



ERIDENEN DU ZERZAZ KONTENTA:
SAILKIDEEN OMENALDIA
HENRIKE KNÖRR IRAKASLEARI
(1947-2008)

MARIA-JOSÉ EZEIZABARRENA & RICARDO GÓMEZ (ARG.)

eman ta zabal oazu



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

ARGITALPEN
ZERBITZUA
SERVICIO EDITORIAL

ERIDENEN DU ZERZAZ KONTENTA:
SAILKIDEEN OMENALDIA
HENRIKE KNÖRR IRAKASLEARI
(1947-2008)

Maria-José Ezeizabarrena & Ricardo Gómez (arg.)



Universidad del País Vasco Euskal Herriko Unibertsitatea

ARGITALPEN
ZERBITZUA
SERVICIO EDITORIAL

2015
Bilbo

CIP. Biblioteca Universitaria

Eridenen du zerzaz kontenta : sailkideen omenaldia Henrike Knörr irakasleari (1947-2008) / Maria-José Ezeizabarrena & Ricardo Gómez (arg.) - Bilbo : Universidad del País Vasco / Euskal Herriko Unibertsitatea, Argitalpen Zerbitzua = Servicio Editorial, 2015. – 668 p. ; 24 cm.

Textos en euskara, español e inglés.

D.L.: BI-92-2015. – ISBN: 978-84-9082-092-6

1. Knörr, Henrike (1947-2008). 2. Euskara (Lengua). 3. Literatura vasca. I. Ezeizabarrena Segurola, María José, coed. lit. II. Gómez, Ricardo, coed. lit.

809.169

891.69.09

Argitalpen hau ondoko ikerketa proiektu eta taldeen barruan sartzen da:

Esta publicación se enmarca en los siguientes proyectos y grupos de investigación:

This publication has been developed in the framework of the following research projects and groups:

- «Aproximación al estudio de la competencia léxica y gramatical de los bilingües: metodologías experimentales y no experimentales» (Espainiako MINECO, FFI2012-37884-C03-02).
- «*Monumenta Linguae Vasconum* (IV): textos arcaicos vascos y euskera antiguo» (Espainiako MINECO, FFI2012-37696).
- «Historia de la lengua vasca y lingüística histórico-comparada» (HLMV-LHC) (Eusko Jaurlaritza, GIC. IT-698-13).
- «Hizkuntzaren jabekuntza, irakaskuntza eta erabilerak. Euskara, elebitasuna eta eleaniztasuna» (HIJE) (Eusko Jaurlaritza, GIC. IT-676-13).
- «Adquisición de lenguas, literatura y traducción en la formación plurilingüe» (UPV/EHU, UFI11/06).
- «Hizkuntzalaritza Teorikoa eta Diakronikoa: Gramatika Unibertsala, Hizkuntza Indoeuroparrak eta Euskara» (HiTeDi) (UPV/EHU, UFI11/14).

UPV/EHUko Euskararen arloko Errektoreordetzak lagundutako argitalpena

© Euskal Herriko Unibertsitateko Argitalpen Zerbitzua
Servicio Editorial de la Universidad del País Vasco

© Maria-José Ezeizabarrena & Ricardo Gómez (arg. / eds.)

© Egileak / Los autores / The authors

ISBN: 978-84-9082-092-6

Lege gordailua / Depósito legal: BI-92-2015

Aurkibidea

Aurkezpena	
MARIA-JOSÉ EZEIZABARRENA & RICARDO GÓMEZ	XI
Hitzaurrea	
IÑAKI GOIRIZELAIA, UPV/EHUko errektorea	XV
Henrike Knörr eta UPV/EHUko Letren Fakultatea	
IÑAKI BAZÁN, Letren Fakultateko dekanoa.	XVII
Henrike Knörr (1947-2008). Bibliografia	
GARIKOITZ KNÖRR (Universitat de València)	XXI
Diskurtso-markatzaile berri batzuk: <i>beste hitz batzuetan (esanda/esateko)</i> , <i>beste hitzetan (esanda/esateko)</i>	
XABIER ALBERDI LARIZGOITIA (UPV/EHU)	1
Eguna egunkariko idazleak (1937)	
XABIER ALTZIBAR ARETXABALETIA (UPV/EHU)	27
Lazarragaren eskuizkribuko adizki batzuez (euskal aditz jokoaren kronologia erlatibo baterantz)	
BORJA ARIZTIMUÑO LÓPEZ (UPV/EHU)	49
Euskararen azentuera: hizkera tradizioaletatik euskara estandarera	
GOTZON AURREKOETXEA, IÑAKI GAMINDE, LEIRE GANDARIAS & AITOR IGLESIAS (UPV/EHU)	71
<i>Work in conflict</i> : (meta)literatura y escritura en novelas que abordan el conflicto vasco	
MIKEL AYERBE SUDUPE (UPV/EHU)	97
<i>Hain zuzen (ere)</i> diskurtso-markatzailea: lokailu ala operadore?	
MIREN AZKARATE (UPV/EHU)	113

Erreferentziakidetasun-sareen etiketaze-metodologia EPEC Corpusen tratamendu konputazionalari begira KLARA CEBERIO (UPV/EHU), ITZIAR ADURIZ (Universitat de Barcelona), ARANTZA DÍAZ DE ILARRAZA (UPV/EHU) & INES M. GARCIA-AZKOAGA (UPV/EHU)	129
Zubereraren /y/: bere sorrera eta hedapenaren inguruan ANDER EGURTZEGI (UPV/EHU)	149
New media – new challenges for language policy: Rethinking the role of public media in language revitalization AGURTZANE ELORDUI (UPV/EHU)	163
EPEC-RolSem: Ingelesezko <i>PropBank-VerbNet</i> eredura etiketatutako euskarazko corpusa. Erabakiak, egokitzapenak eta bereizgarriak AINARA ESTARRONA, IZASKUN ALDEZABAL, ARANTZA DÍAZ DE ILARRAZA & MARÍA JESÚS ARANZABE URRUZOLA (UPV/EHU) .	179
<i>Kode alternantzia</i> elebidun txikien hizkuntzak bereizteko gaitasunaren adierazgarri MARIA-JOSÉ EZEIZABARRENA (UPV/EHU)	207
Narratze-trebetasunen garapena. Ipuin bera eta ume bera hiru adin ezberdinetan INES M. GARCIA-AZKOAGA, ITZIAR IDIAZABAL & LUIS M. LARRINGAN (UPV/EHU)	227
Birformulatzailerik zuzentzaileak testu idatzietan JULIO GARCIA GARCIA DE LOS SALMONES & MIKEL IRUSKIETA (UPV/EHU)	245
Partizipioa erlatiboa zenekoa: partizipio kategoria XIX. mendeko zenbait euskalariren lanetan RICARDO GÓMEZ LÓPEZ (UPV/EHU)	267
Sobre una estampilla ibérica de Ruscino (Castell Rosselló, Perpignan) JOAQUÍN GORROCHATAGUI (UPV/EHU)	281
Galderen eta erlatibozkoen ulermena haur euskaldunengan M. JUNCAL GUTIERREZ-MANGADO (UPV/EHU)	289
Euskarazko galdera erdibituak: analisi bitar baterantz ARITZ IRURTZUN (CNRS-IKER & HiTT)	307
Euskarazko silaba egiturak zen.ba.tzen OROITZ JAUREGI (UPV/EHU)	325

