

eman ta zabal zazu



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

EUSKAL FILOLOGIA SAILA

Euskarazko errorearen eta desbideratzearen analisirako lan-ingurunea

Determinatzaile-errorearen azterketa eta prozesamendua

Larraitx Uria Garinek

Doktore titulua eskuratzeko aurkezturiko

TESIA

Donostia, 2009ko iraila

eman ta zabal zazu



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

EUSKAL FILOLOGIA SAILA

Euskarazko errorearen eta desbideratzearen analisirako lan-ingurunea

Determinatzaile-errorearen azterketa eta prozesamendua

Larraitz Uria Garinek Igone Zabala
Unzaluren eta Montse Maritxalar
Angladaren zuzendaritzapean egindako
tesia, Euskal Herriko Unibertsitatean
Doktore titulua eskuratzeko aurkeztua.

Donostia, 2009ko iraila.

Aitari

eskerrak ematen

(1)

Eskerrik asko Montse ta Igone
lan honen zuzendariak,
zuen ondoan ikasi ditut
gauza interesgarriak
buruhaustekak, une gozoak,
eman dizkigu tesiak
lanerako ta bizitzarako
ikasgai paregabiak
nahita're ezingo nituen izan
nik zuzendari hobiak!!

(2)

IXA taldeko kide guztiei
ta nere azpitaldekoei
eskerrak eman nahi dizkizuet
lagundu didazuenei
“aguantatu” behar izan nauzuen
bulegokide kuttunei
omenajia ingo didaten
IXAgym-eko lagunei
prozesu hontan gertu sentitu
zaituztedan lankideei.

(3)

Maite ta Itziar zuen laguntza
izan da Excel-entia
erregelak, Bertol, zuri esker
lortut ebaluatzia
eskertzekoa EDBLko
kontuak a-Klara-tzia
zuzenketekin, formatuakin
batzuen parte hartzia (*)
zuen laguntzik gabe ez nuke
ez idatziko tesia.

(4)

Eskerrak ere IXAtik kanpo
animatu nauzuenei
gure lana ulertu ez arren
hor euki zaituztedanei:
Txeki, Iratxe, Tokaia, Kami,
Viki, Aran ta Julenei,
mendiria ihesi nerekin joaten
zeaten mendizaleei,
ta ahaztu gabe muxu handi bat
gure izeba Axunei.

(5)

Zure “animo ttikiya!!”k Idu
bai bitamina bikaiña
Ozenkirekin kantatutzia
terapia atsegiña
denen laguntza eta babesa
izan det behar adiña
nekez beharrean indar betez
in det hola ahalegiña,
ta azkenian, ta azkenian:
“eziña ekinenez egiña”.

(6)

Azkeneko bertsoa gorde det
aipatzeko familia
beti ondoan izan zaituztet
hori bai gauza haundia
Aiheko, Xela, Laia ta Xiba,
aita ta amatxo maitia
eskertzeizuet bihotz-bihotzez
zuen maitasun guztia!!!
NERE ESKER ONAK DENEI EMANAZ
hemen bukatzet tesia!!!!!!

(*) Maxux, Maite, Ruben, Izas, Nere, Mikel...

*Ez badakizu bereizten letra
xehea eta larria,
ez badakizu zerbaiten faltan
uzten duzun elbarria,
ez badakizu zetaz apaindu
behar zenukeen txerria...
hona zuretzat euskara batu
sinplearen aldarria.*

*Euskaldun askok txistukaririk
bereizten ez duen gisan,
gustiak ese egin ditsagun,
es atserakoiak isan:
“atso etorri ta etsi joan”
“apaia prest dago mesan”
Biskaian elkar ulertsen dute
ta sertan es hala esan?*

*Hegoaldean bada hatsea
esabatseko ustea,
es aoskatus sentsuskoa da
es jartsea obestea,
artara erres ulertuko da
esaldi onen fustea:
“sein gogorra den astelen goises
astea berris astea”.*

*Nola euskaldun berri gustien
traba ergatiboa den,
gu uste dugu ori euskaratik
kendu bear dela lenbailen:
“Ni i maite aut” esaten badut
se “k” bear da ba emen?
Nola demontre es dut jakingo
or sein sein maitatsen duen!*

*Iparraldean erri “egi” da
eta ala isan bedi;
Biskaian jota “iota” diote
eta es al da silegi?
A itsatsia kentsea ere
es da isango geiegi,
len ere maio entsun baitugu
“makina bat tontakeri”.*

*Kontraksioak aplikatuta
amaitu gurot jolasa:
naisuna motso esan al basu,
es iardun denborapasa!
Beingos kenduta gramatikaren
neugi gabeko desfasa
ea oin nogbait esaten dian
euskara estela errasa!*

Aurkibidea

I.	Sarrera eta aurkezpen orokorra.....	15
I.1.	Motibazioa	15
I.2.	Helburuak	18
I.3.	Metodologia	20
I.4.	Txostenaren antolakuntza	23
II.	Erroreen eta desbideratzeen analisia	25
II.1.	Sarrera.....	25
II.2.	Hizkuntzen i(ra)kaskuntza: ikerketa-lerro nagusiak	27
II.2.1.	Analisi Kontrastiboa.....	27
II.2.2.	Erroreen Analisia	30
II.2.3.	Tarte-hizkuntza	35
II.3.	Errore kontzeptua.....	36
II.3.1.	Euskararen egoera soziolinguistikoa.....	36
II.3.2.	Euskara batua eta euskalkiak	38
II.3.3.	Errore kontzeptua definitzeko hainbat ikuspuntu	43
II.3.4.	Gure ikuspuntua: erroreak eta desbideratzeak	47
II.4.	Analisi tradizioaletik ordenagailuz lagundutako analisira.....	50
II.5.	Laburbilpena	53
III.	Erroreak eta desbideratzeak aztertze lan-ingurunea	57
III.1.	Sarrera.....	57
III.2.	Euskarazko erroreak eta desbideratzeak aztertze lan-ingurunea	59
III.3.	Ordenagailuz lagundutako euskararen i(ra)kaskuntzarako aplikazioak	70
III.4.	Laburbilpena	75
IV.	Corpusgintza eta hizkuntza-ikasleen corpora.....	77
IV.1.	Sarrera.....	77
IV.2.	Corpusgintza: mugak eta abantailak	79
IV.3.	Corpusgintzaren metodologia orokorra	81
IV.4.	Corpus motak.....	84
IV.5.	Euskarazko erroreen eta desbideratzeen analisirako corpusak	89
IV.5.1.	Euskara-ikasleen testuak identifikatzeko kode-sistema	92
IV.5.2.	Baliatutako corpusen inguruko datuak.....	93
IV.6.	Laburbilpena	97
V.	Euskarazko erroreen eta desbideratzeen sailkapena.....	101
V.1.	Sarrera.....	101
V.2.	Beste hizkuntza batzuetan egin diren errore-sailkapenak.....	104
V.3.	Euskarazko erroreen eta desbideratzeen sailkapena	109
V.3.1.	Sailkapenaren egitura	109
V.3.2.	Sailkapeneko kategoria nagusien definizioa	112
V.3.3.	Sailkapena egituratzeko irizpideak	119
V.4.	Sailkapenaren ebaluazioa.....	120
V.5.	Laburbilpena	122

