Seddah, Kübler & Nivre, 2013).

Euskararen kasuan, esate baterako, analisi morfosintaktikoaz (Aduriz *et al.*, 2000) hitz egitea ezinbestekoa dugu. Hizkuntza eranskaria izanik, atzizkiek informazio morfologikoarekin batera dakarte informazio sintaktikoa ere. Hori dela eta, corpusa etiketatzerakoan, kategoriez gain, beste informazio gramatikalaz eta sintaktikoaz ere arduratu beharko gara: bai izenen flexio-morfemek dakarten informazio gramatikalaz (kasua, numeroa, mugatutasuna), bai aditzen inflexioan aurkituko dugun informazioaz, esaterako, denbora, modua, aspektua eta komunztadura. Horren informazio aberatsa izanik, corpusa etiketatzerakoan anbiguotasun morfosintaktikoari aurre egin beharko diogu, gerta baitaiteke forma bat baino gehiago anbiguo izatea aipatutako maila horietako edozeinetan (Aduriz *et al.*, 2003).

HNPan etiketatze eta desanbiguazio morfosintaktikoak oinarritzko tresnak lantzearen estrategian kokatzen dira. Horrela, 90etan sortu zen lehen desanbiguatzailea euskarako, EUSLEM izenekoa (Aduriz *et al.*, 1996), eta ordutik hona asko hobetuz joan den tresna giltzarria izan da, ondoren etorri diren lan aurreratuagoen oinarri: ANALHITZA (Otegi, Imaz, Díaz de Ilarraza, Iruskieta & Uria, 2017).

Urte hauetan izandako esperientziak erakutsi digu badela konpontzen zaila den kategoria-anbiguotasunaren mota bat, hizkuntzak berak eragindakoa, adjektibo-adberbioen kasua (*azkar*), adibidez, edo izen-adjektiboen kasua (*garai*), eta bestalde, oinarri lexikoaren informazioak eragindako beste bat, Euskararen Datu-base Lexikalean (EDBL) (Aldezabal *et al.*, 2001) oinarritzetik dator. Azken horri dagokionez, *aurrekoan* hitza analizatzen badugu, bi sarreretatik emango digu analisia, esanahi bera izanik: analisi morfologikoaren bidetik datorrena (*aurre + ko + an*) eta sarrera lexikalaren bidetik eratorria (*aurrekoan* sarreratik). Bai batak eta bai besteak estrategikoki soluzio-bide desberdinak planteatzen ditu desanbiguatzailearen helburuaren aurrean, hau da, ahalik eta hitz gehienak desanbiguatua izatea.

Artikulu honetan desanbiguate-lanetan urtetan aritu ostean topatutako kasu arazotsuak azalduko ditugu, alegia, azalduko diren arrazoi desberdinengatik desanbiguatzen zailak diren kasuak. Kasu horiek azal dezakete, parte batean, desanbiguate-lanetan doitasuna hobetzeko saiakeretan maiz % 100era ez iristearen hormarekin jo izana. Era berean, lan honetan aurkezten dugun hausnarketak, gure iritziz, lagun dezake POS etiketatzeari begira hurbilpen linguistiko batetik emaitzak hobetzeko har daitezkeen aukerak azaltzen. Izan ere, euskararako landu dugun POS etiketatzaileak dituen zenbait gaitz EDBLn dagoen informazioa moldatuz konpondu edo, behintzat, arindu daitezkeela iruditzen zaigu. Manningek (2011) bezala, uste dugu oraindik badela informazio linguistikorako tarterik. Hasieran, sarrera orokorrean, esparrua kokatu dugu. Gero, bigarren puntuan, *POS tagging* delakoaren garrantzia eta hurbilpen nagusiak aurkeztuko ditugu laburki; hirugarren puntuan, POS etiketatzea analisi-katean kokatu eta kate horretako zenbait osagai aurkeztuko ditugu. Laugarrenean, desanbiguate-lanak zailtzen dituzten zenbait kasu azalduko ditugu, eta, bukatzeko, zenbait ondorio eta epe erdirako helburu finkatuko ditugu.

2. POS etiketatzearen garrantziaz eta hurbilpen nagusiez

HNPan corpusak etiketatzeko modu ohikoena hitzen kategoria gramatikala etiketatzea da. Zeregin hori testuko hitz bakoitzari kategoria zuzena esleitzean datza, testuingurua kontuan izanik. Halaber, maiz, prozesamendu sintaktiko azalekoaren lehen urratsa ere bada askorentzat. Izan ere, HNPko oinarrietako bat dugu, hitz-adiera desanbiguazioarekin eta preposizio-sintagmen anbiguotasunaren (*PP attachment* izenez ezagutua ingelesez) fenomenoekin batera. Ildo beretik, gaur egunean puri-purian dagoen txioen aztergaiari ekiterakoan ere, urrats horren garrantzia nabarmentzen dute Derczynskik, Ritterrek, Clarkek eta Bontchevak (2013, 198. or.): «Part-of-speech tagging is a central problem in natural language processing, and a key step early in many NLP pipelines». HNPan garrantzitsua den urrats hori, zertarako behar dugu? Informazioaren berreskurapenerako (ingelesez, *Information Retrieval*, IR), testua ahots bihurtzeko (*Text to Speech* ingelesez), hitz polisemiko baten adiera zehatza zein den zehazteko (*Word Sense Disambiguation*, WSD) eta *parsing* edota analizatzaile sintaktikoen lehenengo urrats gisa; azken batean, abiapuntuan hitz bakoitzeko analisi bakarra izateak analisi sintaktikoaren aukerak

murritzuko baititu. Halere, Kirovek, Sylak-Glassmanek, Knowlesek, Cotterellek eta Postek (2017, 112. or.) dioten moduan, tradizionalki analisi morfologikoa eta *POS tagging* delakoa morfologia aberatseko hizkuntzetara mugatu da. Horrela, gehienetan analizatzaile sintaktikoetan integraturiko analizatzaile morfologikoen POS etiketatzaileen azaleko informazioa erabiltzen dute edota hitz anitzeko unitate lexikalen zerrenda lauak, esaterako Standfordeko *CoreNLP tool* tresnak egiten duen bezala (Mckinlay & Baldwin, 2005).

Euskararen kasuan, sarreran aipatu bezala, analisi morfosintaktikoa nahitaezkoa dugu HNPko edozein aplikazioren aurrean. Lehenengo urrats batean, analisi morfologikoen bitartez (ik. 1. irudia) hitz-forma bakoitzeko lema eta kategoriak ematen dira. Halaber, ondorengo urratsetan, maila bakoitzean gero eta informazio xeheagoa desanbiguatzeaz arduratuko da, hala nola azpikategoria, kasua, mugatutasuna eta mendeko perpaus-mota. Beraz, desanbiguazioa dugu prozesamendu sintaktiko sakonagoei ekin ahal izateko nahitaezko lehen urratsa.