Aitzol Aramaioren <i>Terminal</i> laburmetraia. Literatura eta zinema JON KORTAZAR (UPV/EHU)	337
Hiru hasperen haboro JOSEBA A. LAKARRA (UPV/EHU)	349
Zaraitzuko oikonimia JUAN KARLOS LOPEZ-MUGARTZA IRIARTE (NUP)	379
Ehulku eta Jagonet kontsulta-zerbitzuak aztergai JESUS MARI MAKAZAGA EIZAGIRRE (UPV/EHU)	401
Gaztelania H ₁ eta H ₂ -ren garapena 5 eta 8 urte artean: ipuinetako aditz kohesioa aztergai IBON MANTEROLA GARATE (UPV/EHU)	419
Irudikapen fonologikoa gaztelania-euskara afasia elebidunean AMAIA MUNARRIZ (UPV/EHU)	437
<i>Gauza ziurra</i> eta <i>gezur hutsa</i> : euskal izenondo adberbialen kolokazioak JUAN CARLOS ODRIOZOLA PEREIRA (UPV/EHU)	457
Sites of memory in the Narrative of Bernardo Atxaga MARI JOSE OLAZIREGI (UPV/EHU)	475
Tipologia eta hizkuntzaren erritmoa: Stampe eta Doneganen ideien inguruan MIREN LOURDES OÑEDERRA (UPV/EHU)	491
<i>Awkward birds</i> . Cosmopolitanism and globalisation issues in mi- norised cultural communities LOURDES OTAEGI IMAZ (UPV/EHU)	499
Oloroeko diozesako euskal liburuak eta zuberera idatzia MANUEL PADILLA MOYANO (UPV/EHU)	519
Koherentzia diskurtsiboa euskal baladen exordioetan PATXI SALABERRI MUÑOA (UPV/EHU)	537
Hizkuntza corpusak eta euskara PELLO SALABURU (UPV/EHU)	559
Landucciren hiztegien ordena eta honen eragina euskarazko hiztegian BLANCA URGELL (UPV/EHU)	577
Henrike Knörr y su relación con Turquía CARMEN URIARTE (Turkiaren ohorezko kontsula Euskal Herrian) . .	603

Zenbait ohar <i>bata bestea</i> anafora elkarkariaren inguruan ARITZ URRUTIA, BEÑAT GOITIA & XABIER ARTIAGOITIA (UPV/ EHU)	617
Arabako euskararen gaineko ikerketak KOLDO ZUAZO (UPV/EHU)	643
Tabula Gratulatoria	667

Erreferentziakidetasun-sareen etiketatze-metodologia EPEC Corpusean tratamendu konputazionalari begira

Klara Ceberio^a

klara.ceberio@ehu.eus

Itziar Aduriz^b

itziar.aduriz@ub.edu

Arantza Díaz de Ilarraza^a

a.diazdeillaraza@ehu.eus

Ines M. Garcia-Azkoaga^c

ines.garciaazkoaga@ehu.eus

^a Hizkuntza eta Sistema Informatikoak Saila, UPV/EHU

^b Departament de Lingüística General, Universitat de Barcelona

^c Euskal Hizkuntza eta Komunikazioa Saila, UPV/EHU

1. Sarrera

Definizio onartuenak dio erreferentziakidetasun-erlazioak pertsona, objektu edo ekintza berari erreferentzia egiten dioten egitura linguistikoen artean ematen direla (Hirschman *et al.* 1997). Erlazio horiek argitzea diskurtsoaren ulermenaren zati garrantzizkoa da, izan ere, hizkuntzaren erabiltzaileei aukera ematen die elementu edo entitate berberarekin lotuta dauden diskurtsoko beste elementu batzuk identifikatzeko, eta horrenbestez, jatorrizkoaren gaineko informazioari jarraitutasuna emateko.

Horiek horrela, hizkuntzaren prozesamendu automatikoak aurrera eramaten dituen atazen artean erreferentziakidetasunaren ebazpenak berebiziko tokia eta garrantzia izan du. Are gehiago, gai hau landuta edukitzea ezinbestekotzat jotzen da hainbat proiektutan, tresna askoren aurrerapena edo arrakasta baldintzatzen duelako, testuen ulermenaren giltzatako bat baita. Izan ere, testuaren esanahia testu hori osatzen duten hitzen artean harilkatzen denez, derrigorrezkoa da, alde batetik, hitz bakoitzaren esanahia zein den garbi izatea, eta bestetik, esanahi horien artean dauden erlazioak azalaraztea.

Goi mailako ataza da, inondik ere, hizkuntzaren prozesamendu automatikoko beste mailetako lanak ebazita daudela suposatzen duelako, hala nola, morfosintaktikoa, sintaktikoa, (azalekoa gutxienez) eta hitzen adiera semantikoak desanbiguatuta, bi prozesuek elkar elikatu baitezakete neurri handi batean.

Euskararen prozesamendu automatikoaren bidean ere egin beharreko lana zela ikusi zen, eta badira urte batzuk IXA taldeko¹ estrategiaren barruan sartu zela. Horrela, ikerketa honen hastapenak eta tarteko egoera Aduriz *et al.* (2005) eta Ceberio *et al.* (2008) artikuluetan azaldu ziren.

Urte batzuetako eskarmentuaren ondoren, ikerketaren azken emaitzak azaltzea da artikulua honen helburua. Horrela, lehen puntuari erreferentziakidetasunaren inguruko kontzeptuak argituko ditugu, beti ere gure helburuari begira lagun dezaketen teoriak aztertuz. Jarraian erreferentziakidetasuna automatikoki ebazteko beharrezkoa den informazio linguistikoaren inguruan arituko gara, eta helburu horretarako egin dugun anotazioaren nondik norakoak azalduko ditugu. Ondoren, corpora etiketatzean automatikoki egindako pausoak azalduko ditugu eta bukatzeko lan honetan atera ditugun ondorioak.

2. Aztergaiaren mugak

Gauzak horrela, lehendabizi aztergaiaren mugei eskainiko dizkiegu lerro batzuk. Irakurleak testu bat irakurtzen duenean, testua gramatikalki zuzena izan arren, gerta liteke ulergaitza izatea. Testua ondo ulertuko bada, testuan agertzen diren gaiak eta azpigaiak modu egokian harilkatu behar dira, hau da, testua osatzerakoan kohesio-mekanismo egokiak erabili beharko dira.

Testua bere osotasunean kontuan hartzen dugula, kohesioa deritzogunak gaien garapena eta azpigaien artikulazioa ziurtatzen du. Lan honetan *gaia* diogunean diskurtsoan agertzen diren entitateei edo izateei buruz ari gara. Ahozko zein idatzizko ekoizpen batean, hainbat entitate eta diskurtsoko objektu agertzen dira, eta horrenbestez, hainbat adierazpide erreferentzial, batzuetan elkarren arteko harreman edota loturak erakusten dituztenak. Testuaren kohesioa ziurtatzen duten erlazio eta lotura horietako batzuk dira guk aztertuko ditugunak. Zehatzago esanda, aztertuko dugun fenomenoak *erreferentziakidetasuna* izango da eta horrekin hertsiki lotua dagoen *anafora*. Adibide baten bitartez definitzen saiatuko gara:

- (1) [*Alfredo Salazar Baskoniako orduko idazkari teknikoak*] lan handia egin zuen. Batik bat gauza bat deskubritu zuen [*hark*]: Argentinako harrobia. Hain zuzen, orain dela hamar urte egin zuen Argentinarako lehen bidaia [*Salazarrek*] [...] Gaur egun txapelketa sendoak dira hangoak. Jokalari gazte ugari parte hartzen duela da ezberdintasun nabariena Andres Nocioni da horren adibide argienetakoa: hama-

¹ <http://ixa.si.ehu.eus>

bost urte zituenerako egin zuen debuta Argentinako txapelketan; orain hogei urte ditu”. [Salazarrek] normalizat du Nocionik hain gazte egitea debuta, lehen esan bezala Argentinako saskibaloia ez delako hain serioa ere, eta, [Baskoniako idazkari tekniko ohiaren] ustez, zaila da horrelakorik gertatzea ACBn, lehia handiko liga delako.

Oro har, esan dezakegu erreferentzia dela hizkuntza-zeinuak edo adierazpide batek errealitate batera bidaltzeko duen propietatea, eta errealitate horri *erreferente* esango diogu; gainera, pertsona-erreferentziaz gain, espazio-erreferentziaz eta denbora-erreferentziaz ere hitz egin daiteke. Horretaz gain, erreferentziakidetasunari buruz hitz egin behar da, hau da, Moirand-en (1990: 19) definizioari jarraituta: bi elementuren artean erreferentziakidetasun harremana egongo da elementu horiek testuaren unibertsoaren erreferente berbera irudikatzen dutenean. Horrela, goiko adibidean letra etzanez agertzen diren adierazpide erreferentzial guztiek pertsona berari egiten diote erreferentzia, *Alfredo Salazar* pertsonari, hain zuzen. Haien artean erreferentziakidetasuna dago, eta ondorioz, ondoko izen-kateek erreferentzia-kate bat osatzen dute: [Alfredo Salazar Baskoniako orduko idazkari teknikoak], [hark], [Salazarrek], [Salazarrek], [Baskoniako idazkari tekniko ohiaren].