VI. Determinatzaile-erroreen eskuzko etiketatzea eta azterketa linguistikoa	125
VI.1. Sarrera.....	125
VI.2. Euskal determinatzaileen deskripzio linguistikoa	127
VI.2.1. Euskal determinatzaileen sailkapena	127
VI.2.2. Euskal determinatzaileen eta zenbatzaileen jokaera sintaktikoa	133
VI.3. Determinatzaile-erroreak etiketatzeko irizpideak.....	138
VI.4. Determinatzaile-erroreen sailkapena	143
VI.5. Determinatzaile-erroreen eskuzko etiketatzea: datuak	153
VI.6. Laburbilpena.....	162
VII. Determinatzaile-erroreen tratamendu automatikoa.....	165
VII.1. Sarrera.....	165
VII.2. Euskararako analizatzaile sintaktikoa.....	168
VII.2.1. Analisi-katea: urratsak	169
VII.3. Baliatutako formalismoa: Murriztapen Gramatika.....	177
VII.4. Determinatzaile-erroreen gramatika	179
VII.5. Detekzio automatikoan aurkitutako arazo batzuk	185
VII.5.1. Errore ortografikoak.....	185
VII.5.2. Hitz-segida ez-gramatikalak	187
VII.5.3. Aurreko urratsetan gertatutako akatsak	188
VII.5.4. HAULen tratamendua	189
VII.5.5. Hizkuntzaren konplexutasuna	191
VII.5.6. Hizkuntza-ekoizpen gramatikalak aztertzeko baliabideak erabiltzea	192
VII.6. Erregelen ebaluazioa	194
VII.6.1. Erroreen detekzioa euskara-ikasleen testuetan	196
VII.6.2. Alarma faltsuak testu zuzenetan	199
VII.7. Hitzak deklinatzean determinatzaile atzizkia ongi ez erabiltzeagatik sortutako erroreen azterketa.....	201
VII.8. Laburbilpena.....	204
VIII. Ondorioak.....	209
VIII.1. Ekarpen eta ondorio nagusiak	209
VIII.2. Tesiarekin lotura duten lanak	218
VIII.3. Etorkizuneko lanak	219

Eranskinak

A eranskina: Beste hizkuntza batzuetan garatu diren errore-sailkapenak	251
B eranskina: Euskarazko erroreen sailkapena	263
C eranskina: Determinatzaile-erroreen sailkapena	267
D eranskina: Euskara-ikasleen testuetan etiketatutako determinatzaile-erroreak	269
E eranskina: Determinatzaile-erroreak detektatzeko erregelak	287

Irudien aurkibidea

1. irudia.	<i>Erroreak</i> datu-basearen ingurunea.	21
2. irudia.	<i>Ikasleak</i> datu-basearen ingurunea.	22
3. irudia.	Erroreak eta desbideratzeak tratatzeko baliabideak.	48
4. irudia.	XUXEN-ek okertzat hartu arren, mendealdeko euskaran zuzenak diren hitzak. ...	49
5. irudia.	UCLEE errore-editorearekin etiketatutako adibidea.	60
6. irudia.	EtikErro editorearen ingurunea.	61
7. irudia.	EtikErro bidez datu-basera inportatzen den informazio linguistikoa.	62
8. irudia.	<i>Erroreak</i> datu-basean gordetzen den informazio teknikoak.	63
9. irudia.	<i>Erroreak</i> datu-basearen ingurunea.	64
10. irudia.	<i>Ikasleak</i> datu-basean gordetzen den informazio psikolinguistikoa.	65
11. irudia.	<i>Ikasleak</i> datu-basearen ingurunea.	65
12. irudia.	<i>Erroreak</i> eta <i>Ikasleak</i> datu-baseen lotura.	66
13. irudia.	Euskarazko erroreen eta desbideratzeen analisirako lan-ingurunea.	69
14. irudia.	HIKAS aplikazioaren interfazea.	73
15. irudia.	Idazlanen ebaluazio automatikorako sistemaren interfazea.	74
16. irudia.	Erroreen eta desbideratzeen sailkapenaren egitura hierarkikoa.	110
17. irudia.	Determinatzaile-erroreei dagokien egitura hierarkikoa.	111
18. irudia.	Determinatzaile-sintagmaren egitura, (Artiagoitia, 1998) lanetik hartua eta moldatua.	130
19. irudia.	Determinatzaileen eta zenbatzaileen kokagune sintaktikoa, Artiagoitia (2004) lanetik hartua eta moldatua.	130
20. irudia.	Determinatzaile-erroreen sailkapena.	144
21. irudia.	MORFEUS analizatzaile morfosintaktikoaren irteera.	171
22. irudia.	EUSTAGGER desanbiguatzaile morfosintaktikoaren irteera (3. mailan desanbiguatuta).	173
23. irudia.	IXATI zatitzailearen analisiak.	174
24. irudia.	Euskararako definitutako mendekotasun-erlazioak edo dependentziak.	175
25. irudia.	Analisi-katearen urratsak.	176
26. irudia.	Determinatzaile-erroreak detektatzeko idatzi diren erregelen aplikazioa.	181
27. irudia.	EZEZAG etiketa duen lexikorik gabeko hitzean erregela aplikatzea, aplikatu behar denean.	186
28. irudia.	EZEZAG etiketa duen lexikorik gabeko hitzean erregela aplikatzea, aplikatu behar ez denean.	186
29. irudia.	EZEZAG etiketa duen lexikorik gabeko hitzean erregela ez aplikatzea, aplikatu behar denean.	187
30. irudia.	Erregela aplikatzea, aplikatu behar denean.	187
31. irudia.	Hitz-segida ez-gramatikaletan sortutako alarma faltsuak.	188
32. irudia.	Desanbiguatzaileak utzitako analisisia egokia ez denean sortutako alarma faltsuak.	189
33. irudia.	Ongi detektatzen den determinatzaile-errorea.	190
34. irudia.	' <i>hitz egin</i> ' HAUL gisa analizatuta ez egoteagatik sortuko litzatekeen alarma faltsua.	190
35. irudia.	' <i>hitz egin</i> ' HAUL gisa analizatuta ez egoteagatik sortuko liratekeen alarma faltsuak.	191
36. irudia.	* <i>auzia hau</i> eta <i>auzi hau</i> sintagmen analisiak.	193
37. irudia.	Ongi detektatutako errore-adibideak.	198

Taulen aurkibidea

1. taula.	Euskara-ikasleen testu-fitxategiak identifikatzeko kodeen adibideak.	93
2. taula.	Orain artean bildutako euskara-ikasleen corpora.	94
3. taula.	Definitu izan diren hizkuntza-mailen arteko parekotasunak.	95
4. taula.	Euskara-ikasleen corpora, HEOKeko hizkuntza-maila bakoitzean.	95
5. taula.	Erroreen eta desbideratzeen analisirako baliatzen ditugun corpusen inguruko datuak.	97
6. taula.	Erroreen eta desbideratzeen sailkapeneko kategoriak eta horien tratamendurako tresnak.	118
7. taula.	Determinanteen sailkapena Goenagaren (1980) arabera.	128
8. taula.	Determinatzaileen sailkapena Euskaltzaindiaren (1993) arabera.	129
9. taula.	Euskarazko determinatzaileen sailkapena (Zubiri eta Zubiri, 1995).	129
10. taula.	Euskal determinatzaileak eta zenbatzaileak Lakaren (1996) arabera.	131
11. taula.	Euskal determinatzaileak eta zenbatzaileak Trask-en (2003) arabera.	132
12. taula.	Oinarritzat hartu dugun determinatzaileen sailkapena.	133
13. taula.	Etiketaturako errore motak eta bakoitzari dagokion errore-kodea.	153
14. taula.	Eskuz etiketatu den euskara-ikasleen corpora.	154
15. taula.	Hizkuntza-maila bakoitzean etiketatutako determinatzaile-erroreak, corpuseko IS kopuruarekiko.	155
16. taula.	Hizkuntza-maila bakoitzean etiketatutako determinatzaile-erroreak, guztira etiketatutako errore kopuruarekiko.	155
17. taula.	Etiketaturako determinatzaile-errore motak, guztira etiketatutako errore kopuruarekiko.	156
18. taula.	Hizkuntza-maila bakoitzean etiketatutako errore motak, maila bakoitzeko IS kopuruarekiko.	157
19. taula.	Hizkuntza-maila bakoitzean etiketatutako errore motak, maila bakoitzeko errore kopuruarekiko.	158
20. taula.	Euskaltegi bakoitzeko testuetan etiketatutako determinatzaile-erroreak, maila bakoitzeko IS kopuruarekiko.	158
21. taula.	Euskaltegi bakoitzeko testuetan etiketatutako determinatzaile-erroreak, guztira etiketatutako determinatzaile-errore kopuruarekiko.	159
22. taula.	Euskaltegi bakoitzeko testuetan etiketatutako errore motak, euskaltegi bakoitzean egindako determinatzaile-errore kopuruarekiko.	159
23. taula.	Euskara Teknikoko ikasleen testuetan etiketatutako determinatzaile-erroreak, corpuseko IS kopuruarekiko.	160
24. taula.	Euskara Teknikoko ikasleen testuetan etiketatutako determinatzaile-erroreak, guztira etiketatutako determinatzaile-errore kopuruarekiko.	161
25. taula.	Desanbiguazio-mailak.	172
26. taula.	MG formalismoa erabili duten zenbait gramatika-zuzentzaile.	178
27. taula.	Erregela-adibidea.	183
28. taula.	Landutako errore-kategoria bakoitzean dauden azpikategorien kopurua (VI.4. atala).	183
29. taula.	Errore mota bakoitzeko erregelen kopurua.	184
30. taula.	Garapenerako eta probarako baliatu dugun euskara-ikasleen corpora.	196
31. taula.	Euskara-ikasleen probarako corpusaren gainean egindako ebaluazioaren emaitzak.	196
32. taula.	Errore mota bakoitzeko erregelen kopurua.	197
33. taula.	Probarako corpusaren gainean egindako ebaluazioaren emaitzak, errore mota bakoitzeko.	197