Nahitaezkoa eta oinarritzkoa izateaz gain, esan dezakegu desanbiguatze-prozesu hori ez dela hutsala; izan ere, morfologiaren prozesamendua eskatzen du: prozesu morfologiko konplexuak, inflexioa nahiz eratorpen-morfologia kontuan hartzea eta hainbat anbiguotasuni aurre egitea (kategoria gramatikala, azpikategoria, kasua, mugatutasuna, perpaus-mota), besteak beste. Halere, gaur egun kategoria gramatikala etiketatzen duten analizatzaile (*POS tagger*) sendoak eta azkarrak ditugu. Manningek (2011) dioen bezala, *POS tagger* gehienek % 97ko doitasuna erdiesten dute *tokeneko*. Beraz, nolabait ez ote da jada konponduriko arazo bat? Baliteke hori izatea pertzepzio nagusia HNPko komunitatean, baina, beharbada, dena ez da hain arrosa-kolorekoa (Manning, 2011, 171. or.):

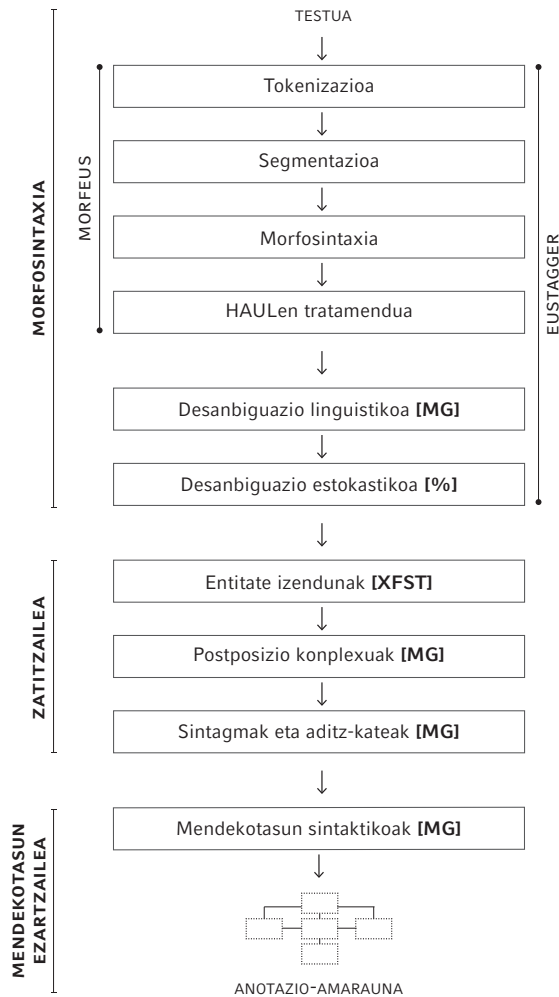
It is perhaps more realistic to look at the rate of getting whole sentences right, since a single bad mistake in a sentence can greatly throw off the usefulness of a tagger to downstream tasks such as dependency parsing. Current good taggers have sentence accuracies around 55-57%, which is a much more modest score. Accuracies also drop markedly when there are differences in topic, epoch, or writing style between the training and operational data.

Halaber, lan horretan aipatzen diren hiru baieztapen ekarriko ditugu hona: (1) % 97ko doitasun hori ezin dela hobetu, (2) giza etiketatzaileen arteko adostasuna ere ez dela % 97 baino handiagoa eta (3) emaitzak hobetzeko ikasketa automatikoko sistemak deskribapen linguistikoa hobetu behar dutela. Alegia, etiketatzearen kontsistentzia sendotzeko, oinarri linguistikoak jorratu behar direla proposatzen du. Hain zuzen ere, guk ere alde linguistikoaren ikuspegitik heldu nahi diogu ikerlerro honi. Izan ere, doitasuna hobetzeko saiakeretan maiz jo izan dugu emaitzak % 100era iritsi ezineko hormarekin. Saiakera horietan metodo sinbolikoak eta estokastikoak konbinatu izan ditugu. POS etiketatzeari ekiteko hurbilpenak askotarikoak izan badira ere, hauek dira ohikoenak: erregetan oinarritutakoa edo sinbolikoa, estokastikoa, ikasketa automatikoa (*machine learning*, ingelesez) eta hurbilpen hibridoak, adibidez, metodo sinbolikoen eta estokastikoen konbinazioa eginez (Ezeiza, Aduriz, Alegria, Arriola & Urizar, 1998). Halere, modu batera edo bestera, hurbilpen guztietan etiketa-sistema (*tagset*, ingelesez) definitzea oso

garrantzitsua da eta hainbat proiektutan garatu izan dute *POS tagging*a egiteko etiketa-sistema berezia, nahiz eta etiketa-sistema horiek definitzeko printzipioen lanketan gutxi batzuk aritu (Baskaran *et al.*, 2008). *Constraint Grammar* formalismoan (Karlsson, Voutilainen, Heikkilä & Anttila, 1995) egin izan da lanik handienetakoa etiketen taxonomia hobetzeko etiketatzailer automatikoen esparruan.

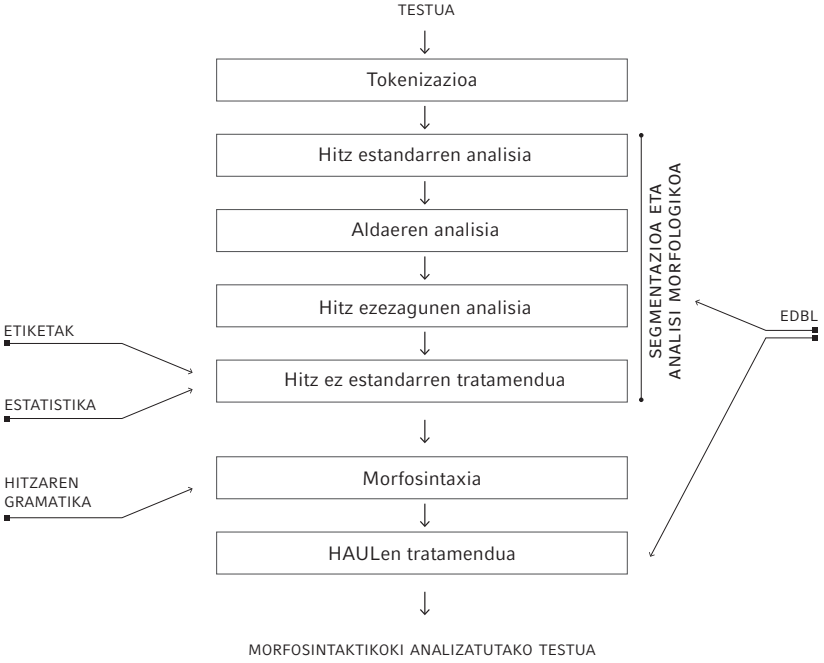
3. Analisiaren katea

Atal honetan etiketatze-lanak aurrera eramateko ditugun analisi linguistikoen oinarriak izango ditugu mintzagai, alegia, analisi-katea (ik. 1. irudia).