Diskurtsoaren eta testuen azterketarekin batera anaforaren nozioa ere agertu da, hura baita kohesioa eragiten duen elementu garrantzitsuenetarikoa. Testuko elementuen arteko erreferentzia horiek interpretatzeko moduarekin lotuta daude: “elementu bat aurretik edo atzetik (katafora) aipatu den beste elementu batekin harremanetan jartzen duen fenomeno adierazten du” (Garcia-Azkoaga 2003: 77). Gainera, erlazio anaforikoa dagoen kasuetan adierazpide anaforikoak semantikoki mendeko dira aurrekariarekiko (Cornish 1999), eta ondo interpretatzeko aurrekariaren informazioa ezinbestekoa izango da. Lehengo testuan agertzen den *hark* izenordainak adibidez, ez du berez inongo esanahirik, aurrekarira jo beharko dugu *Alfredo Salazar* idazkari teknikoari buruz ari dela jakiteko, eta horregatik bi elementuen arteko erlazioa anaforikoa dela esaten da.

Aurrekari baten eta hura izendapen ezberdinarekin berreskuratzen duen adierazpide erreferentziakideren baten artean harreman anaforikoa egon daiteke, baina horrek ez du esan nahi anafora eta erreferentziakidetasuna beti parekoak direnik (Garcia-Azkoaga, prestatzen). Aurrekaria eta adierazpide anaforikoaren arteko harremana ezin dela beti simetrikoa izan onartu beharko dugu. Batzuetan erlazio anaforikoa osatzen duten bi terminoek erreferentzia berdina izango dute (*Alfredo Salazar Baskoniako orduko idazkari teknikoak* eta *hark*), baina benetako ekoizpenen azterketa enpirikoak

erakusten digunez ez da beti horrela izango. Hor daude, esate baterako, asoziazio bidezko anaforak (2) edo erreferente ebolutiboen kasuak (3) (García-Azkoaga 2003).

- (2) *Herri batera* iritsi ginen. *Eliza* tontor batean zegoen.
- (3) Har itzazu *lau sagar*. Zuritu eta zatitu. Eduki egosten ordu erdiz. Txiki-txiki egin. Hoztu ondoren, zerbitzatu *konpota hori* gaileta-txoekin.

Era berean, litekeena da adierazpideak erreferentziakide izatea baina haien artean harreman anaforikorik ez egotea. Hori gertatzen da, hain zuzen, izen berezien kasuan edota errepikapen hutsezko kasuetan (lehenengo adibideko *Salazarrek*), bigarren terminoa aurrekar ra jo gabe interpreta baitezakegu. Era independentean interpreta daitezkeen adierazpideak dira eta horrenbestez, adibide horretan autonomoak diren bi adierazpideren arteko erreferentziakidetasun erlazioa besterik ez genuke izango (Kleiber 1994: 22).

Kontzeptu hauek guztiak hizkuntzalaritza konputazionalaren ikuspegitik ere aztertu izan dira (Botley & McEnery 2000, Mitkov 2002, Stede 2011), eta arlo horretan ohikoa den moduan, fenomeno horiek corpus errealetan nola gauzatzen diren aztertzen da. Horretarako, corpusak eskuz etiketatzen dira erreferentzia mailan, eta horixe da lan honen helburuetako bat, euskarazko etiketatzean kontuan izan ditugun adierazpide erreferentzial izateko aukera duten hautagaiak zehaztea (3.1 puntua) eta horrez gain, erreferentziakideak diren adierazpide horiek osatzen dituzten erreferentzia-kateak azaltzea (3.2. puntua).

Era berean, Mitkovek (2002) esan bezala, corpusa etiketatua izatea ezinbestekoa da, erreferentziakidetasunaren eta anaforaren ebazpen automatikorako tresna sendo bat garatzeko prozesuaren lehen urrats moduan. Corpus horien bidez ikasiko du ordenagailuak bere funtzionamendua, hau da, erlazio horiek gidatzen dituzten erregelak eta hala, markatuta ez dauden beste corpus batzuetan aplikatu ahalko dira automatikoki. Euskarazko corpusa garatzeko orduan, erreferentziakidetasunaren inguruko teoriak aztertzear gain, arlo honetan beste hizkuntzetan egindako lanak aztertu genituen aurreko lan batzuetan (Aduriz *et al.* 2005, Ceberio *et al.* 2008).

3. Erreferentzia automatikoki ebazteko alderdi linguistikoak

Hizkuntza edozein delarik ere, testu bateko elementuak bi multzotan bana ditzakegu: lehen multzoan zerbait erreferentziatzen duten adierazpide erreferentzialak izango dira (*referring expressions*) eta bigarren multzoan ezer erreferentziatzen ez duten adierazpideak ditugu (*non-referring expressions*) (Saeed 2009).

Desberdintasun hori bi modutara azal daiteke. Batetik, badira espresio linguistiko batzuk zerbaiti erreferentzia egiteko inoiz erabiliko ez direnak, besteren artean maiztasuna adierazten duten adberbioak (*sarri*), aditzak (*nai*) edota partikulak (*ote*). Hitz horiek diskurtsoaren parte dira eta semantikoki esanahia badute ere, ez diote munduko entitate bati erreferentzia zuzenik egiten, horregatik erreferentzialak ez diren espresioak direla esan ohi da. Aldiz, hizlari batek (4) adibidean bezala *zaldi* hitza erabiltzen badu, denok dakigu animalia jakin bati egiten ari zaiola erreferentzia, horregatik orokorrean izenak entitate bat identifikatzeko erabiltzen diren adierazpideak direla esan dezakegu.

(4) Ikaragarri sufritu zuen *zaldia*k lasterketan.

Beste modu batera esanda, erreferentziagai diren testu-zatiek ezaugarri gramatikal jakin batzuk dituztela ikusiko dugu. Goian aipatu ditugun osagai erreferentziadunak markagai edo aipamen erreferentziagai izango dira (*zaldi*), hau da, munduko objektu edo objektu-multzo bati erreferentzia egiteko gaitasuna duten hautagaiak. Horretarako, erreferentziakidetasun-kate batean parte har dezaketen elementuak identifikatzeaz gain, haien ezaugarri morfosintaktikoak zehaztuko ditugu.

3.1. Markagaiak edo aipamenak: ezaugarri morfosintaktikoak

Erreferentziakidetasuna izen-kateen² mailan gertatzen da, batez ere. Hala, gure etiketate-lana egiteko izen-katea osatzen duten elementuen arteko erreferentziakidetasunezko erlazioan zentratzeko erabakia hartu dugu. Ondoren, guk aipamen erreferentziakidegaitzat zein elementu hartzen ditugun azalduko dugu eta elementu horien ezaugarri morfosintaktikoak zehaztu. Sailkapenaren oinarri gisa euskarazko gramatikaz gain (Euskaltzaindia 2002), beste hizkuntzetan egin diren anotazio-eskemak izan ditugu kontuan (Recasens 2010, Pradhan *et al.* 2007).

3.1.1. Izenordainak

Aipamen erreferentziakideen artean, pertsona-izenordainak dira testuetan sarrien aurkitzen direnak. Pertsoneri erreferentzia egiten diete eta eurek bakarrik osatzen dute izen-katea, beste elementuren baten laguntzarik gabe.

Euskaltzaindiaren gramatika (Euskaltzaindia 1985) aipatzen denaren arabera, pertsona-izenordain arruntak lehen pertsonakoak (*ni*, *gu*) eta biga-

² *Izen-kate* diogunean izen-sintagma arruntez gain, izena ardatz gisa hartzen dituzten hitz-segida luzeagoak ere kontuan hartzen ditugu, hala nola, mendeko perpausa duten izenak barne hartzen dituzten izen-kateak etab.

renekoak (*hi, zu, zuek*) baino ez dira, euren forma indartuez gain (*neu, geu, heu, zeu, zeuek*).