34. taula.	Alarma faltsuen kopuruak.	199
35. taula.	Probarako corpusean egindako ebaluazioaren emaitzak, erregelei ez dagozkien alarma faltsuak kontuan hartu gabe.	199
36. taula.	Euskara-ikasleak ez diren hiztunen testuetan detektatutako adibideen kopuruak, sintagma-kopuruarekiko.	200
37. taula.	Euskara-ikasleak ez diren hiztunen testuetan detektatutako egiazko positiboen eta gezurrezko positiboen kopuruak, IS kopuruarekiko.	201
38. taula.	Errore mota batzuk detektatu eta diagnostikatzeko EDBLn sartutako deklinabide-atzizki okerren zerrenda.....	203

Tesi-lan honen helburu nagusia izan da euskarazko errorearen eta desbideratzearen azterketa eta prozesamendua bideratzeko lan-ingurune bat garatzea gure hizkuntza-komunitaterako ekarpenak egiteko bi ikerketa-alarretan: errorearen tratamendu automatikoan eta ordenagailuz lagundutako hizkuntzen i(r)a-kaskuntzan.

I. Sarrera eta aurkezpen orokorra

I.1. Motibazioa

Errorearen eta desbideratzearen azterketa eta prozesamendua da tesi-lan honi dagokion ikerlerroa. Euskarazko errorearen analisiaren eta tratamendu automatikoaren inguruan gogoeta bat egin nahi izan dugu: gure hizkuntza-komunitatearen egoera kontuan hartuta, Hizkuntzaren Azterketa eta Prozesamenduaren (HAP) alorrean garatu behar ditugun bitartekoak eta egin behar ditugun ekarpenak zer-nolakoak izan behar duten ikertzea izan da gure xedeetako bat. Euskarak bizi duen egoera soziolinguistikoa dela eta, eta ordenagailuak gaur egun gure gizarteko hainbat arlotan duen presentzia handia denez, alor horretan ekarpen garrantzitsuak egin daitezkeela aurreikusten dugu.

Gure hizkuntza-komunitatearen egoerari dagokionez, normalizazio-prozesuan dagoen hizkuntza gutxitua da euskara, ia erabat hiztun elebidunez osaturiko komunitatea da geurea eta bertan, gainera, hainbat dialekto hitz egiten dira. Euskal idazleek mendeetan errepikatutako kezka izan zen kode estandar baten beharra; izan ere, euskal komunitatea elkarren artean oso komunikazio ahula zuten komunitate txikietan banatuta zegoen, eta hizkuntzaren biziraupena ziurtatzeko ezinbestekoa zen batez ere komunikazio-egoera formal idatzietarako euskaldun guztiek partekatuko zuten kode bat garatzea. Horrela, 1968 urtean ekin zitzaion euskara batuaren sorrerari, eta berari esker lortu da, besteak beste, euskal hiztunen kopurua handitzea eta gure hizkuntza-komunitatea trinkotzea, hizkera desberdinetako hiztunen arteko

ulergarritasuna bermatzen baitu kode estandarrak. Euskara batuaren gauzatzeak ekarri du, halaber, euskal hizkuntzaren barruan dauden kode desberdinen banaketa funtzionala (Zuazo, 2000; Euskaltzaindia, 2004).

Hizkuntzen erabilerari erreparatzen badiogu, aldakortasuna da nabarmentzen den ezaugarria: hiztunen jatorriaren arabera (dialektoak, soziolektoak) eta zereginen edota komunikazio-egoeren arabera (estiloak, erregistroak) bereizi ohi dira. Gainera, aldaera mota horiek gurutzatu egiten dira erabileran: dialekto eta soziolekto desberdinetako hizkuntza-baliabideak funtzio jakin batzuetarako aukeratzen ditugu komunikazio-egoeraren arabera. Euskararen kasuan, komunikazio-egoera formaletan eta batez ere idatzietan erabiltzen den kodea euskara batua da: zuzentasunaren erreferentzia. Formaltasun maila baxuagoa behar duten testuinguruetan zuzentasunaren erreferentzia-ardatz horretatik urruntzen gara, baliabide dialektalak erabiliz edota euskara batuaren kodea erabili arren, estilo zaindurako ematen diren estilo-arauetatik urrunduz. Hizkuntzaren erabilera zuzena, beraz, kode estandar idatziarekin identifikatu ohi da (Milroy eta Milroy, 1991) eta testuinguru formaletan ekoizten diren testuek (gutun ofizialak, txostenak...) kode estandarrari eta estilo zainduari dagozkion ezaugarriak bete behar izaten dituzte gizartean arrakastatsuak izateko. Egoera informaletan (lagunartean...), berriz, estilo zaindugabea edota kode dialektalak erabili ohi dira. Beraz, kode estandarra eta ez-estandarra (dialektala), estilo zaindua eta zaindugabea, erregistro formalak eta informalak, guztiak dira beharrezkoak hizkuntza-komunitate batean; guztiak izan daitezke egokiak komunikazio-egoeraren edota eremu geografikoaren arabera. Eta hori kontuan hartzekoa da errorearen azterketa eta prozesamenduaren alorrean, erroretzat zer hartuko den eta zer ez finkatzeko unean.

Bestalde, komunikazio-bide berriak etengabe garatzen ari dira, egun, informazioaren eta komunikazioaren gizartea esaten zaion honetan, eta ordenagailuaren presentzia ezinbestekoa da, jada, gure gizarteko hainbat arlotan. Idatziz komunikatzen gara maiz, formaltasun maila desberdinetako testuinguruetan (posta elektronikoa, foroak, *chatak*...). Askotan ordenagailu bidez idazten dugunez, laguntza-tresna informatikoak funtsezkoak dira bai hizkuntza-ikasleentzat, bai hiztun arruntentzat. Hizkuntza-ikasleei euskara batua edota testuinguru formaletan erabiltzen diren baliabideak ez ezik, testuinguru informaletan erabili ohi diren hizkuntza-baliabideak ere irakatsi behar zaizkie, euskaraz egoera desberdinetan

komunikatzeko gaitasuna garatuko badute. Euskara estandar formala soilik irakastea hizkuntzaren eredu mugatuegia ematea litzateke. Hizkuntzen i(ra)kaskuntzaren alorrean, beraz, gramatika menderatzeko eta hizkuntza aberasteko baliagarriak diren tresnak eskaini behar dira: baliabide estandarrak hizkuntzaren gramatikaltasuna ikasteko izango zaizkio baliagarriak ikasleari; eta kodeei (estandarra/dialektala) eta erregistroei (formala/informala) dagozkien baliabide funtzionalak hizkuntzaren aberastasuna eskuratzeko balioko dute. Hiztun arruntei, aldiz, hizkuntzaren eredu estandarra eskaintzen dizkieten tresna (erdi)automatikoak interesatuko zaizkie gehienbat (ortografia- eta gramatika-zuzentzailea¹, esaterako).