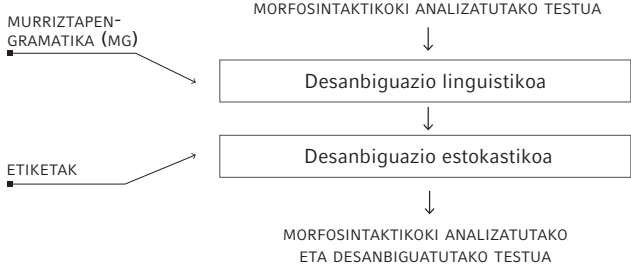
1. irudia. Analisi-katea.

Analisi-katean ikus daitekeen bezala, mailaka egiten da testuen analisia, modulu bakoitza aurrekoaren analisisan oinarritzen baita. Oinarrian analisi morfosintaktikoa eta desanbiguatze morfosintaktikoa ditugu. Analisi morfosintaktikoaz, MORFEUS izeneko tresna (ik. 2. irudia) arduratzen da. Tresna horrek hitz bakoitzeko analisi morfologikoa- ren aukera guztiak ematen ditu.



2. irudia. Analisi morfosintaktikoa.

Bestalde, desanbiguazio morfosintaktikoari aurre egiteko bi modulu ditugu: bata linguistikoa eta bestea estokastikoa. Bi modulu horiek osatzen dute EUSTAGGER (ik. 3. irudia). EUSTAGGER izeneko tresnak analisi morfosintaktikoan emandako aukeren artetik analisi bat hautatuko du, alegia, desanbiguatze morfosintaktikoa gauzatuko du.



3. irudia. EUSTAGGER, lematizatzailea/etiketatzailea.

Prozesamenduaren katearen hasieran (Ezeiza, 2002) lau urrats nagusi egiten dira: tokenizazioa, segmentazioa eta analisi morfologikoa, morfosintaxia eta hitz anitzeko unitate lexikalen (HAUL) tratamendua. Atal honetan, bada, *handiena* hitzaren adibidea baliatuko dugu (ik. (1) adibidea) hasierako prozesaketan esku hartzen duten hiru moduluak azaltzeko, labur bada ere. Ondoren, EUSTAGGERen zereginari helduko diogu, zehazki, desanbiguazio linguistikoari. Izan ere, desanbiguazio linguistikoaren atal horretan POS etiketatzeaz arduratzen gara.

Hartara, *handiena* hitzaren analisiari erreparatzen badiogu, batetik, hitzaren morfemakako osakeraren berri dugu eta, bestetik, osagai bakoitzari dagokion informazio morfologikoa, alegia, lema dagokion kategoria gramatikala eta atzizkiei dagokien informazio morfologikoa. Hitz beraren kategorizazio desberdin guztiak ematen zaizkigu, izan ere, hitza ager daitekeen testuinguruak aurreikusten baitira. Analizatzaile morfologikoak informazio hori eman ahal izateko oinarri lexikala EDBL dugu, prozesamendu orokorrerako ezinbesteko oinarria eta adibidean ikus daitekeen bezala, atzizkiak ere maila lexikoan gordetzen dituen. Kategoria gramatikal bakoitzari dagokion morfema-segida ere EDBLn adierazten zaio sarrera bakoitzari eta informazio hori guztia bi mailatako formalismoaren bidez antolatzen da (Koskeniemi, 1983).

Hau da, segmentatzaile morfologikoaren emaitzak hitzaren osakeraren berri ematen digu (adib. *handiena* hitzaren morfemakako osaera: *handi + en + a*). Horretan oinarrituta, analizatzaile morfologikoak hitza osatzen duen osagai bakoitzari dagokion informazio morfologikoa esleitzen dio:

(1) *handiena*:

- a. *handi* [adjektibo] + *en* [gradu][superlatibo] + *a* [mugatu][singular]
- b. *handi* [adjektibo] + *en* [deklinabide][genitibo][mugatu][plural] + *a* [mugatu][singular]

Alegia, *handi* lema bera izanda ere, atzizkiaren sinkretismoa ikusten dugu: a) *-en* morfema, graduatzaile superlatibo edo b) genitibo mugatu plurala izan daiteke.

Analisi morfosintaktikoaren helburua, beraz, testu-hitzaren azken analisi morfosintaktikoa ematea da, betiere emaitza osoaren berri emanaz, barruan dagoen informaziozirik galdu gabe; alegia, sintaxiaren atarian utzi behar ditugu analisiak gerora baliagarri izan daitezen (ik. 3. adibidea). Hain zuzen ere, analisi hori dugu hurrengo urratsaren abiapuntu (ik. 3. irudia).

Lan honi begira, dagokigun atala 3. irudian azaltzen dena izango litzateke, hain zuzen ere, EUSTAGGER. Abiapuntu gisa morfosintaktikoki analizatutako testua izango dugu, hau da, hitzen segmentazioa oinarri hartuta, osagai bakoitza dagokion informazio morfosintaktikoarekin hornituko da eta osagai horien arteko kateatzeak formalizatuko dira zehatzuz zer morfema-segida diren zilegi kasu bakoitzean (adib. *baitator* onartuko dugu, baina **baitatorrela* ez da onartuko). Horretarako bi mailatako erregelak baliatuko dira. Era berean, morfemen pilaketaren ondorioz dugun informazioa optimizatzeke, analizatzaile morfosintaktikoa arduratuko da jaso beharreko informazioaren kudeaketaz. Desanbiguatze linguistikoari heldu aurretik, izango dugun informazioa ilustratze

aldera, *bizkor* hitzaren adibidean kategorizazio desberdin guztiak kontuan hartutako analisia ikus dezakegu:

(2) *bizkor*:

```
bizkoR[[bizkortu0][aditza][aditzoina_bizkor]]+[[aditz_mota_morfema][aditzoina]]
bizkoR[[bizkor1][adberbio][arrunt]]
bizkoR[[bizkor2][adjektibo][arrunt][IZAUR1-]]
bizkoR[[bizkor2][adjektibo][arrunt][IZAUR-]]+[[deklinabide][absolutibo][mugagabe]]
```

Beraz, (1) eta (2) adibideetan ikusi dugun bezala, hitzak segmentatuta eta morfologikoki analizatuta ditugu. Funtsean, analisi horiek dira oinarri gero etorriko diren analisi-urrats edo aplikazioetarako (lematizatzailea, analisi sintaktikoa...). Halere, morfologikoki analizaturiko hitz horiek, ondoren tratamendu morfosintaktikoa izango dute (Aduriz *et al.*, 2000), morfema-pilaketaren ondorioz izango den informazioa optimizatzeko eta osagai bakoitzetik zer informazio goratu erabakitzeke. Hona hemen, adibidez, *lan-baldintzen* hitz elkartuaren segmentazioa eta analisi morfosintaktikoa:

(3) *lan-baldintzen*

Sarrera: lan-baldintzen

Irteera:

Segmentazioa

(lan)	LEMA	IZEN ARRUNT	-BIZIDUN
(-)	MARRATXOA		
(baldintza)	LEMA	IZEN ARRUNT	-BIZIDUN
(-en)	ATZIZKI GENITIBO PLURALA	+MUGAT (@<IZECOMPL @IZECOMPL> ²)	

Analisi morfosintaktikoa:

[lan + - + baldintza + en]

IZE ARR -BIZ GEN PL + MUG IZE+IZE (@<IZECOMPL @IZECOMPL>)

Hitz elkartu horren adibidean ikus daiteke osagai bakoitzetik zer informazio hartuko den hitzaren analisiaren azken emaitza gisa. Informazio hori guztia antolatzeaz arduratzen da analizatzaile morfosintaktikoa. Nolanahi ere, azpimarratu behar dugu EDBL dela informazio horren iturri, beraz, ezinbestekoa da haren informazioa antolatzeko zenbait irizpide kontuan hartzea.