- (5) [*Gu*] ere gogor defenditzen saiatuko gara, hor erabakiko baita nor-gehiagoka.

Hirugarren pertsonari dagokionez, euskaraz erakusleak erabiltzen ditugu; beraz, testuetan honako forma hauek agertuko zaizkigu: singularrean, *hau/hori/hura* (nor kasurako) eta *hon-/hor-/har-* (gainontzeko kasuetarako); eta, pluralean, *hauek/horiek/haiek*. Forma indartuak ere kontuan hartzekoak dira: *bera/berori/berau*, singularrean, eta *berauek/beroriek/beraiek*, pluralean.

- (6) *Jacken* bila joko dute, nahiz eta [*hark*] nahiago duen elurretan jolasten segitu.

Izenordain elkarkariak ere multzo honetan sartuko ditugu: *elkar* eta *batak bestea*. Horiek ere, euren bakarrik sor dezakete izen-katea, eta beti egiten diote erreferentzia perpaus berean lehenago agertutako elementuren bati.

- (7) *Hiru automobilek* [*elkar*] jo zuten Gornizko Andra Mari auzoan.

Izenordain bihurkarien kasua ere antzekoa da, baina hauek nahitaez behar dute beste elementuren bat izen-katea osatzeko: *X-ren burua*.

- (8) Gizon hark ez zuen [*bere burua*] besterik maite.

3.1.2. *Posesiboak*

Posesiboak genitibo marka duten izenordainak eta determinatzaileak ditugu. *Izenordain posesiboen* portaerari erreparatzen badiogu, beste izenordainen-gandik pixka bat aldentzen dira. Pertsona-izenordainen eta elkarkarien antzera, euren bakarrik osa dezakete sintagma, baina erreferentzia bikoitza egiten dute: batetik objektuari edo elementuari egiten diote erreferentzia, eta bestetik, norena den adierazten dute.

- (9) *Escudero* euskal musika tradizionala eraberritu eta indartu zuen. [*Harenak*] dira, esate baterako, *Illeta, Pinceladas Vascas eta Eusko Salmoa obrak*.

Goiko adibidean *harenak* izenordain posesiboak, batetik, *Escudero* musikariaren jabetza adierazten du, eta bestetik, musikariaren obrei “*Illeta, Pinceladas Vascas eta Eusko Salmoa obrak*” egiten die erreferentzia.

Posesiboak determinatzaile ere izan daitezke, eta, oro har izen-kate osoak kontuan hartuko baditugu ere, posesiboak agertzen diren kasuetan

kate osoaz gain, posesiboa bera bakarrik ere etiketatuko dugu. Erakusle horrek testuan lehenago aipatu den zerbaiti erreferentzia egiten dio, hurrengo adibidean *bere* posesiboak egiten duen moduan.

- (10) *Epitieren* kasuan, [[*bere*] helburua] lortu dezakeela dirudi eta bali-teke denboraldia Lehen Mailan hastea.

3.1.3. *Izen-kateak*

Izen-kate diogunean, izena ardatz duen sintagmari buruz ari gara. Multzorokor honen barruan berezitasun batzuk dituzten izen-kateak azalduko ditugu puntuz puntu.

3.1.3.1. *Izen-kate arruntak*

Izen-kate arruntak hiru multzo nagusitan banatu ditugu: artikulua mugagabea dutenak (11), artikulua mugatua daramatenak (12) eta izen-katea determinatzaile erakuslearekin mugatzen dutenak (13).

- (11) Inguruko etxeetan [*zenbait* beira] hautsi arren, kristalek ez zuten inor zauritu.
- (12) Txileko Gorte Gorenak baimena eman dio Juan Guzman epaileari [Pinocheten erregimenak ustez Argentinan egindako *hilketak*] ikertzeko.
- (13) Egunen batean lortzen badugu alderdien arteko mahai bat osatzea, eta [alderdi *horiek*] abiapuntu bat behar izatea, [abiapuntu *hori*] izan daiteke guk lor dezakegun proposamena.

Multzoak egiteko mugatasunean oinarritu gara, oso ohikoa baita elementu bat testuan lehen aldiz aipatzen denean mugagabeen erabiltzea. Mugatzailea edo erakuslea, aldiz, dagoeneko testuan aipatu den elementu bati buruz ari garenean erabiliko da, Zabalak (1996: 35) esan bezala “bai artikulua (edo deklinabide) mugatua bai erakusleak ere, *bat/batzuk* bezalako artikulua mugagabeen osagarriak dira erreferentzia-kateetan”.

3.1.3.2. *Izen bereziak dituzten izen-kateak*

Izen-katearen burua izen berezia denean (*Argentina*), orokorrean, ez dute determinatzailearik behar baina badaude kasu jakin batzuk determinatzailea onartuko dutenak (*Frantziako Gobernua*). Egiturari dagokionez ez dute eskuarki determinatzailearik behar, baina badaude kasu jakin batzuk determinatzailea onartuko dutenak (*Frantziako Gobernua*).

- (14) Epailea [*Argentinara*] joango da.

3.1.3.3. *Mendeko perpausa duten izen-kateak*

Mendeko perpausa duten izen-kateak ere aipamen erreferentziagaitzat hartuko ditugu. Izen-kateak buru bat izango du (15. adibidean *baimena*) eta burua modifikatzen duen mendeko perpausa izen-kate osoaren parte izango da.

- (15) [[[*DINAK*] [*Argentinan*] egindako krimenak] ikertzeko *baimena*] eman du Txileko Gorte Gorenak.

Izen-katearen burua *baimena* litzateke eta azpian duen perpausa, *DINAK Argentinan egindako krimenak ikertzeko*. Guk kate guztia kontsideratzeaz gain, aldi berean, azpian dituen izen-kateak aintzat hartuko ditugu:

- [*DINAK* Argentinan egindako krimenak ikertzeko *baimena*]
 [*DINAK* Argentinan egindako *krimenak*]
 [*DINAK*]
 [*Argentinan*]

Erlatibozko esaldiak ere izenaren mendeko perpaus gisa ager daitezke. Horrela, erlatibozko perpausa eta izena ere hartuko dugu aipamena markatzerakoan:

- (16) [[*Igandeko partiduak*] duen *garrantzia*] dela eta, lasai egotea beharrezkoa dutela esan zuen Lotinak.

3.1.3.4. *Izen-kate koordinatuak*

Izen-kate koordinatuak markagai izango dira, hau da, *eta*, *ala*, *edo*, *edota*, *nahiz* edo *zein* juntagailuen bidez osaturikoak. Izen-kate osoa ez ezik, barruko izen-kate bakunak ere kontuan hartuko ditugu.

- (17) Bazkal ondoren [[*siesta*] *eta* [*atsedena*]] besterik ez zuten egin.

Dena den, juntagailuz loturiko elementu bakunak ez dira beti aipamenak izango. Izan ere, batzuetan, koordinatutako izen-katea beste izen-kate baten modifikatzaile izan daiteke, eta, elementu bakoitza bere aldetik hartuz gero, informazioa galdu egingo genuke.

- (18) [*Txile, Argentina, Bolivia, Brasil, Paraguai eta Uruguai*ko diktadurarako *polizia sekretuek*] bat egin zuten ezkerreko militanteak atzeman eta hiltzeko.

Adibide horretan, herrialdeen izenak bakoitza bere aldetik hartuz gero, ez genuke esanahi osoa jasoko, berez esan nahi duena *Txileko/Argentinako/...-ko diktadurarako polizia sekretuak* baita, eta horregatik, aipamena guretzat osoa izango da.

3.1.3.5. *Postposizio-lokuzioetako markagaiak*

Euskaltzaindiak onartzen du ez dagoela *postposizioaren* definizio bateratu bat gramatikarien artean baina hona hemen *Euskal Gramatika. Lehen Urratsak I* liburuan ematen duen definizioa “perpauaren sintagmen arteko erlazio gramatikalak adierazten dituzten forma askeak direla esan daiteke” (Euskaltzaindia 1991: 297).