Testuinguru honetan, erroreen tratamendu automatikoak ekarpen garrantzitsua egin dezake laguntza-tresna informatikoak garatzeari begira. Hizkuntza i(ra)kasteko, lantzeko edota hobetzeko prestatzen ditugun tresnek akats mekanikoak (mekanografia-akatsak kasu), akats ortografikoak edota gramatika-akatsak markatzeaz gain, gai izan behar lukete estilo-oharrak eskaintzeko edota kode estandarretik urruntzen ari garela ohartarazteko. Izan ere, hizkuntzaren erabilerari dagokion aldakortasuna kontuan hartuz (estiloei, dialektoei eta erregistroei dagozkien baliabideak bereiziz), jakin ahalko genuke noiz erabiltzen ari garen euskaraz inolako kodetan existitzen ez den baliabide bat eta noiz euskaraz existitzen den baliabide bat.

Horrek lotura du, halaber, errore kontzeptuaren definizioarekin. Erroreen analisiaren alorrean erroretzat zer hartu behar den eta zer ez definitu beharra dago, eta horretarako ezinbestekoa da ikusmiran dugun hizkuntza-eredua kontuan izatea. Gisbert-ek (1998) dioen moduan, egoera batean erroreduna denak beste egoera batean ez du zertan errore izan. Eta erroreen azterketa eta prozesamenduaren alorrean islatu behar den ideia da hau: kode dialektalei edota estilo zaindugabeari dagozkion formak erroretzat hartu behar dira erregistro formalak eskatzen dituen komunikazio-egoeretan, baina ez erregistro informalak eskatzen dituenetan; eta, alderantziz, kode estandarri eta estilo zainduari dagozkion baliabideak erabiltzea ez da egokia erregistro informalak eskatzen dituzten kasuetan, baina ezinbestekoa da

¹ Gramatika-zuzentzailea terminoa erabiltzen dugun arren, gramatika-zuzentzaileek, oro har, ez dituzte erroreak zuzentzen, detektatu baino ez dituzte egiten, eta batzuetan zuzenketak proposatu.

komunikazio-egoera formalean. Horren aurrean, erroreak eta desbideratzeak bereizten ditugu guk: errore esaten diegu gramatika-arauak betetzen ez dituzten egitura okerrei; eta desbideratzetzat hartzen ditugu gramatikalak izan arren, testuinguru edota egoera jakin batean egokiak ez diren formak. Euskararen azterketa eta prozesamenduaren alorrean, beraz, errorearen eta desbideratzearen tratamendu automatikoa bideratzeko tresnak (kode estandarrekoak diren baliabideak eta estandarrak ez diren baliabideak bereizten dituztenak) garatzeko premia aurreikusten dugu, gure hizkuntza-komunitatearen beharrak asetzeko ekarpen bat egingo badugu.

I.2. Helburuak

Tesi-lan honen xede nagusia da euskarazko errorearen eta desbideratzearen analisirako ezinbestekoak diren baliabideekin lan-ingurune bat osatzea, errorearen tratamendu automatikoaren alorrean eta ordenagailuz lagundutako hizkuntzen i(r)a(k)askuntzarenean ekarpenak egiten jarraitzeko. Ondoko baliabide hauek osatzen dute lan-ingurunea: corpusek (hizkuntza-ikasleen edota hiztun arrunten testuek), adibide erroredunak etiketatzeko editoreak, errorearen eta desbideratzearen sailkapenak eta etiketatutako adibideen inguruko informazioa gordetzeko datu-baseek. Halaber, errorearen azterketa eta prozesamendua bideratzeko hainbat irizpide finkatu behar izan dira garatzen ditugun tresnak edota bitartekoak gure hizkuntza-komunitaterako baliagarriak izan daitezen.

Tesi hau Hizkuntzaren Azterketa eta Prozesamendua (HAP) esaten zaion ikerketa-arloan EHUko IXA taldeak² egiten duen lanaren barruan kokatzen da; zehazkiago, euskarazko errorearen tratamendu automatikoa eta ordenagailuz lagundutako hizkuntzen i(r)a(k)askuntza bideratzea helburu duen azpitaldean.

IXA taldea 1987 urtean sortu zen euskararen azterketa eta prozesamenduaren alorrean ikertzeko. Taldearen azken helburua euskararako hain beharrezkoak eta onuragarriak izan daitezkeen sistema (erdi)automatikoak egitea edo egiten laguntzea da. Euskararen azterketa eta tratamendu automatikoa bideratzeko lanetan, gure

² <http://ixa.si.ehu.es/Ixa>

hizkuntza prozesatzeko tresna eta bitarteko sendoak sortu ditugu, hizkuntza-corpus idatzien analisi morfologiko, sintaktiko, semantiko eta pragmatikoetatik abiatuta. Erroreen azterketa eta prozesamenduaren alorrean, IXA taldean aspaldi bideratu zen hitz-mailako erroreen detekzioa (Agirre *et al.*, 1992), Euskararen Datu Base Lexikalean (EDBL) (Aldezabal *et al.*, 2001) biltegiatutako informazioan eta bi mailatako morfoloian (Koskenniemi, 1983) oinarrituta. Gramatika-errore batzuen detekzio automatikoa ere egiten da jada (Gojenola, 2000; Oronoz, 2009). Erroreen azterketa eta prozesamendua, ordea, lan konplexua da eta egiteko asko dago oraindik alor horretan.

Erroreen analisisian ondoko urrats hauek bereizten ditugu: *detekzioa*, *etiketatzeari*, *diagnosi linguistikoa*, *diagnosi psikolinguistikoa* eta *zuzenketa*. *Detekzioa* deritzo hitz, esaldi edota egitura jakin batean hizkuntza-arau bat hautsi dela aurkitzeari. *Etiketatzeari* detektatutako adibideak markatzean datza. *Diagnosi linguistikoa* detektatutako errorea edo desbideratzea linguistikoki deskribatzea da, zein motatakoa den zehaztea. *Diagnosi linguistikoa* egiteko, euskarazko erroreen edota desbideratzeen sailkapena garatu dugu. *Diagnosi psikolinguistikoa*, aldiz, errore-adibideak ikasleen edota hizkuntza-mailaren ezaugarriak kontuan hartuz aztertzean datza; hizkuntza-maila desberdinetako ikasleen erroreak, maila bereko ikasleek egindakoak, hainbat ikastetxetako ikasleenak, ikasle batek ikasketa prozesuan zehar egin dituenak... aztertzea eta alderatzea da. *Diagnosi linguistikoa* ezinbestekoa da bai erroreen tratamendu automatikoaren alorrean, bai hizkuntzen i(r)a-kaskuntzarenean ere; *diagnosi psikolinguistikoa*, berriz, hizkuntzen i(r)a-kaskuntzaren alorrari dagokio gehienbat. Azkenik, *zuzenketa* erroreari dago(z)kion baliokide zuzena(k) ematea da.

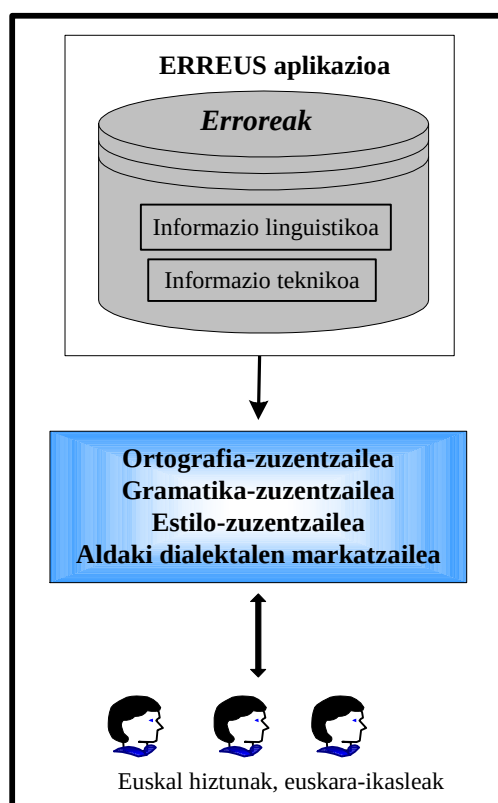
Tesi-lan honetan aurkezten dugun euskarazko erroreen eta desbideratzeen analisirako lan-ingurunea edozein errore edota desbideratze motaren detekzioa, etiketatzea, diagnosa (linguistikoa edota psikolinguistikoa) eta zuzenketa bideratzeko diseinatu da. Nolanahi ere, eskura ditugun bitartekoekin errore eta desbideratze guztien analisisa aldi berean egitea ezinezkoa zaigu eta, beraz, fenomeno linguistiko baten azterketa zehatza egin dugu tesi-lan honetan: determinatzaile-erroreak aztertu ditugu. Lehenik, determinatzaile-erroreak, testu errealean etiketatu ditugu eskuz; eta etiketatutako determinatzaile-errore batzuen detekzio automatikoa bideratu dugu ondoren.