¹ IZAUR ezaugarriak izenaren aurretik joan daitekeen adierazten du, beraz, balioa «-» denean ez dela gai izenaren aurretik agertzeko esan nahi du.

² @<IZECOMPL: funtzio sintaktiko honen bidez modifikatzen duen izena ezkerrean izango duela adierazten da eta @IZECOMPL> funtzioarekin, berriz, modifikatzen duen izena eskuinean agertzen dela.

3.1. EDBLko informazio linguistikoa

EDBL euskararen tratamendu automatikorako euskarri lexikal eta gramatikal orokorra da. Beraz, analisisetan baliatuko dugun informazio linguistikoa EDBLtik jasotzen dugu.

Artikulu honetan aurkeztu nahi ditugun arazoak eta horren adibide diren kasu arazotsuak ongi ulertzeko nahitaezkoa da datu-base lexikal horrek ematen digun informazioari buruzko zenbait ezaugarri kontuan hartzea. Izan ere, (4) adibidean ikusten den bezala, *zurrutean* hitzak bi analisi posible ditu: a) EDBLko sarrera gisa (*zurrutean* osorik) eta b) analizatzaile morfologikoaren emaitza (*zurrut + INE*):

(4) *zurrutean*

Analisi-1: *zurrutean*N[[Sarrera_zurrutean—1—+][adberbio][arrunt]]

Analisi-2: *zurrut*[[Sarrera_zurrut—1—+][izen][arrunt][BIZ_-]]

+0[[Sarrera_0—5—-][deklinabide] [mugatu][singular]]

+Ean[[Sarrera_an—-][deklinabide][inesibo][funtzio sintaktikoa[@ADLG]]]

Hitz honek bi analisi-iturri ditu EDBLN: lehenengoa adberbio sarrera da, *zurrutean* hitza hiztegi-sarrera gisa dagoelako datu-basean, eta bigarren analisisian, berriz, *zurrut* izena agertzen da, inesibo kasuan.

Informazio hori guztia EDBLtik jasotzen du analizatzaile morfologikoak eta erabili hitz bat analizatzen dugunean.

Ikusiko dugu gero, 4. puntuan, kasuistika zabalagoa aztertuz, zenbait hitzek azaltzen duten analisi bikoiztasun hori beharrezkoa ote den.

Ikus dezagun aurretik EDBLko informazio lexikalaren iturri nagusia zein den eta lexikoa lantzeko zein irizpide erabiltzen dugun, hobeto ulertzeko nondik datozen (4) bezalako kasuak.

3.2. EDBLko informazio lexikalaren iturriak eta sarrerak lantzeko irizpideak

Gaur egunean EDBL lantzeko iturri nagusiak hauek dira: 1) *Euskaltzaindiaren hiztegia* (Euskaltzaindia, d. g. a) eta 2) Lexikoaren Behatokia corpus lematizatua (Euskaltzaindia, d. g. b). Azken hau Ixa taldeko tresnekin lematizatzen da eta hortik hautematen da zer dagoen corpusean EDBLN ez dagoena.

*Euskaltzaindiaren hiztegi*ko informazioa erabiltzeko irizpideei dagokienez, batez ere zein azpisarrera landuko den zehazteko, EDBLko azpitaldeak landutako txosten bat erabiliko dugu eta haiek zehaztutako ondorioak azalduko ditugu puntuka³.

3.2.1. EDBLN sarrera dutenak

Oro har, sarrera ematen zaie honako hauei:

- Eratorpen atzizkia duten hitzei, kategoria gramatikala aldatzen baitute gehienetan, adibidez: *bortxaka*, *bulartsuki*.
- Kasu-marka dutenei, ez beti, ordea. Halako baldintzak dituztenei:

³ Komeni da argitzea EDBLko egituran ez dagoela azpisarrerarik eta sarrera guztiak maila berekoak direla.

- . Batez ere, zentzua aldatzen bazaio (ez-konposizionala): *hankapetik* (= erraz), *begietaraino* (= erabat, osoki).
- . Jakina, kategoria aldatzen bada: *gogoetan*, *hizorian*, *bidegabez*, *egarriz*, *bizkorrik* (izenetik eta adjektibotik adberbiora, kasu hauetan bezala).
- . -zkodunak, oro har, sartu dira sarrera moduan, adibidez: *nahigabezko*, *nahizko*, *nirezko*, *partezko*, *presazko*, *printzipiozko*, *semezko*.
- . Aditz-izenei, batez ere kasu hauetan:
 - . Jakintza-alor batekoa bada, adibidez: *erdiratze-kirol*.
 - . Oro har, OEHN adiera berezi batekin (terminoa) agertzen bada (ez ekintza soila, alegia), adibidez: *arintze*: *zerga-arintze*, *zigor-arintze*.

3.2.2. EDBLn sarrerarik ez dutenak

Ez zaie sarrera eman, ordea, honako hauei:

- . Honelako atzizkiak hartzen dituztenak: *-na* (*-naka*, *-nakako*), *-garren* bezalakoak hartzen dituztenak, adibidez: *bana*, *hiruzpalauna*, *bosnaka*, *binakako*, *bosgarren*; nahiz kasu-marka izan, kasu hauetan bezala: *-tan*, *-tara*, adibidez: *bitan*, *hirutara*. Izan ere, atzizki hauek zenbatzaile guztietara orokortzen ahal dira.
- . *-xe* atzizkia hartzen dutenei ere ez, adibidez: *hauxe*, *horixe*, *huraxe*.
- . *-tzedunei*, oro har: *-tze* atzizki lexikala EDBLn dagoelako aditz-izenak sortzeko analisi morfologikoaren bidez.
- . Izenordain posesiboak: *nire*, *heure*, *geure*.
- . Adberbioen gainetik osatuak, esanahi berezirik gabe:
 - . *-ko*: *hango* (*han* + *-ko*), *bezainbateko* (*bezainbat* + *-eko*), *etziko* (*etzi* + *-ko*).
 - . *-danik*: *noizdanik*, *aspaldidanik*, *betidanik*.
 - . *-tik*: *noiztik*, *hemendik*, *iaztik*, *handik*, *aspalditik*.
 - . *-raino*: *haraino*, *beranduraino*.
- . Leku-denborazko kasu-marka askorekin erabiltzen badira: *jendartean*, *-tik*, *-ra*...; *barruan*, *-tik*, *-ra*, *-raino*...
- . Baina bai: *lekutan*, *-tara*, *-tako*, *-tatik*... esanahi berezia hartzen baitu kasu hauetan (= *urrun*).
- . Postposizio gisa edo menderagailu-testuinguruan bakarrik erabiltzen direnetan: *iritzian* (*-en* atzizkiaren eskuinean).
- . Esamolde osoago batean ageri diren formak (*bitsetan* → *zorion-bitsetan*; *izerdi-bitsetan*).
- . *Euskaltzaindiaren hiztegian* erabilera bereziren baten berri emateko erabilia, baina konposizionala eta EDBLn sistematizatua, adibidez: *betidanik*, *betitik*.