Egungo ikuspegitik, postposizioen artean bi mota nagusi bereizten dira: alde batetik, lehari erantsirik ageri diren atzizkiak: *-an* (*etxean*), Zabala eta Odriozolaren (2004) arabera atzizki postposizioak lirakeenak; eta bestetik, bi elementuz osatutako postposizio-lokuzioak: *-tik kanpo* (*etxetik kanpo*) (Zabala & Odriozola 2004, Aduriz *et al.* 2008). Guk postposizioaren lehenengo elementua bakarrik hartuko dugu aipamen erreferentziagaitzat, lehari erantsirik ageri diren atzizkiak barne (*etxetik*), postposizio-lokuzioaren elementu askeak (*kanpo*) ez baitu ezeri erreferentzia egiteko gaitasunik.

- (19) Zaragozaren beherakada ere horrelaxe iritsi da: [*etxetik*] *kanpo* eskas aritu arren, [*etxean*] partida guztiak irabazten hasi zen, azken boladan etxeko partidetan ere kale egiten hasi diren arte.

3.1.3.6. *Aditz-izenak*

Aditz-izenei buruz hitz egitean, “jokatu gabeko perpau baten aditzaz ari gara, aditz hori izen baten itxurarekin agertzen baldin bazaigu ere” (Euskaltzaindia 2002: 97). Batzuetan aditz eta izen horien arteko mugak lausoak dira, aditzen eta izenen ezaugarriak erakusten baitituzte aldi berean. Izaera bikoitza izaten dutela esan daiteke.

Horregatik, guk bi baldintza hauetakoren bat betetzen dutenean hartuko ditugu izentzat: aditz-izenaren subjektua edutezko genitiboan edo lekuzko genitiboa doanean (20), edo ondoren determinatzaile aske bat dutenean (21).

- (20) [EBren *zabaltzeak*] arazo asko konpontzera behartuko du.

- (21) Onik baizik ez dezake ekar [*zabaltze horrek*].

3.1.3.7. *Elipsia*

Badira zenbait izen-kate, esplizituki izenik agertu gabe ere, izenaren arrastoa dutenak, fenomeno horri maila morfosintaktikoan gertatzen den *elipsia* deritzogu. Agerian izenik ez duten arren, aipamen erreferentziagaitzat hartu ditugu hitz-segida horiek. Adibide bat erlatibozko perpauaren eta ondorengo kasu-marken artean gertatzen dena dugu. Aditzari zuzenean lotzen zaio kasu marka (22. adibidean, *zen-Ø-ak*), eta, izena isiltzen bada ere, sintagma osoak izen-kate gisa jokatzeko du. Hortaz, horrelako kasuak aipamentzat hartuko ditugu.

(22) [*Bigarren sailkatu zenak*] segundo bakarra kendu zion.

Bestalde, aditzarekin komunztadura egiten duten osagarriak isil daitezke euskaraz. Kasu horietan ere elipsia gertatzen dela esan ohi da, baina guk kasu hauek bazter utzi ditugu, lehenago aipatutako etiketatze morfosintaktikoan oraingoz kontuan hartu ez den elementua izan baita.

3.1.4. *Adberbioak*

Orain arte izen-kateez aritu bagara ere, zenbait kasutan, adizlagunek ere funtzio kohesiboa izaten dute, eta, corpuseko testuetan, ugariak dira funtzio hori duten leku-adberbioak. Euskaraz, benetako lekuzko adberbioak hiru dira (*hemen, hor, han*), eta horiei erakusle indartutik sortzen den adberbioa ere gehitzen zaie (*bertan*). Etiketatze-lanean, beraz, aipamentzat hartuko ditugu lekuzko adberbioak ere.

(23) Etxeko atezainarenean giltza bat utziko ziola, eta apartamentura igo eta [*han*] itxaroteko.

3.2. **Erreferentzia-sareak**

Artikuluaren hasieran aipatu dugu anafora eta erreferentziakidetasuna biak batera gertatzen diren fenomenoak izan daitezkeela. Batzuetan erreferentziakidetasunean oinarritutako anafora-erlazioak izango dira, baina badira kasuak, anafora-erlazorik agertzen ez dutenak. Horregatik, guk erreferentziakidetasuna etiketatzen dugula esango dugu.

Erreferentziagai diren aipamenen arteko lotura mota zehazterakoan, elementu bakoitzak bere aurrekoarekiko duen harremanari erreparatu zaio. Bada, kate batean entitate berberari erreferentzia egiten dioten hiru elementu badaude (*a, b* eta *c*), bi lotura mota zehaztu beharko ditugu: *b*-k *a*-rekiko duen erlazioa eta *c*-k *b*-rekiko duena (*ez, ordea, c* eta *a*-ren artekoa). Adibidea argigarri gertatuko zaigu:

[Markagaia1]_a ← [Markagaia2]_b ← [Markagaia3]_c

(24) [*Peter Gentzel atezain suediarra*]_a ez da, azkenean, Portland San Antoniora etorriko. Nafarrek egindako eskaintzari ezezkoa eman dio [*jokalariak*]_b, eta Granollers taldearekin jokatu du hurrengo denboraldian. [*Gentzelek*]_c atzo goizean hitz egin zuen Portlandeko entrenatzaile Zupo Ekisoainekin, Granollersek egindako eskaintza onartzeko arrazoiak azaltzeko.

[Peter Gentzel atezain suediarra]_a ← [jokalariak]_b ← [Gentzelek]_c

Adibide horretan ezarri beharreko erlazioak hurrengo hauek dira: [jokalariak]-ek aurretik lotu dugun [Peter Gentzel atezain suediarra]-rekiko duena eta [Gentzelek]-ek [jokalariak] izen katearekiko ezartzen duen lotura zein den.

3.2.1. Erreferentziakidetasun-erlazioak

Askotan esan ohi da korreferentziak etiketatzea ez dela ataza xumea (Recasens 2010). Urteetan zehar anotazio-eskema proposamen ugari egon dira eta ohikoena corpus bakoitzak bere beharretara egokitutako eskema erabiltzea izan da. Beraz, esan dezakegu ez dagoela estandartzat har daitekeen proposamenik, eta gainera, sarritan eskema horiek kontraesanean erori izan dira (van Deemter & Kibble 2000). Horregatik, puntu honetan gure sailkapenaren inguruko zehaztasunak emango ditugu. Hurrengo 1. taulak guk ezarriko ditugun lotura semantikoak laburtzen ditu.

1. taula
Etiketatzeko-sailkapena

Lotura-erlazioak			
Erreferentziakidetasuna	Pronominalak		
	Nominalak	Errepikapenezkoak	
		Errepikapen partzialak	
		Errepikapenik gabeak	Sinonimoak
			Hiponimoak
			Hiperonimoak
			Hiperonimo generikoak
			Alderantzizko instantziak
	Bestelakoak		
	Izen bereziak	Errepikapenezkoak	
		Errepikapen partzialak	
		Instantziak	
		Bestelakoak	
	Eliptikoak		
Lekuzko adberbioak			
Bestelako erlazioak	Predikatiboak		
	Aposizioak		

3.2.2. *Pronominalak*

Izenordainek aurreko izen-katearekin ezartzen duten erlazioari *pronominala* deritzo. Erlazio hori ondoko hitz-kategoriek osa dezakete: izenordain funtzioa betetzen ari den determinatzaile erakusleak (25), pertsona-izenordainak (26), izenordain bihurkariak eta posesiboak (27, 28) (genitiboan doazen izenordainak).

- (25) Aimar Olaizolak pazientzia asko dauka. [*Kontrarioa*]_i estu hartzen saiatzen da beti, [*hura*]_i nekarazten. Pilota lau eta erdiko arrastoan erabiltzen du askotan, eta tantoa amaitzen du aukera duen bakoitzean.