Euskarazko erroreen eta desbideratzeen analisia bi ikerketa-alorretan egiten dugu: erroreen tratamendu automatikoaren alorrean, euskararako ortografia-, gramatika- eta estilo-zuzentzailea edota aldaki dialektalen markatzailea bezalako tresnak garatzea dugu helburu; eta ordenagailuz lagundutako hizkuntzen i(r)a kaskuntzarenean, berriz, hizkuntzaren i(r)a kaskuntza prozesua bideratzen laguntzeko baliabideak prestatzea. Alor batean zein bestean azterketak egin eta erabiltzaileen benetako beharrei erantzuteko laguntza-tresna eraginkorrak garatu ahal izateko, ezinbestekoa da euskaraz egiten diren erroreen eta desbideratzeen analisi osoa egitea, gure hizkuntza-komunitateak bizi duen egoera soziolinguistikoa kontuan hartuta.

I.3. Metodologia

Erroreen Analisia psikolinguistikaren alorrean sortu zen 70.eko hamarkadan. Ordenagailuaren erabilerak azken urteotan izan duen hedaduraren ondorioz, aukera berriak zabaldu dira eta egun, jada, beste hainbat arlotan egiten diren ikerketen oinarri bilakatu da. HAPen alorrean, esaterako, erroreen analisia ezinbestekoa da gramatika- eta estilo-zuzentzaileak garatzeko (Ornoz, 2009) eta ordenagailu bidez hizkuntzak i(r)a kasteko, hobetzeko edota lantzeko lagungarriak izan daitezkeen baliabide linguistikoak sortzeko (Maritxalar, 1999; Tono, 2003; Granger, 2002, 2003 eta 2004, besteak beste). Oro har, baina, erroreak bereiz aztertu ohi dira aipatu bi ikerketa-alorretan, batean zein bestean helburuak desberdinak direlako. Guk egiten dugun euskarazko erroreen eta desbideratzeen azterketak, ordea, badu berezitasun bat: analisi berbera baliatzen dugu erroreen tratamendu automatikoaren alorrean eta ordenagailuz lagundutako hizkuntzen i(r)a kaskuntzarenean; batean egiten dugun lanaren zati bat baliagarria zaigu besterako, eta alderantziz. Horrela, bi ikerlerroak uztartzen dituen lan-ingurune oso bat garatu dugu, Maritxalarren (1999) lana abiapuntutzat hartuta, euskarazko erroreen eta desbideratzeen analisia bideratzeko ezinbestekoak zaizkigun bitartekoekin.

Erroreen eta desbideratzeen tratamendu automatikoari dagokionez, euskararako ortografia-, gramatika- eta estilo-zuzentzailea edota aldaki dialektalen markatzailea bezalako tresnak garatzea da gure xedeetako bat. Horretarako, testuetan detektatutako adibideak *Erroreak*³ deritzon datu-basean biltegitratzen ditugu, adibide bakoitza automatikoki tratatzeko beharrezkoa den informazio linguistikoarekin eta teknikoarekin batera. Bertako informazioa eskuragarri dago ERREUS (Aldabe *et al.*, 2005a) izeneko web aplikazioaren bidez (1. irudia).

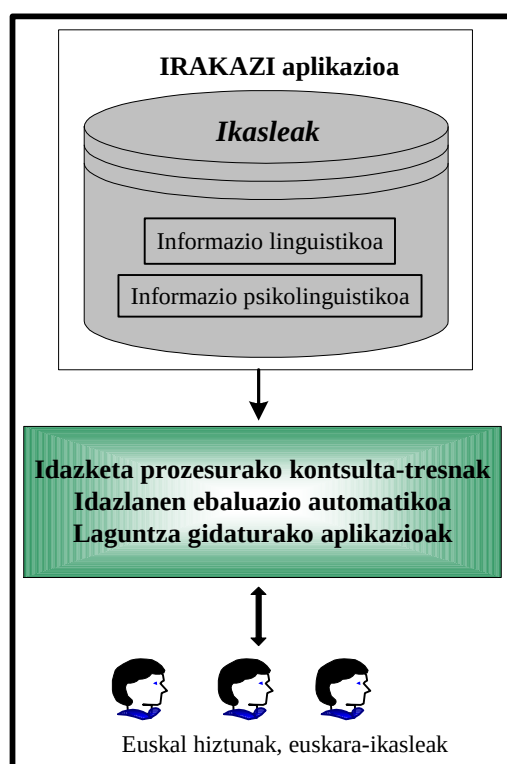


1. irudia. *Erroreak* datu-basearen ingurunea.

Ordenagailuz lagundutako hizkuntzen i(r)akaskuntzaren alorrean, berriz, gure helburuak dira hizkuntzaren ikasketa-prozesua ezagutzeko baliabideak prestatzea eta erabiltzaileen beharretara egokitutako laguntza-tresnak sortzea (idazketa prozesurako kontsulta-tresnak, idazlanen ebaluazio automatikorako sistemak, laguntza gidatua

³ Datu-basearen izena *Erroreak* den arren, erroreak eta desbideratzeak jasotzen ditugu bertan.

eskaintzen duten aplikazioak...). Horretarako, *Ikasleak* izeneko datu-basean, IRAKAZI web aplikazioaren bidez (Aldabe *et al.*, 2005b), hizkuntza-ikasleek egindako errorean eta desbideratzeen adibideak gordetzen ditugu, adibide bakoitzaren inguruko informazio linguistikoarekin eta psikolinguistikoekin batera (2. irudia).



2. irudia. *Ikasleak* datu-basearen ingurunea.

Bi datu-baseak oinarritzkoak dira euskaraz idatzitako testuetan egiten diren erroreak eta desbideratzeak biltzeko, aztertzeko eta tratatzeko. Izan ere, datu-base horietan biltegitratutako informazioa da aztergai ditugun ikerketa-alorretan ekarpenak egiteko eta laguntza-tresnak sortzeko abiapuntu nagusia. Datu-base horiekin batera, corpusak, adibideak etiketatzeko errore-editorea, euskarazko errorean eta desbideratzeen sailkapena edota HAP alorrean garatutako hainbat tresna dira aipatu lan-ingurunea osatzen duten gainerako bitartekoak.

I.4. Txostenaren antolakuntza

Tesi-txostena zortzi ataletan banatu dugu. Lehenengo kapitulua sarrera honi dagokio. Bertan laburbildu ditugu gure lanaren motibazioa, helburuak eta metodologia. Hemen esandakoak zehazten eta garatzen joango gara tesian barrena. Erroreen eta desbideratzeen analisia izango dugu hizpide II. kapitulan: alor horretan egin diren lanak aipatu eta euskarazko erroreen eta desbideratzeen azterketaren inguruan gure lanak eta ikuspuntuak zein diren azalduko dugu, errore kontzeptua definituz eta zehaztuz. Ondoren, III. kapitulan, erroreen eta desbideratzeen analisirako garatu dugun lan-ingurunea eta euskararen i(ra)kaskuntzaren alorrerako baliagarriak izango diren bi aplikazio aurkeztuko ditugu. Corpusgintzaz, oro har, eta gure azterketak egiteko baliatzen ditugun corpusez hitz egingo dugu IV. kapitulan. Euskarazko erroreen eta desbideratzeen sailkapenari dagokio V. kapitulua: beste hizkuntza batzuetan garatu diren errore-taxonomia batzuk aurkeztuko ditugu eta gure sailkapena azalduko dugu. VI. kapitulan determinatzaile-erroreen inguruan egin dugun eskuzko etiketatzea eta azterketa linguistikoa izango dira gai nagusiak eta VII. kapitulan, berriz, eskuz etiketatutako determinatzaile-errore mota batzuen detekzio automatikoa nola bideratu dugun kontatuko dugu. Azkenik, VIII. kapitulan, ekarpen eta ondorio nagusiak laburbildu eta aurreikusten ditugun etorkizuneko lanak zehaztuko ditugu, lan honi amaiera emateko.