3.2.3. Zalantzazkoak

Nolanahi ere, duda asko sortzen dira honako kasu hauetan, esate baterako:

- . Esanahia konposizionala den ala ez ebaztean: *agerian* formaren kasuan, adibidez. *Ageri* (adj.) sarrera badago EDBLn eta baita *Euskaltzaindiaren hiztegian* ere. Zalan-

tza da konposizionala ez den irakurketa ba ote duen hitz honek, sarrera berezitua eskatzen duena: «*Agerian* hitz egin: jendaurrean hitz egin», adibidez.

- Postposizioen kasuan, adberbio beregain gisa ere erabil ote daitekeen zehaztean (aurrerago aipatutako *iritzian* formaren kasuan edo *arabera* formaren kasuan, adibidez).
- Oro har, zenbateraino dagoen lexikalizaturik erabakitzean.

4. Sorkuntza morfologikoa vs sarrera lexikoak

EDBL elikatzeko erabiltzen den hiztegitik datorren informazioari begira batez ere, sarrera lexikotzat jotzeko irizpideak zeintzuk diren aztertu dugu aurreko puntuan.

Ikusi bezala, zalantza handiak egoten dira, irizpide-mailan, erabakitzeko iturriko hiztegian sarrera ala azpisarrera gisa agertzen diren hitzak EDBLn sarreratzat hartu behar diren ala ez. Gogoratu behar dugu EDBLko sarrerek dagozkien numero-mugatutasun eta kasu-markak hartzen dituztela, baina behar denean, batez ere semantikak aginduta; alegia, lexikalizatutako formak morfosintaxiak eragindako analisiak ez bezalako adiera duenean, numero-mugatutasun eta kasu-marka horiek barne hartzen dituzten sarrerak ere lantzen dira.

Horren adibide, aurretik aipatutako *zurrutean* hitz-forma aurkitzen dugu (4. adibidea), non batetik, *zurrut* formatik *zurrutean* eratortzen den analisi bidez eta, bestetik, *zurrutean* forma osorik landuta dagoen, semantikoki berezitua delako.

Irizpideen azalpenean azaldu bezala, askotan zaila da bereizten zein formak duen adiera berezitua eta, ondorioz, sarrera behar duen ala ez.

Horrela, *zurrutean* hitzaren kasuan garbi samar baldin badago bi sarrerak behar dituela⁴, beste kasu askotan ez dago arrazoi morfosintaktikorik, kategoria berririk ez duelako sortzen numero-mugatutasun ala kasu-marka duen sarrerak, eta semantikoki ere ez duelako hartzen esanahi bereziturik. Horregatik desanbiguatu ezin diren formak sortzen dira. Alegia, analizatzaileak bi analisi aukera ematen dizkigu, nahiz eta linguistikoki ez diren anbiguo, eta, beraz, nekez desanbiguatuko dira. Adibidez, *kanpora* hitzarekin gertatzen den bezala:

- Sorkuntza morfologikoa: EDBLn dagoen sarrera lexikala: *kanpo* + kasu-marka: *-ra* → *kanpora*.
- EDBLko sarrera lexikala, kasu-markaduna, adberbio kategoriaduna, hiztegitik azpisarreratik datorrena: *kanpora*.

Desanbiguatu ezin diren formak egoteak dakarren bikoiztasunak ez dio mesederik egiten sistemaren zurruntasunari eta txukuntasunari, oro har. Baina bada aplikazio bat arazo hori begi-bistakoago eta larriago bilakatzen duena: desanbiguatzailearena (ik.

⁴ *Zurrut* + *-ean* → «Bere *zurrutean* aitaren antza ikusten zitzaion». *Zurrutean* (= *zurrut eginez*): ik. 4. adibidea.

3. irudia). Izan ere, EDBLren informazio lexiko eta morfosintaktiko guztiaren gainean egiten da formen desanbiguazioa eta bikoiztasun horiek kalte egiten diote sistemari, desanbiguaezinak diren formak geratzen baitira bikoiztuta arrazoirik gabe. 1. irudian ikusten den bezala, sistemaren arkitektura dela-eta, desanbiguatzailearen irteera, aldi berean, analizatzaile sintaktikoaren sarrera da, eta puntu horretan, alferrikako forma bikoiztuak egoteak kalte egiten du. Izan ere, anbiguotasun-kasuetan, ahal den neurrian, desanbiguatzaileak analisi-lerro bakarra utzi beharko luke hurrengo urratsean modu eraginkorrean lan egin ahal izateko (Karlsson *et al.*, 1995).

Desanbiguazio-lanetan urtetan aritu ondoren (Aduriz *et al.*, 2003), aipatzen ari garen arazoaren kasuistika hiru multzo nagusitan jasotzen dela ikusi ahal izan dugu:

- *-ko* genitiboaren bidez sortzen diren formek osatua.
- Leku-kasuen bidez sortzen direnek osatua.
- Instrumentalaren bidez sortutakoek osatua.

4.1. *-ko* genitiboaren bidez sortzen direnak

Euskaraz, oso bide emankorra da izenetatik edo adberbioetatik adjektiboak ala izenak sortzea *-ko* genitiboaz atzizkia erabilita (Euskaltzaindia, 1991; Odriozola, d. g.). Oinarritzko lan horietan ez ezik, euskarari buruzko bibliografian maiz hitz egin izan da prozesu honi buruz. Orain, ordea, ez gara hasiko xehetasunak ematen, baina egia da oso maiz azpisarrera gisa jasotzen direla hiztegitan, dagokion izenaren azpian, nagusi-nagusiki. Asko leudeke aipatzeko, adibidez: *aspaldiko*, *eguneroko*, *gaurko*, *nolabaiteko*, *etxeko*, etab.

Horietako askok esanahi berezitua hartzen du, gehien-gehienetan lexikalizazioen kasuetan. Horrela, *etxeko* edo *eguneroko* hitzek, adibidez, izen eta adjektibo semantikoki beregainak sortu dituzte: *etxeko*, izenak, ‘famiarteko’ eta adjektiboak, ‘hurbileko’ adierak hartzen ditu. *Eguneroko* hitzak berriz, ‘egunkari’ (iz.) adierakoa, ‘egun guzietakoa’ (adj.) adieradunaren ondoan.