Corpus honetako testuak nagusiki egunkarietakoak dira eta batzuetan elkarrizketak tartekatuko dira. Testu zati horietan ohikoa da lehenengo eta bigarren pertsonako izenordainak agertzea. Pertsona-izenordain horiek, maiz balio deiktikoa duten arren, testu barruko elkarrizketetan agertzen dira eta testua bere osotasunean kontuan hartzen badugu, deiktikotasun hori galdu eta balore anaforikoa bereganatzen dute, izenordain horiek testuinguru linguistikoa kontuan harturik bakarrik interpreta ditzakegulako.

- (26) Egun hauetan Atlantako zilarraz galdetu diogun bakoitzean hura errepikatzea ia ezinezkoa dela erantzun du [*Abraham Olanok*]_i. [...] “Gustura egoteko moduko lana egin dut, baina laugarrengo postuak beti uzten dizu tristura bat, podiumaren atarian gelditzen baitzara. Baina, bai, [*nik*]_i uste dut nahikoa saio ona egin dudala”.

Markagaien atalean aipatu ditugun posesiboak ere, pronominaltzat hartuko ditugu, erakusle gisa jokatzeko dutenak batetik:

- (27) [*Carlos Prats*]_i Txileko Armadako burua eta presidenteordea izan zen Salvador Allenderen Gobernua indarrean zegoenean. 1974an DINAK Buenos Airesen hil zuen atentatuan, [*bere*]_i emaztearekin batera.

Eta bestetik izenordain funtzioa betetzen duten posesiboak ditugu:

- (28) Aretoa obretan dagoela eta, Bokaleko aretoan iraganen dira Scene Nationale erakundearen partaidetzarekin muntatutako hitzordu nagusiak. [*Orkestrak*]_i, adibidez, martxoan zehar egingen ditu [*be-reak*]_i han.

3.2.3. *Nominalak*

Korreferentzia *nominala* dela esaten dugu erreferentziakidetasunezko erlazioa izen-kate baten bitartez gauzatzen denean. Izen-kate hauek denetarikoa

ezaugarriak izan ditzakete eta hemen aurrekariekin ezartzen dituzten erlazioen arabera sailkatuko ditugu.

3.2.3.1. *Errepikapenezkoak*

Testuan zehar izen-katea errepikatzen denean, errepikapen bezala sailkatzen dugu, nahiz eta deklinabide-marka ezberdinak hartu, ez baita inongo informazio berririk gehitzen. Multzo honetan, bi azpimota bereizi ditugu: *errepikapen hutsa*, kate osoa errepikatzen denean (29); eta *errepikapen partziala* (30), bi izen-kateen gunea errepikatzen denean: errepikapena informazio urriagoarekin edo gehiagorekin ager daiteke.

- (29) Taldeetako zuzendariak eta [txirrundulariak]_i erabaki horren aurka agertu ziren, batik bat erabaki hori babesik gabeko gune batean hartu zelako, eta [txirrundulariek]_i hotz handia pasatu behar izan zutelako.
- (30) Clerc nagusi [esprintean]_i. [...] Eztabaidak alde batera utzita, etapa garaipena [azken esprintean]_i erabaki zen.

3.2.3.2. *Errepikapenik gabeak*

Bigarren izen-katean aurrekoaren elementurik errepikatzen ez denean, bi markagaien guneen arteko erlazio lexiko-semanticoko sakondu dugu euren arteko erlazioa zein izan daitekeen ebazteko. Testuan erreferentziakide diren bi izen-kate topatzen ditugunean euren artean maizen gertatzen diren erlazio motak ondorengoak ditugu:

Sinonimoak: Bigarren izen-katearen osagaia lehen kateko gunearen sinonimoa denean.

- (31) Peruko Gobernuak esan du OEAK ezin dituela indargabetu [hauteskundeak]_i [...] Peruri gerta dakioken okerrera [bozak]_i indargabetzea litzatekeela esan zuen De Trazegniesek.

Hiperonimoak: Kasu honetan ezartzen den erlazioa hiperonimo bidezkoa da, hau da, aurreko izen-katearen esanahia adiera zabalagoko hitz batek jasotzen badu, beti ere, erreferente berberari buruz ari garenean.

- (32) Errepide ugari moztu zituzten [manifestariak]_i, eta denda eta lantegi asko itxi [...] [Herritarrei]_i laguntza eskaini zien agintariengandik kolpatuak izanez gero.

Hiperonimo generikoa: Bigarren izen-kateko buruak izen generiko baten bidez (gai, kontu, gertaera...) eguneratzen denean.

- (33) Peruko Gobernuak esan du OEAK ezin dituela indargabetu [*Peruko hauteskundeak*]_i, [...] [*Herrialde barruko gaia*]_i, direla azpimarratu zuen De Trazegniesek.

Hiponimoa: Bigarren izen-kateko eta lehen kateko gunearen artean hiponimiazko erlazioa ezartzen denean.

- (34) Txileko Gorte Gorenak baimena eman dio Juan Guzman epaileari [*Pinocheten erregimenak*]_i, ustez Argentinan egindako hilketak ikertzeko [...] 1991n, ordea, Retting komisioa eratu zen, [*Pinocheten diktadurapean*]_i.

Alderantzizko instantzia: Aurreko elementua izen berezi bat denean agertzen da, gehienbat, eta bigarren kateak aurrekoaren ezaugarri bat ematen digu.

- (35) Baina ostegunean [*Males Zenawik*]_i, adierazi zuen amaitzear zegoela orain bi urte] hasi zen gatazka. Suetena eztabaidatzeko prest dago Addis Abebako Gobernua, [*Etiopiako lehen ministroak*]_i, esan zuenez.

3.2.4. *Izen bereziak*

Erreferentzia-sare bateko bi elementuren artean bigarrena izen berezia bada, lau motatako loturak zehaztu ditugu.

3.2.4.1. *Errepikapenezkoak*

Izen berezia osorik errepikatzen denean.

- (36) Etxera itzultzean tirokatu zuten [*Zugic*]_i, asteazken gaueko 23:15ean. [*Zugicek*]_i, autoa aparkatu ondoren...

3.2.4.2. *Errepikapen partzialekoak*

Errepikapen partzian izen bereziaren zati bat errepikatzen da. Bigarren elementua luzeagoa edo motzagoa izan daiteke.

- (37) [*Milo Djukanovicek*]_i, lehendakaritzarako hauteskundeak irabazi zituenean [...] Kosovoko krisian eta gerran, [*Djukanovicek*]_i, babesa eman zien Kosovotik ihes egin zuten albaniarrei.

3.2.4.3. *Instantzia bidezkoak*

Aurreko izen-kateak izen bereziari buruzko ezaugarriren bat ematen badigu.

- (38) [*Lehendakari*]_i hautatu zutenetik, [*Djukanovic*]_i aldaketa handia eman dio bere ildo politikoari.

3.2.5. *Eliptikoak*

Testuetan ohikoa da izena ezkutuan duten izen-kateak aurkitzea. Lotura eliptikoa dela esango dugu, erreferentzia-sareko bigarren elementuak izena ezkutuan badarama eta haren aurreko elementuan agerian dagoen izenari erreferentzia egiten badio.

- (39) [*Denboraldi honetan*]_i ordaindu du ACBra egokitu ahal izateko ordaindu beharrekoa. Ikusiko dugu [*datorrean*]_i zer eman diezaioken taldeari.

3.2.6. *Lekuzko adberbioak*

Baliteke kokagune bat izendatzeko izen-kate bat erabiltzea, eta, horren ondoren, lekuzko adberbio batez berriro egitea erreferentzia kokagune horri.

- (40) Etiopiak kontrolpean izango ditu iparraldeko frontean berreskuratu dituen lurraldeak eta [*Zalambessa hiria*]_i. Nazioarteak segurtasuna bermatu ondoren erretiratuko da [*handik*]_i.

3.3. **Beste erlazio batzuk**

Atal honetan erreferentziakidetasunezkoak ez diren erlazio batzuk kontuan hartu ditugu. Egitura predikatiboan eta aposizioen informazioa baliagarri gerta daiteke erreferentziakidetasunaren ebazpen automatikorako (Recasens 2010), izan ere, egitura horien bitartez, aipamen erreferentziagaia beste modu batera izendatzen da edota erreferentearen definizio zehatzagoa eman ohi da.