Tesi-lan hau bukatzeko, kontsultatu ditugun erreferentziak eta hainbat eranskin gehituko ditugu: beste hizkuntza batzuetan garatu diren errore-sailkapenak (A eranskina); euskarazko erroreen eta desbideratzeen sailkapena osatzen duten kategoria eta azpikategoria guztiak (B eranskina); determinatzaile-erroreen sailkapena (C eranskina), euskara-ikasleek eta Euskara Teknikoa irakasgaiko ikasleek egindako determinatzaile-erroreen bilduma (D eranskina) eta determinatzaile-erroreak automatikoki detektatzeko idatzi ditugun erregelak (E eranskina).

Erroreen eta desbideratzeen analisia ezinbesteko ikerlerroa da Hizkuntzaren Azterketa eta Prozesamenduaren alorrean aurrerapausoak ematen jarraitu ahal izateko.

II. Erroreen eta desbideratzeen analisia

II.1. Sarrera

Erroreen Analisia Hizkuntzalaritza Aplikatuaren adar bat da, Hizkuntzalaritza Konputazionalaren eta Hizkuntzen I(ra)kaskuntzaren barruan lantzen den alorretako bat.

Hizkuntzalaritza Aplikatua deritzo hizkuntzari dagozkion arazo praktikoak diziplinarteko ikuspuntu batetik konpontzen saiatzen diren hizkuntzalaritzaren ikerketa-ildoek osatzen duten alorrari, hainbat jakintza-alorretako hatsarreak kontuan hartuz hizkuntzari lotutako arazoak konpontzeko motibazio aplikatua duten ikerketa-ildoei. *“La LA es defineix per tant, al nostre entendre, com una orientació de la teoria i la recerca en el conjunt de camps d’estudi integrats en les ciències del llenguatge que, partint d’un marc teòric interdisciplinari, té una finalitat pràctica (aplicada) consistent en la resolució de problemes i conflictes que sorgeixen en la praxi lingüística”* (Cabré, 1990). Hizkuntzalaritza Aplikatuak hainbat jakintza-alor barne hartzen ditu, besteak beste: hizkuntza-teknologiak edo hizkuntzalaritza teknikoa, psikolinguistika edo hizkuntzen i(ra)kaskuntza, itzulpengintza, soziolinguistika, etnolinguistika, terminologia... (Hoffman, 1998).

Tesi-lan honetan Hizkuntzalaritza Aplikatuaren bi alor tradizionalenak uztartzen ditugu: hizkuntzalaritza teknikoari dagokion Hizkuntzaren Azterketa eta Prozesamendua (HAP) eta psikolinguistikari dagokion ordenagailuz lagundutako hizkuntzen i(ra)kaskuntza. Biek dute, ordea, oinarri berbera: Erroreen Analisia. Eta

horixe da, hain zuzen, lan honen gai nagusia: euskarazko erroreen eta desbideratzeen analisia. Gure azterketen helburua edota motibazio aplikatua bikoitza da:

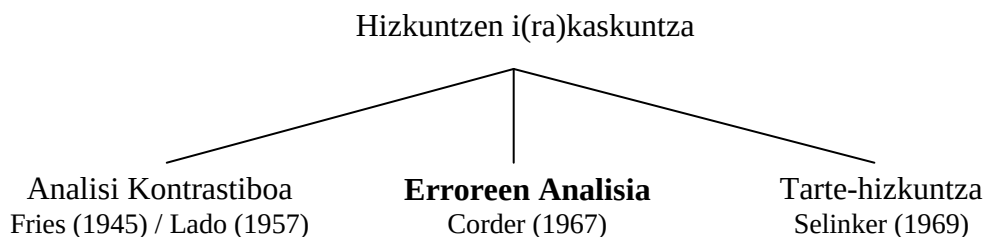
- erroreen tratamendu automatikoaren alorrean, euskararako ortografia-, gramatika- eta estilo-zuzentzailea edota aldaki dialektalen markatzailea bezalako tresnak garatzea;
- ordenagailuz lagundutako hizkuntzen i(ra)kaskuntzaren alorrean, hizkuntzaren ikasketa-prozesua bideratzen laguntzeko baliabideak prestatzea.

Xede nagusi horiek erroreen eta desbideratzeen analisi sakon bat eskatzen dute. Horretarako ezinbestekoak diren irizpideak finkatu eta lan-ingurune oso bat garatu dugu guk, tesi-lan honen ekarpen nagusi gisa.

Kapitulu honetan erroreen eta desbideratzeen analiziaz hitz egingo dugu. Lehenik eta behin, Erroreen Analisiak Psikolinguistikaren edo hizkuntzen i(ra)kaskuntzaren alorrean izan duen bilakaeraz mintzatuko gara II.2. atalean, Analisi Kontrastiboaren (II.2.1.), Errore Analisiaren (II.2.2.) eta Tarte-hizkuntzaren (II.2.3.) ikuspuntuak azalduz. Erroreen Analisiaren alorrean, ezinbestekoa da errore kontzeptua definitzea, errore zeri esaten diogun zehaztea. Horretaz mintzatuko gara, hain zuzen, II.3. atalean: euskarak bizi duen egoera soziolinguistikoa izango dugu hizpide (II.3.1.) eta zenbait gogoeta egingo ditugu euskara batuaren eta euskalkien arteko funtzio-banaketatik (II.3.2.), eragin zuzena baitu horrek gure lanerako finkatzen ditugun irizpideetan; errore kontzeptuaren inguruan hainbat adituren ikuspuntua zein den ikusi ostean (II.3.3.), guk finkatu ditugun irizpideak aipatuko ditugu II.3.4. atalean: erroreak eta desbideratzeak bereizten ditugu, eta ekarpen garrantzitsua iruditzen zaigu hori gure hizkuntza-komunitateko erabiltzaileentzat baliagarriak izango diren laguntza-tresnak sortu ahal izateko. Ondoren, II.4. atalean, erroreen analisi tradizionalaren aldean ordenagailuz lagundutako analisiak eskaintzen dizkigun abantailak azalduko ditugu, ordenagailuari esker aurrerapen handiak egin baitira bai erroreen tratamendu automatikoaren alorrean eta bai ordenagailuz lagundutako hizkuntzen i(ra)kaskuntzarenean ere. Bukatzeko, kapituluaren laburbilpena egingo dugu II.5. atalean.

II.2. Hizkuntzen i(ra)kaskuntza: ikerketa-lerro nagusiak

Erroreen Analisiak azken urteotan zenbait diziplina edota ikerketa-lerrotan aplikazio berriak izan dituen arren (HAPen barruan, esate baterako), hizkuntzen i(ra)kaskuntzaren alorrean du sorrera, non hiru ikuspuntu edo ikerlerro nagusi garatu diren: Analisi Kontrastiboa, Erroreen Analisia eta Tarte-hizkuntzaren teoria:



II.2.1. Analisi Kontrastiboa

Erroreen Analisiaren aurrekaria dugu Analisi Kontrastiboa. Fries (1945) eta Lado (1957) aitzindari izan zituen teoria honek 50. eta 60.eko hamarkadetan hartu zuen indarra. Bi hizkuntza-sistemaren arteko konparazioan oinarritzen da: ikaslearen ama-hizkuntzaren eta xede-hizkuntzaren arteko alderaketa sistematikoa egitea da Analisi Kontrastiboaren helburua. Bi hizkuntzen arteko interferentziak eta baliokidetasunak aztertu eta hizkuntza berriaren i(ra)kaskuntza azterketa horren baitan bideratu behar dela diote teoria honen jarraitzaileek, azterketa kontrastiboetatik ondorioztatu edota aurreikusi ahal izango baita hizkuntza jakin bat ikastea ikasleek aurki ditzaketen zailtasunak eta arazoak zein diren.