Garbi dago bigarren adibide honen kasuan izenari lotutako adiera lexikalizatuagoa dagoela adjektiboari lotutako adierarena baino. Izan ere, bigarrena, adjektibo kategoriaduna konposizionalki lortzen ahalko litzatekeen adiera da.

Egunero eta *etxe* adberbio gisa eta izen gisa agertzen dira EDBLn landuta, hurrenez hurren, eta *-ko* atzizkia hartzen dute. Horrez gain, *eguneroko* eta *etxeko* sarrerak ere landuta daude EDBLn, *Euskaltzaindiaren hiztegitik* datozelako. Berrero ere, morfologikoki desanbiguatzen zailak diren kasuen aurrean geundeke.

4.2. Leku-denborazko atzizkiak dituztenak

Leku-denborazko atzizkiak hartzen dituzten kasuekin antzeko portaerak topatzen ahal ditugu hainbat eta hainbat kasutan. Horrela, adibidez, *behera* edo *kanpora* kasuetan, sarrera moduan landuta daude EDBLn, adberbio kategoriadun sarrera moduan, esanahi bereziturik gabe (esanahi konposizionala baitute) eta bestetik, *behe* eta *kanpo* hitzek, bere aldetik, izen kategoriadun sarrerak dituzte, dagozkien atzizkien

artean leku-denborazko atzizkiak hartzen dituztelako, izenek ohiko duten gisan, eta hortik ere *behean*, *behetik* eta *kanpora* formak eratortzen dira, adberbio funtzio sintaktikoa dutenak.

Adibide moduan eta zenbakietara joaz, EDBLko 457 kasutan ematen da bikoiztasun hau: adberbio sarrera duten hitzak eta izen-kategoriadunak inesiboan⁵.

4.3. Instrumentalaren bidez sortutakoak

Instrumentalaren bidez sortutako hitzetan, azkenik, aurreko bi puntuetan aipatu ditugun kasuetan gertatzen den kasuistika bera gertatzen da, leku-kasuetan baino maiztasun txikiagoarekin gertatzen bada ere. Horrela, 239 kasutan gertatzen dira bikoiztasunak.

Adibidez, *irribarrez* eta *indarrez* sarrerak daude landuta EDBLn, adberbio kategoria baitute, eta bestetik, *indar* eta *irribarre* izenak, instrumental kasua hartzen dutenak.

5. Ondorioak

Lan honetan euskararen prozesamenduaren alorreko oinarrizko atal batean kokatu gara, corpusen etiketatze eta desanbiguazio morfosintaktikoan, hain zuzen ere. Etiketatze-lan hori hainbat hurbilpenetatik egin badaiteke ere, irizpide linguistikoei erreparatu nahi izan diegu: kategoria gramatikalaren eta azpikategoriaren desanbiguazioari. Prozesu konplexu honetan, desanbiguatzeko-lana eragozten duten hainbat zailtasunekin egin dugu topo. Bereziki, EDBLtik datorren informazioak eragindako zenbait analisiren bikoiztasuna azaldu dugu, zergatiak aztertu eta adibideak eman. Analisi bikoitz horien azterketa egin ondoren, uste dugu euskararen tratamendu automatikorako oinarri den EDBLren informazioa hornitzeko irizpide batzuk moldatu eta findu behar direla.

Irizpide moldatze horrek etiketatze-lanez gain, ondorengo prozesamenduko urratsetan izan dezakeen eragina kontuan izanik, hausnarketa sakona eskatzen du. Azken einginkizun hau etorkizunerako lanen artean sartuko litzateke.

Erreferentziak

Aduriz, I., Agirre, E., Aldezabal, I., Alegria, I., Arregi, X., Arriola, J. M., ... Urkia, M. (2000). A Word-grammar based morphological analyzer for agglutinative languages. In *Proceedings of the International Conference on Computational Linguistics. COLING 2000* (1-7. or.). Saarbrücken: Universität des Saarlandes.

Aduriz, I., Aldezabal, I., Alegria, I., Artola, X., Ezeiza, N. & Urizar, R. (1996). EUSLEM: A lemmatiser/tagger for Basque. In M. Gellerstam, J. Järborg, S.-G. Malmgren, K. Norén, L. Rogström & C. Rödger Pappmehl (arg.), *Euralex '96. Proceedings* (1, 17-26. or.). Göteborg: Göteborg University.

⁵ Kontuan hartu EDBLn 4636 adberbio sarreradun hitz daudela.

- Aduriz, I., Aranzabe, M., Arriola, J. M., Atutxa, A., Díaz de Ilarraza, A., Ezeiza, N., ... Urizar, R. (2006). Methodology and steps towards the construction of EPEC, a corpus of written Basque tagged at morphological and syntactic levels for the automatic processing. In A. Wilson, P. Rayson & D. Archer (arg.), *Corpus Linguistics Around the World. Language and Computers 56* (1-15. or.). Amsterdam-New York: Rodopi.
- Aduriz, I., Arriola, J. M. & Diaz de Ilarraza, A. (2003). Desanbiguazio morfologikoa, azterketa sintaktikoaren lehen urratsak eta aplikazioak Murritzapen Gramatikaren eredu konputazionala jarraituz. In J. M. Makatzaga & B. Oyharçabal (arg.), *P. Lafitteren sortzearen mendemugako biltzarra. Euskaltzaindiaren XV. Biltzarra. Euskal Gramatikari eta Literaturari buruzko ikerketak XXI. mendearen atarian. Gramatika Gaiak. Iker 14* (3-35. or.). Bilbo: Euskaltzaindia.
- Aldezabal, I., Ansa, O., Arrieta, B., Artola, X., Ezeiza, N., Hernández, G. & Lersundi, M. (2001). *EDBL: a General Lexical Basis for The Automatic Processing of Basque*. IRCS Workshop on Linguistic Databases ekitaldian eginiko hitzaldia, Philadelphia, AEB.
- Baskaran, S., Bali, K., Bhattacharya, T., Bhattacharyya, P., Nath Jha, G., Rajendran, S., ... Subbarao K. V. (2008). Designing a Common POS-Tagset Framework for Indian Languages. In *Proceedings of The 6th Workshop on Asian Language Resources* (89-93. or.). Hyderabad, India.
- Derczynski, L., Ritter, A., Clark, S. & Bontcheva, K. (2013). Twitter part-of-speech tagging for all: overcoming sparse and noisy data. In G. Angelova, K. Bontcheva & R. Mitkov (arg.), *Proceedings of the International Conference Recent Advances in Natural Language Processing (RANLP)* (198-206. or.). Shoumen: INCOMA.
- Euskaltzaindia. (1991). *Euskal gramatika lehen urratsak-I (EGLU-I)*. Bilbo: Euskaltzaindia.
- Euskaltzaindia. (d. g. a). *Euskaltzaindiaren hiztegia*. https://www.euskaltzaindia.eus/index.php?option=com_hiztegiabilatatu&view=frontpage&Itemid=410&lang=eu helbidetik eskuratua.
- Euskaltzaindia. (d. g. b). *Lexikoaren Behatokia [datu-basea]*. <http://lexikoarenbehatokia.euskaltzaindia.eus/> helbidetik eskuratua.
- Ezeiza, N. (2002). *Corpusak ustiatzeko tresna linguistikoak. Euskararen etiketatzaile morfosintaktiko sendo eta malgua* (doktore-tesia). UPV/EHU, Donostia.
- Ezeiza, N., Aduriz, I., Alegria, I., Arriola, J. M. & Urizar, R. (1998). Combining stochastic and rule-based methods for disambiguation in agglutinative languages. In *Proceedings of the International Conference on Computational Linguistics and Annual Conference of the Association for Computational Linguistics COLING-ACL'98* (380-384. or.). Montreal, Kanada.
- Giesbrech, E. & Evert, S. (2009). Part-of-speech tagging – a solved task? An evaluation of POS taggers for the Web as a corpus. In I. Alegria, I. Leturia & S. Sharoff (arg.), *Proceedings of the 5th Web as Corpus Workshop (WAC5)* (27-35. or.). Donostia: Elhuyar Fundazioa.