3.3.1. *Aposizioa*

Aposizioa izen-sintagma bat bestearen ondoren agertzen denean gertatzen da, eta orokorrean batak bestea azaldu edo zehazten du. Aposizio mota honi ez-murrizgarria deitu ohi zaio (Euskaltzaindia 1991: 203) eta gehienetan bi izen-sintagmak koma ikurraren bidez banatzen dira.

- (41) Zeregin horretan, [*Alfredo Salazarrek*]_i, [*Baskoniako orduko idazkari teknikoak*]_i lan handia egin zuen.

3.3.2. *Predikazioa*

Egitura predikatiboak oso maiz agertzen dira testuetan, egitura horiek subjektuaren ezaugarriak deskribatu ohi dituzte.

(42) [*Bulatovic*]_i [*Jugoslaviako Gobernu Federaleko lehen ministroa*]_i da.

4. Corpusaren prozesamendu automatikoa

Artikulu honetan aurkezten dugun etiketatze korreferentziala EPEC (*Euskarearen Prozesamendurako Erreferentzia Corpora*) izeneko corpusaren gainean egin da (Aduriz *et al.* 2006a). Oinarri gisa maila morfosintaktikoan etiketatutako corpora izango dugu. Lehenik *Morfeus* analizatzaile morfolo-
gikoa erabiliz (Aduriz *et al.* 1998) analizatutako hitzei informazio morfosin-
taktikoa esleituko zaie.

Aipatutako analisi morfosintaktikoa testuingurua kontuan hartu gabe egiten da eta hitz bakoitzak izan ditzakeen analisi posible guztiak ematen dizkigu. Anbiguotasun morfosintaktikoa murrizteko *Eustagger* izeneko analizatzailea (Ezeiza 2002) erabiltzen da eta testuinguru bakoitzean zuzena den aukera uzten du. Funtzio-etiketa sintaktikoa ere esleitzen dio fase honetan.

Hurrengo urratsean *Ixati* zatitzailea (Aduriz *et al.* 2006b) aplikatzen da. Modulu honek testua kateetan zatitzen du. Ezagutuko diren egitura sintaktiko oinarritzkoak hauek izango dira: *izen-entitateak*,³ *postposizio konplexuak*⁴ eta *kateak*⁵ (izen-kateak eta aditz-kateak). *ML-Ixati* tresnaren bidez, kateez gain, testuko perpausak identifikatzen dira (Arrieta 2010).

Maila morfolo-
gikoa zein sintaktikoa automatikoki analizatzen dira. Maila testualeko etiketatzea aldiz, eskuz burutu dugu.

4.1. Markagaien detekzio automatikoa

Aurreko puntuan zehaztutako irizpideetan oinarrituta, erreferentziagai diren aipamen kopuru bat eskuz etiketatu ondoren, markagaiak automatikoki etiketatzeko aplikazio bat garatu da (Soraluze *et al.* 2012).

Lehenagotik datorkigun izen-kateen informazioa erabiliz eta orain arte azaldu ditugun irizpideak kontuan hartuz automatikoki ezarriko zaio markagai bakoitzari dagokion etiketa. Eskuzko etiketatzailaren lehen lana markagaien kategoriak egiaztatzea eta dagokionean zuzentzea izango da. Hona hemen aipamenei automatikoki ezarriko zaien sailkapena jasotzen duen taula, etiketa bakoitzari buruzko azalpenekin eta adibideekin:

³ Entitateak: *Euskal Herria, Txomin*, etab.

⁴ Postposizio konplexuak: *-ren ondoren, 0 gisa...*

⁵ Katea: izenarekin eta aditzarekin sintaktikoki erlazionatutako hitzez osatua dago eta eskuarki sintagma kategoriako zatia da.

2. taula
Markagaien ezaugarri linguistikoak

Etiketa		Azalpena	Adibidea
Izenordainak	p_per	Izenordain pertsonalak	<i>ni, hi, gu, zu, zuek</i>
	p_ds	Izenordain gisa jokatzten duten erakusleak	<i>hau / hori / hura</i>
	p_enf	Izenordain gisa jokatzten duten erakusle indartuak	<i>bera, berau</i>
	p_rec	Izenordain elkarkariak	<i>elkar</i>
Posesiboak	defnp_pos	Izen-kate mugatuak, aurretik posesiboa dutenak	<i>honen / horren / haren etxea</i>
	defnp_pos_enf	Izen-kate mugatuak, posesibo indartuak	<i><u>bere</u> helburua</i>
	p_pos	Izenordain posesiboak	<i>harenak</i>
Izen-kate arruntak	defnp	Izen-kate mugatuak, artikuludunak	<i>hilket<u>a</u>k</i>
	indefnp	Izen-kate mugagabeak	<i>zenbait beira</i>
	defnp_dem	Izen-kate mugatuak, erakusledunak	<i>abiapuntu hori</i>
	ne_p	Pertsona izen bereziak	<i>Salazar</i>
	ne_l	Leku-izen bereziak	<i>Argentina</i>
	ne_o	Erakunde-izen bereziak	<i>Frantziako Gobernua</i>
	post_pre	Postposizio konposatuen aurretik doazen izen-kateak	<i><u>etxetik</u> (kanpo)</i>
Adberbioak	adv_l	Lekuzko adberbioak	<i>hemen / hor / han</i>

Etiketa guzti hauek automatikoki esleitu ondoren, testuan agertzen diren erreferentzia-sareak eskuz etiketatu dira aurreko 3.2 puntuan azaldu dugun moduan.

4.2. Erreferentzia-kateen ebazpen automatikoa

Erreferentzia-kateen ebazpen automatikora bidean hastapen gisa har dezakegu anafora pronominalen eta errepikapenezko izen-kateen ebazpen automatikoaren inguruan aurkeztu zen lana (Goenaga *et al.* 2012).

Lan honetan aurkeztu dugun etiketatzeak oinarri gisa balioko du, ebazpen automatikoa aipatu ditugun beste erlazio mota batzuetara zabaltzeko.

5. Ondorioak eta etorkizuneko lana

Duela urte batzuk hasitako ikerketari jarraipena emateko asmoz (Aduriz *et al.* 2007 eta Ceberio *et al.* 2008), erreferentziakidetasunaren inguruko erlazioetan sakondu eta horien inguruko azterketa egin dugu artikulu honetan. Honela, azken urteetan egindako ikerketa aztertuz eta hura oinarri gisa hartuz, erreferentziakidetasun-erlazioan oinarritutako sailkapen berria proposatu dugu EPEC corpusa maila erreferentzialean etiketatzeko helburuarekin.

Corpusa maila erreferentzialean etiketatzeko lehen pausoak automatizatzeke oinarriak ezarri ditugu. Soraluze *et al.* (2012) lanean aurkeztu genuen bezala, aipamen erreferentziagaiak automatikoki etiketatzeko tresna garatzeko balio izan duen corpusa osatu dugu.

Aipatutako markagaien etiketatze honek hurrengo pausoa aurrera eramateko, hau da, etiketatze erreferentziala egiteko bidea nabarmen erraztu digu, eskuzko etiketatzea azkarrago eta erosoago egin ahal izan baitugu. Horrez gain, aurkeztu dugun lan honek erreferentziakidetasuna (erdi-)automatikoki ebatziko duen tresna garatzeko oinarri sendoa izango da, izan ere, etiketatutako corpuseko informazioa erabiliko du aplikazioak, erlazio erreferentzialen portaera ikasi eta behar bezala ebatzi ahal izateko.

Lanaren oinarrian dagoen azterketa linguistikoak erreferentziakidetasunaren parte hartzen duten elementuen portaera hobeto ezagutzen eta sortzen diren sarea argitzen lagunduko digulakoan gaude.

Bukatzeko, begiak epe luzeagora jarriz, modulu hori bera euskarazko hainbat aplikazio informatikoetan txertatzeko aukera ikusten dugu, hala nola, laburpen sistemetan, galdera-erantzun sistema automatikoetan edota itzulpen automatikoko aplikazioetan.