Skinner-en teoria behaviorista (konduktista) du oinarrian Analisi Kontrastiboak. Autore horren arabera (Skinner, 1985), hizkuntzak *Estimulu-Erantzun* sistema baten bidez ikasten ditugu; hau da, hizkuntza berri bat ikasteko, ama-hizkuntzako ohiturak (egiturak, hiztegia...) alde batera utzi eta xede-hizkuntzako ohiturak hartu behar izaten ditugu. Ondorioz, xede-hizkuntzan egiten diren errore gehienak ama-hizkuntzaren transferentziak eragiten dituela defendatzen dute Fries eta Lado-k, eta ondorioztatzen dute ama-hizkuntzan eta xede-hizkuntzan antzekoak diren fenomenoak erraz ikasten direla eta bi hizkuntzetan desberdinak diren egiturak ikastea, aldiz, zailagoa gertatzen zaigula (Fries, 1945; Lado, 1957).

Analisi Kontrastiboaren alorrean sortu ziren oinarritzko bi kontzeptu dira *interferentzia* eta *errorea*. Interferentzia esaten zaio xede-hizkuntzan ama-hizkuntzari dagozkion ezaugarri fonetiko, morfologiko, sintaktiko edota lexikoak erabiltzeari. Lado-k bi interferentzia edo transferentzia mota bereizten ditu: negatiboa (alderatzen diren egiturak bi hizkuntzetan desberdinak direnean) eta positiboa (bi hizkuntzetan egiturak antzekoak direnean). Errorea, berriz, xede-hizkuntzako arau-haustea da, ama-hizkuntzaren interferentziak eragin ohi duena. Hizkuntzak kanpo-estimulu, *feedback* edota errefortzu egokiari erantzunez ikasten badira eta erroreak kanpo-estimulu horiei emandako erantzun desegokiak badira, errore horiek berehala zuzentzea komeni da; zuzendu ezean, ohitura bihurtzen baitira xede-hizkuntzan. Bada, Analisi Kontrastiboaren ikuspuntutik erroreak onartezinak dira, lehen mementotik saihestu edo ekidin beharrekoak. Teoria honen arabera, lehen hizkuntzaren interferentziak eragiten ditu hizkuntza berri baten ikaskuntza-prozesuan sortzen diren eragozpenik handienak. Baina hizkuntza bat ikasteen sortzen diren errore, interferentzia edota zailtasun horiek aurreikus eta ekidin daitezke azterketa kontrastiboen bidez, ama-hizkuntzaren eta xede-hizkuntzaren arteko berdintasunak eta desberdintasunak alderatuz.

Analisi Kontrastiboa material didaktikoak prestatzeko metodologia baliagarria ere badela defendatzen dute teoria honen jarraitzaileek: azterketa kontrastibo baten bidez finka daitezke bi hizkuntzen arteko desberdintasun nagusiak eta horren araberrako baliabide pedagogikoak garatu. Bada, ikasmaterial onenak izango dira i) xede-hizkuntzaren deskripzio zientifikoetan oinarritzen direnak eta ii) xede-hizkuntza ikaslearen ama-hizkuntzaren deskripzio paralelo batekin alderatzen dutenak.

60.eko hamarkadatik aurrera Analisi Kontrastiboa kritikak jasotzen hasi zen. Hainbat zailtasun eta hutsune egotzi zitzaizkion, baita oinarri sendo baten gabezia ere:

- Batetik, elkarren artean antzekotasunik ez duten hizkuntzak konparatzen direnean, adibidez, asko izan daitezke aurreikusi ezin diren errore, egitura edota ezaugarriak. Izan ere, dena aurreikustea ezinezkoa da, aurreikusitakoa ez da beti zuzena eta aurreikusitakoa askotan ez da ikasketa-prozesuko egoera

errealetan agertzen (hau da, ikasleek benetan egiten ez dituzten erroreak aurreikusi ahal dira).

- Bestetik, errore guztiak ez dira ama-hizkuntzaren interferentziak eragindakoak. Dulay eta Burt-ek (1974), esaterako, ikerketa baten bidez frogatu zuten errore gehienak garapen-erroreak direla, eta ez ama-hizkuntzaren transferentziak eragindakoak; egindako esperimentutik ondorioztatu zuten erroreen % 3 baino ez zegokiola ama-hizkuntzaren interferentziari, % 85 garapen-erroreak zirela eta gainerako % 12 ikasleen ama-hizkuntzarekin zerikusirik ez zuten erroreek osatzen zutela. Ikertzaile horien arabera, beraz, bigarren hizkuntzetako ikasleek egiten dituzten errore berberak egin ohi dituzte hizkuntza horretako haurrek hitz egiten hasten direnean, eta kasu horretan ez dago ama-hizkuntzaren interferentziarik.
- Horrekin batera, bi hizkuntzen arteko desberdintasunak ikasle-talde homogeneoentzat soilik izango direla baliagarriak egotzi zitzaion Analisi Kontrastiboari, ama-hizkuntza ezberdina duten ikasleek batzuetan errore mota berberak egin ohi dituztela kontuan hartu gabe.
- Analisi Kontrastiboa irakasleek dagoeneko ezagutzen dituzten erroreak aurreikusteko baino ez dela gai eta ikaslearen gramatika-gaitasuna baino ez duela kontuan hartzen (funtzio komunikatiboari, testuinguruari edota erregistroari erreparatu gabe) ere kritikatu zitzaion teoria honi.

Euskarazko erroreen azterketa eta prozesamenduaren alorrean ditugun helburuak lortzeko, Analisi Kontrastiboa ez zaigu ikerketa-bide baliagarria: batetik, zaila ikusten dugu ama-hizkuntzan eta xede-hizkuntzan egindako ekoizpenak konparatuz ikasleek egiten dituzten erroreen edota desbideratzeen aurreikuspen errealak egitea; bestetik, errore edota desbideratze posibleak aurreikustea baino bideragarriagoa ikusten dugu testuetan detektatzen diren adibideak⁴ abiapuntutzat hartzea, erabiltzaileek dituzten zailtasunen eta beharrean arabera laguntza-tresnak garatu ahal izateko; gainera, analisi kontrastiboetatik eskuratutako emaitzak ez lirateke

⁴ Guk testu idatziak baliatzen ditugun arren, ahozko ekoizpenak ere erabil daitezke erroreen analisisirako.

guztiz aplikagarriak erroreen tratamendu automatikorako, eta gure helburua azterketa bera bi ikerketa-alorretan baliatzeko aukera izatea da.

Ezin ahantz dezakegu, ordea, Analisi Kontrastiboa izan zela Erroreen Analisia eta Tarte-hizkuntza oinarri duten azterketen abiapuntua, eta Analisi Kontrastiboaren ondotik hizkuntzen i(r)a(k)askuntza-prozesuak norabide berri bat hartu zuela, arreta osoa ikaslearengan jarritz, irakasleari garrantzia kenduz eta erroreari balio positiboa emanez.

II.2.2. Erroreen Analisia

70.eko hamarkadan Erroreen Analisia Analisi Kontrastiboari lekua hartzen hasi zitzaien, Analisi Kontrastiboari egotzitako arazo, hutsune eta ahuleziak gainditzen zituela aldarrikatuz. Corder izan zen Erroreen Analisiaren teoriaren aitzindari nagusia, 1967an *The Significance of Learners' Errors* argitaratzearekin batera sortu baitzen hizkuntzen i(r)a(k)askuntzaren alorrean teoria edo ikerketa-lerro berri hau. Hasieran kritika ugari jaso zituen, ikasleek egindako erroreen garrantzia aldarrikatzen baitzuen erroreak onartezinak eta saihestu beharrekoak ziren garaian. Pixkanaka, ordea, indarra eta sinesgarritasuna irabaziz joan zen.