- Ixa taldea & Elhuyar Fundazioa. (2007). Testu-corpusak: ezaugarriak, eraketa eta tresnak. In M. J. Arrieta Sagarna (arg.), *Hizkuntza, komunikazioaren eta teknologiaren garaiak* (15-28. or.). Gasteiz: HAEE/IVAP.
- Karlsson, F., Voutilainen, A., Heikkilä, J. & Anttila, A. (arg.). (1995). *Constraint Grammar: A language-independent system for parsing unrestricted text. Natural Language Processing (vol. 4)*. Berlin-New York: Mouton de Gruyter.
- Kirov Ch., Sylak-Glassman, J., Knowles, R., Cotterell, R. & Post, M. (2017). A rich morphological tagger for English: Exploring the cross-linguistic tradeoff between morphology and syntax. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics* (II, 112-117. or.). Valenzia, Espainia.
- Koskenniemi, K. (1983). *Two-level morphology: a general computational model for word-form recognition and production* (doktore-tesia). University of Helsinki, Helsinki.
- Manning, C. D. (2011). Part-of-speech tagging from 97% to 100%: Is it time for some linguistics? In A. F. Gelbukh (arg.), *Proceedings of the 12th International Conference on Computational Linguistics and Intelligent Text Processing* (I, 171-189. or.). Berlin: Springer-Verlag.
- McKinlay, A. & Baldwin, T. (2005). Tagging with a more informative tagset. In T. Baldwin, J. Curran & M. Van Zaanen (arg.), *Proceedings of the Australasian Language Technology Workshop 2005* (40-48. or.). Sydney: University of Sydney.
- Odriozola, J. C. (d. g.). Adjektibo adberbialak. In *Sareko euskal gramatika (SEG)*. www.ehu.eus/seg helbidetik eskuratua.
- Otegi, A., Imaz, O., Díaz de Ilarraza, A., Iruskieta, M. & Uria, L. (2017). ANALHITZA: a tool to extract linguistic information from large corpora in Humanities research. *Procesamiento del Lenguaje Natural*, 58, 77-84.
- Tsarfaty, R., Seddah, D., Kübler, S. & Nivre, J. (2013). Parsing morphologically rich languages. *Journal of Computational Linguistics*, 39(1), 15-22.

Aurkibidea / Índice

- 11 Aitzinsolasa
- 13 Prólogo
- 15 Testu-corpusen informazio morfosintaktikoaren etiketatze automatikoa
hizkuntz ezagutzan oinarrituz: zenbait arazo, hainbat erronka
ITZIAR ADURIZ, JOSE MARI ARRIOLA
- 31 Bertsolaritzaren genealogia subalternoak
LUZIA ALBERRO, K. JOSU BIJUESCA
- 47 Euskal hiztun berri gazteak muda-prozesuan: ekintza-ikerketa baten
behinbehineko emaitzak
ESTIBALIZ AMORRORTU, ANE ORTEGA, JONE GOIRIGOLZARRI
- 63 Garaziko euskararen azterketa diafasikoa
ALEXANDER ARTZELUS MUXIKA
- 81 Berridazketak Sarrionandiaren poesian eta Saizarbitoriaren *Egunero hasten
delako* nobelan: hurbilpen genetiko bat
MIKEL AYERBE SUDUPE
- 95 Somorrostro, mendebaldeko euskararen muga
GOIO BAÑALES, MIKEL GORROTXATEGI
- 113 Euskara sasoian? Euskararen sozializazioa Gasteizko
pilota-elkarte batean
IÑIGO BEITIA
- 127 Euskal hiztun natiboak, ez-natiboak eta hitanoa
GARBIÑE BEREZIARTUA ETXEBERRIA, BEÑAT MUGURUZA ASEGINOLAZA
- 141 Lingüística Histórica: estado actual
LYLE RICHARD CAMPBELL
- 163 Lenguas y sociolingüística en el interior del País Vasco en el siglo XIX:
testimonios del Archivo Zavala
BRUNO CAMUS BERGARECHE, SARA GÓMEZ SEIBANE

- 177 Les verbes psychologiques du basque : typologie et diachronie
DENIS CREISSELS, CÉLINE MOUNOLE
- 195 Testu-generoaren irudikapena eta erabilera ahozkoaren
didaktikan
LEIRE DIAZ DE GEREÑU LASAGA, ITZIAR IDIAZABAL GORROTXATEGI, LUIS MARI
LARRINGAN ARANZABAL
- 209 Zentsura-ikasketak gaur egun: aplikazioa euskal literaturan
AMAIA ELIZALDE ESTENAGA
- 223 Azentuazioaren eremu sintaktikoak mugatzen
ARANTZAZU ELORDIETA
- 237 Sintaktikoki etiketatutako euskarazko corpus historikoa
eraikitzen
AINARA ESTARRONA, IZASKUN ETXEBERRIA, RICARDO ETXEPARE, MANUEL
PADILLA-MOYANO, ANDER SORALUZE
- 253 Ahozkotasanaren didaktika ikuspegi dialektikotik abiatuta
AINHOA EZEIZA, JAVIER ENCINA
- 267 Ahozko euskararen erabilera eskolan: gako zenbait irakasleen prestakuntzarako
INES M. GARCIA-AZKOAGA, OLATZ BENGOETXEA, JOSUNE ZABALA
- 283 El corónimo navarro *Salazar / Zaraitzu*: origen y desarrollo de su doble
denominación
ROBERTO GONZÁLEZ DE VIÑASPRE
- 295 *Fontes Linguae Vasconum*: orígenes y documentos para una Historia
del Euskara
JOAQUÍN GORROCHATAGUI
- 315 Euskal literatura itzuliaren historiografia bateraturantz
MIREN IBARLUZEA SANTISTEBAN
- 329 Basque among the world's languages: a typological approach
IVÁN IGARTUA
- 351 Familias vascohablantes: propuesta de definición desde la socialización
lingüística
PAULA KASARES
- 363 Formation linguistique de basque aux enseignants, pour un enseignement
bilingue à parité horaire au Pays Basque Nord
BEÑAT LASCANO
- 375 Eñaut Etxamendiren obra narratiboaren ekarpena euskal poetika erruralari
ITZIAR MADINA
- 391 Erdaretarako literatur itzulpena: zeharkako eta zuzeneko itzulpenaren arteko
muga lausoa
ELIZABETE MANTEROLA AGIRREZABALAGA