Bibliografia

- ADURIZ, ITZIAR; AGIRRE, ENEKO; ALDEZABAL, IZASKUN; ALEGRIA, IÑAKI; ANSA, OLATZ; ARREGI, XABIER; ARRIOLA, JOSE MARI; ARTOLA, XABIER, DÍAZ DE ILARRAZA, ARANTZA; EZEIZA, NEREA; GOJENOLA, KOLDO; MARITXALAR, MONTSE; ORONOZ, MAITE; SARASOLA, KEPA; SOROA, AITOR; URIZAR, RUBEN & URKIA, MIRIAM, 1998, "A framework for the automatic processing of Basque Proceedings of Workshop on Lexical Resources for Minority Languages". In A. RUBIO *et al.* (arg.), *Proceedings of the First LREC Conference*. Granada.

- ADURIZ, ITZIAR; ALDEZABAL, IZASKUN; ARANZABE, MAXUX; ARRIOLA, JOSE MARI; CEBERIO, KLARA; ESTARRONA, AINARA; IRUSKIETA, MIKEL; LERSUNDI, MIKEL; POCIELLO, ELISABETE; URIA, LARRAITZ; URIZAR, RUBEN & ALDASORO, EDURNE, 2008, "Euskarazko postposizio-lokuzioen tratamendu konputazionala". Barne-txostena. UPV / EHU LSI / TR 07-2008.
- ADURIZ, ITZIAR; ARANZABE, MARIA JESUS; ARRIOLA, JOSE MARI; ATUTXA, AITZIBER; DÍAZ DE ILARRAZA ARANTZA; EZEIZA, NEREA; GOJENOLA, KOLDO; OROÑOZ, MAITE; SOROA, AITOR & URIZAR, RUBEN, 2006a, "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. Book series: Language and Computers*. Amsterdam/ New York: Rodopi, 56, 1-15.
- ADURIZ, ITZIAR; ARANZABE, MARIA JESUS; ARRIOLA, JOSE MARI & DÍAZ DE ILARRAZA, ARANTZA, 2006b, "Sintaxi Partziala". In B. FERNÁNDEZ & I. LAKA (arg.), *Andolin gogoan: Essays in Honour of Professor Eguzkitza*. Bilbo: UPV/EHU, 31-49.
- ADURIZ, ITZIAR; CEBERIO, KLARA & DÍAZ DE ILARRAZA, ARANTZA, 2005, "Euskarazko anafora pronominala: ikuspuntu konputazionala eta corpus baten garapena". *Gogoia* 5:1, 91-116.
- ADURIZ, ITZIAR; CEBERIO, KLARA & DÍAZ DE ILARRAZA, ARANTZA, 2007, "Nominal Anaphora in Basque: Annotation issues for later computational treatment". In Br. T. MCENERY, R. MITKOV & F. SILVA (arg.), *6th Discourse Anaphora and Anaphor Resolution Colloquium. DAARC 2007*. Porto: Centro de Linguística da Universidade do Porto.
- ARRIETA, BERTOL, 2010, *Azaleko sintaxiaren tratamendua ikasketa automatikoko tekniken bidez: euskarako kateen eta perpausen identifikazioa eta bere erabilera koma-zuzentzaile batean*. Donostia: UPV/EHUko doktore tesi argitaragabea.
- BOTLEY, SIMON & MCENERY, TONY (arg.), 2000, *Corpus-based and Computational Approaches to Discourse Anaphora*. Amsterdam: John Benjamins.
- CEBERIO, KLARA; ADURIZ, ITZIAR; DÍAZ DE ILARRAZA, ARANTZA & GARCIA-AZKOAGA, INES M., 2008, "Erreferentziakidetasunaren azterketa eta anotazioa euskarazko corpus batean". In X. ARTIAGOITIA, J. A. LAKARRA (arg.), *Gramatika Jaietan. Patxi Goenagaren omenez (ASJUren Gehigarriak, 51)*. Bilbo: UPV/EHU & Gipuzkoako Foru Aldundia, 153-172.
- CORNISH, FRANCIS, 1999, *Anaphora, Discourse and Understanding: Evidence from English and French*. Oxford: Clarendon.
- EUSKALTZAINDIA, 1991, *Euskal Gramatika: Lehen urratsak-I*. Bilbo: Euskaltzaindia.
- EUSKALTZAINDIA, 2002, *Euskal Gramatika Laburra: Perpaus Bakuna*. Bilbo: Euskaltzaindia.
- EZEIZA, NEREA, 2002, *Corpusak ustiatzeko tresna linguistikoak. Euskararen etiketatzaile sintaktiko sendo eta malgua*, Donostia: UPV/EHUko doktore tesi argitaragabea.

- GARCIA-AZKOAGA, INES-M., 2003, *Kohesio anaforikoa hiru testu generotan. Adinaren araberako azterketa*. Bilbo: UPV/EHU.
- GARCIA-AZKOAGA, INES-M., (prestatzen), “Erreferentziakidetasuna eta anafora: Diziplinartekotasunerako zedarriztatze erabilgarri baterantz”.
- GOENAGA, IAKES; ARREGI, OLATZ; CEBERIO, KLARA; DÍAZ DE ILARRAZA, ARANTZA & JIMENO, AMANE, 2012, “Automatic Coreference Annotation in Basque”. In I. HENDRICKX, S. KÜBLER & K. SIMOV (arg.), *TLT11 Proceedings of the Eleventh International Workshop on Treebanks and Linguistic Theories*. Lisboa: Edições Colibri, 115-126.
- HIRSCHMAN, L. & CHINCHOR, N., 1997, “MUC-7 coreference task definition”. In N. CHINCHOR (arg.), *MUC-7 Proceedings*, Science Applications International Corporation.
- KLEIBER, GEORGES, 1994, *Anaphores et pronoms*. Louvain-la-Neuve: Duculot.
- MILNER, JEAN-CLAUDE, 1982, *Ordres et raisons de la langue*. Paris: Seuil.
- MITKOV, RUSLAN, 2002, *Anaphora resolution*, London: Longman.
- MOIRAND, SOPHIE, 1990, *Une grammaire des textes et des dialogues*. Paris: Hachette.
- PRADHAN, SAMEER; HOVY, EDUARD; MARCUS, MITCHELL; PALMER, MARTHA; RAMSHAW, LANCE & WEISCHDEL, RALPH, 2007, “OntoNotes: A unified relational semantic representation”. In *Proceedings of ICSC 2007*, Irvin, California, 517-526.
- RECASENS, MARTA, 2010, *Coreference: Theory, Annotation, Resolution and Evaluation*, Barcelona: Universitat de Barcelonako doktore tesi argitaragabea.
- SAEED, JOHN I., 2009, *Semantics*. Malden, USA: Wiley-Blackwell.
- STEDE, MANFRED, 2011, *Discourse Processing*. San Rafael, California: Morgan & Claypool Publishers.
- SORALUZE, ANDER; ARREGI, OLATZ; ARREGI, XABIER; CEBERIO, KLARA & DÍAZ DE ILARRAZA, ARANTZA, 2012, “Mention Detection: First Steps in the Development of a Basque Coreference Resolution System”. In J. JANCSARY (arg.), *Proceedings of KONVENS 2012*. Vienna: 128-136.
- VAN DEEMTER, KEES & KIBBLE, RODGER, 2000, “On coreferring: Coreference in MUC and related annotation schemes”. *Computational Linguistics* 26:4, 629-637.
- ZABALA, IGONE & ODRIOZOLA, JUAN-CARLOS, 2004, “Los complejos posposicionales en vasco”. In E. PEREZ GAZTELU, I. ZABALA & L. GRÀCIA SOLE (arg.), *Las Fronteras de la Composición*. Donostia: Deustuko Unibertsitatea, 281-315.
- ZABALA, IGONE, 1996, “Testu-lotura: lotura tematikoa eta erreferentzia-sareak testu teknikoetan”. In I. Zabala (arg.), *Testu-loturarako baliabideak: Euskara Teknikoa*. Bilbo: UPV/EHU, 15-44.