- 405 Ahozko euskararen irakaskuntzarako irakasleen prestakuntza: berrikuntza didaktikoa eta soziala?
IBON MANTEROLA
- 421 Enkarterriko PI-(h)aran/(h)uri motako euskal toponimoak
MIKEL MARTÍNEZ ARETA
- 437 Diachronical hypotheses accounting for synchronic variation: the case of the Basque particle *ote*
SERGIO MONFORTE
- 453 Jardueraren azterketa irakasleak prestatzeko bide: debatearen ikas-irakaskuntzaren adibidea
AROA MURCIANO EIZAGUIRRE, ARANTZA OZAETA ELORTZA
- 467 Hausnarketa zenbait euskal literatura-ikerketez
MARI JOSE OLAZIREGI
- 485 Ahozko euskara *Kolegioko ikastresna* ikasmaterialean
ARGIA OLÇOMENDY
- 501 Euskararen postposizioak
JAVIER ORMAZABAL
- 517 Hitz-ordenaren eragina zenbait ezaugarri gramatikalen erabilera-maiztasunean
LUIS PASTOR
- 533 Differential D-marking on proper names? A cross-linguistic study
IKER SALABERRI
- 547 Externalization and morphosyntactic parameters in Basque
HISAO TOKIZAKI
- 561 XIX. mendeko Debagoieneko testuez zenbait argitasun: egiletasuna eta iturriak
OXEL URIBE-ETXEBARRIA
- 579 Euskarazko perpausik gabeko azpikonparazioak
LAURA VELA-PLO
- 595 Latinaren aurreko osagai indoeuroparra Euskal Herriko toponimian: bukaeran -(iz)amo duten leku-izenak
LUIS MARI ZALDUA

Izenburua/Título:

Fontes Linguae Vasconum 50 urte. Ekarpen berriak euskararen ikerketari/Nuevas aportaciones al estudio de la lengua vasca

© Argitaratzaileak/Editores:

Ekaitz Santazilia, Dorota Krajewska, Eneko Zuloaga, Borja Ariztimuño

© Egileak/Autores:

Itziar Aduriz, Jose Mari Arriola, Luzia Alberro, K. Josu Bijuesca, Estibaliz Amorrortu, Ane Ortega, Jone Goirigolzarri, Alexander Artzelus Muxika, Mikel Ayerbe Sudupe, Goio Bañales, Mikel Gorrotxategi, Iñigo Beitia, Garbiñe Bereziartua Etxeberria, Beñat Muguruza Aseginolaza, Lyle Richard Campbell, Bruno Camus Bergareche, Sara Gómez Seibane, Denis Creissels, Céline Mounole, Leire Diaz de Gereñu Lasaga, Itziar Idiazabal Gorrotxategi, Luis Mari Larringan Aranzabal, Amaia Elizalde Estenaga, Arantzazu Elordieta, Ainara Estarrona, Izaskun Etxeberria, Ricardo Etxepare, Manuel Padilla-Moyano, Ander Soraluze, Ainhoa Ezeiza, Javier Encina, Ines M. Garcia-Azkoaga, Olatz Bengoetxea, Josune Zabala, Roberto González de Viñaspre, Joaquín Gorrochategui, Miren Ibarluzea Santisteban, Iván Igartua, Paula Kasares, Beñat Lascano, Itziar Madina, Elizabete Manterola Agirrezabalaga, Ibon Manterola, Mikel Martínez Areta, Sergio Monforte, Aroa Murciano Eizaguirre, Arantza Ozaeta Elortza, Mari Jose Olaziregi, Argia Olçomendy, Javier Ormazabal, Luis Pastor, Iker Salaberri, Hisao Tokizaki, Oxel Uribe-Etxebarria, Laura Vela-Plo, Luis Mari Zaldua

© Argitaratzailea/Edita:

Nafarroako Gobernua/Gobierno de Navarra

Kultura eta Kirol Departamentua/Departamento de Cultura y Deporte

Vianako Printzea Erakundea-Kultura Zuzendaritza Nagusia/Dirección General de Cultura-Institución Príncipe de Viana

Lanak adituek berrikusi dituzte, itsu bikoitzeko sistemaren bidez/Los trabajos han sido revisados por pares doble ciego.

Diseinua eta maketazioa/Diseño y maquetación:

Kö estudio

Imprimaketa/Impresión:

Linegrafic

ISBN: 978-84-235-3561-3

LG/DL: NA 1438-2020

Sustapena eta banaketa/Promoción y distribución:

Nafarroako Gobernuaren Argitalpen Funtsa/Fondo de Publicaciones del Gobierno de Navarra

Navas de Tolosa, 21

31002 Iruña/Pamplona

Tel.: 848 427 121

fondo.publicaciones@navarra.es

<https://publicaciones.navarra.es>

Fontes Linguae Vasconum 50 urte.

2019an 50 urte egin zituen Nafarroako Gobernuaren Vianako Printzea Erakundeak argitaratzen duen *Fontes Linguae Vasconum: studia et documenta* euskal hizkuntzalaritzako aldizkariak.

Horren gorazarre, liburu honek gaur egungo euskal hizkuntzalaritza- eta literatura-ikerketan zertan den erakutsi nahi du. Eskarmentu handiko ikertzaileek eta belaunaldi berriek bat egin dute argitalpen honetan, besteak beste, dialektologia, hizkuntzaren didaktika, filologia, gramatika teorikoa, hizkuntz tipologia, hizkuntzalaritza historikoa, itzulpengintza, literatura, onomastika eta soziolinguistika hizpide dituztela.

La revista de lingüística vasca *Fontes Linguae Vasconum: studia et documenta*, publicada por la Institución Príncipe de Viana del Gobierno de Navarra, cumplió 50 años en 2019.

En homenaje de la efemérides, este libro pretende dar cuenta del estado actual de la investigación en lingüística y literatura vascas. Investigadores de gran trayectoria y nuevas generaciones se reúnen en esta publicación para tratar, entre otros temas, sobre dialectología, didáctica de la lengua, filología, gramática teórica, tipología lingüística, lingüística histórica, traducción, literatura, onomástica y sociolingüística.

ISBN: 978-84-235-3561-3



9 788423 535613