Corder-ek (1967) bestelako iritzi bat eta ideia berriak plazaratu zituen hizkuntzen i(r)a(k)askuntzaren alorrean: Chomsky-ren ikuspuntu estrukturalista eta sortzailea oinarri hartuta, bigarren hizkuntza baten ikaskuntza-prozesua lehen hizkuntzaren jabekuntza-prozesuaren oso antzekoa dela azpimarratzen zuen, hau da, ama-hizkuntza eta bigarren hizkuntza ikasteko jarraitzen diren prozesuek ezaugarri berdintsuak erakusten dituztela, eta bien arteko desberdintasun bakarra hizkuntza ikasteko motibazioa dela. Gizakiak hizkuntzak ikasteko berezko gaitasuna du jaiotzetik, eta barne mekanismo horri esker gai da edozein hizkuntzatan, datu gutxi batzuetatik abiatuta, bere kabuz hipotesiak egin eta erregela edota gramatika berriak osatzen hasteko. Horretarako, baina, hizkuntza berri hori eskura izan behar du, hizkuntza horren inguruko datuak jaso behar ditu. Beraz, ama-hizkuntzaren jabekuntza-prozesuan erroreak saihestu ezintzat hartzen ditugun moduan, bigarren hizkuntza bat ikastean ere ezinbestekoak dira erroreak, beharrezkoak.

Stevens-ek (1969) dioten moduan, erroreak ez dira arazo, ez dira oztopo, ikasleen estrategien adierazgarri diren ezaugarri arrunt eta saihestu ezinezkoak

baizik. Erroreak dira ikaslearen xede-hizkuntzaren isla, hizkuntza berrian ikasleek duten mailaren eta jakintzaren adierazle, ikasleak hizkuntza ikasten ari diren seinale. Erroreen bidez ezagut ditzakegu irakasle edota ikertzaileok ikasleen zailtasun nagusiak, puntu ahulak eta benetako beharrak. Erroreak aztertuz hel gaitzke jakitera zein diren lehen, bigarren edota hirugarren hizkuntza baten ikaskuntza-prozesuaren ezaugarri nagusiak. Eta erroreen azterketatik abiatuta ondorioztatu ahalko dugu nola lagundu ahal zaien ikasleei xede-hizkuntza ikasten, hobetzen eta lantzen.

Corder-ek plazaratutako ideia edo ikuspuntu berriekin, Erroreen Analisisiak Analisi Kontrastiboak aurreikusten ez zituen zenbait errore mota detektatzeko gai dela erakutsi zuen eta erroreak balio negatiboa desagerraraztea lortu zuen: erroreak aintzat hartzen ez ziren akatsak izatetik, positiboki baloratuak izatera pasa ziren. Bada, hizkuntza-ikasleen testuekin osatutako corpus baten azterketa sistematikoa proposatzen du Erroreen Analisisiak, hau da, hizkuntza-ikasleek egindako erroreak ditu aztergai; eta errore horiek abiapuntutzat hartuta hizkuntzaren ikasketa-prozesua ezagutzea du helburu. Erroreen Analisia, gainera, teorikoa eta aplikatua izan daiteke. Analisi teorikoa bigarren hizkuntza bat ikastean ikasleak zer eta nola ikasten duen argitzean datza, zenbat errore egiten diren hizkuntza-maila bakoitzean ikustea, errore mota ohikoenak zein diren jakitea, etab. Analisi aplikatuak, berriz, ikasleen tarte-hizkuntza helburu pedagogikoekin aztertzen du, egindako erroreen azterketetan lortutako emaitzen arabera laguntza mota bat edo beste eskaintzeko xedeaz. Izan ere, ikasleen hutsune, zailtasun eta behar nagusiak ezagutzea nahitaezko urratsa da, i(r)a)kaskuntza prozesuan sortzen diren arazo nagusiei aurre egiteko lagungarriak izango diren ikasmaterialak edota estrategiak prestatu nahi badira. Helburu aplikatua, hala ere, azterketa teorikoan oinarritu ohi da.

Beraz, Corder-ek hizkuntzen i(r)a)kaskuntzaren alorrean plazaratutako ideia berriek eta erroreen inguruan egindako ikerketek erakutsitako emaitza positiboek indarra eta sinesgarritasuna eman zioten Erroreen Analisisaren teoriari 70.eko hamarkadaz geroztik. Eta azken urteotan nabarmen ugartu dira Erroreen Analisisaren inguruan egin diren ikerketak. Hala ere, badira teoria edo ikerlerro honi hainbat kritika eta hutsune kontzeptual leporatu izan dizkioten ikertzaileak (Bell, 1974; Hammarberg, 1974; Schumann eta Stenson, 1974; Schachter eta Celcé-Murcia, 1977; Dulay *et al.*, 1982; Martin del Rey, 2004; Díaz-Negrillo eta

Fernández-Domínguez, 2006, besteak beste). Oro har, erroreen azterketetan hutsune nagusi hauek azpimarratu izan dira:

- Erroreak bakarrik aztertzen dira, egitura zuzenak kontuan hartu gabe: azterketa gehienetan erroreak corpusetik ‘atera’ eta sailkatu egiten dira, testuingurua alde batera utziz eta zuzenak diren egiturei erreparatu gabe; hain zuzen, hizkuntzen i(ra)kaskuntzaren alorrean, ez da ezagutzen ikasleek ongi erabili dituzten egiturak aztertu dituen ikerketarik (Martin del Rey, 2004). Hizkuntzen i(ra)kaskuntzaren alorrean, ordea, erroreak soilik aztertzea ez da nahikoa ikasleen hizkuntza-gaitasuna edota hizkuntzaren i(ra)kaskuntza-prozesua osoki ezagutu ahal izateko. Aitzitik, ikasleek xede-hizkuntzan ekoizten dituzten egitura guztiak dira ezinbestekoak, erroredunak eta zuzenak; testuinguru osoa da beharrezkoa (Hammarberg, 1974; Schumann eta Stenson, 1974; Schachter eta Celcé-Murcia, 1977).
- Errore-kategoriak definitzeko eta erroreak sailkatzeko garaian, ohikoa izaten da zehaztasun falta. Izan ere, errorea zein motatakoa den finkatzea eta kategoria linguistiko jakin batean sailkatzea ez da ataza erraza. Definizio eta termino ugari erabili ohi direnez, errore beraren interpretazio bat baino gehiago aurki daitezke: errorea definitzea edota sailkatzea subjektiboa izan daiteke batzuetan. Ondorioz, ez da irizpide konkreturik jarraitzen, oro har, erroreen sailkapenak egiteko (Schachter eta Celcé-Murcia, 1977). Hala, batzuetan erroreen zerrenda hutsak agertu ohi dira, ongi antolatu gabeak, irizpide jakinik gabe osatuak; beste batzuetan, berriz, erroreen sailkapenetan azpikategoriak sinpleegiak dira eta erroreak deskripzioa, zergatia edota iturburua nahasteko joera egon ohi da (Dulay *et al.*, 1982; Martin del Rey, 2004). Argi dago, beraz, Erroreen Analisiaren literaturan eztabaida gutxi sortu dela errore-sailkapenen egituraren inguruan eta gai horri dagozkion zenbait alderdi argitzeke daudela oraindik. Ondorioz, ikertzaile bakoitzak bere helburuen arabera taxonomiak egin ohi ditu normalean. Irizpide linguistikoa eta deskribatzailea oinarri hartuta osatu dugu guk euskarazko erroreen eta desbideratzeen sailkapena (V. kapitulua), taxonomiaren egitura ahalik eta koherenteena izan dadin.
- Errore sistematikoen zergatia(k) azaltzeko orduan desadostasunak sortzen dira. Errorea egitearen arrazoia zein den jakitea lagungarria izan daiteke materialak prestatzeko eta zenbait ikerketa egiteko (Schachter eta Celcé-Murcia, 1977).

Baina hemen ere desadostasunak azaltzen dira irakasleen eta ikertzaileon artean, ondorio horiek ateratzea zaila eta subjektiboa izan daitekeelako edota errore batek kausa bat baino gehiago izan ditzakeelako. Bestalde, erroreen sailkapenetan errorearen deskripzioa, zergatia edota iturburua nahasteko joera dago. Ordea, errorearen ezaugarriak deskribatzea (zein motatakoa den edo zein kategoria linguistikori dagokion esatea) eta errorea egitearen arrazoi(k) ondorioztatzea (ama-hizkuntzaren eragina, transferentzia negatiboa, itzulpen-estrategia okerren bat erabili izana...) bi zeregin dira, eta ez lirateke sailkapen berean nahastu behar. Erroreen Analisari kritikatzeko zaio, hain justu, oraindik ez dela errore motak ongi bereizteko irizpiderik finkatu eta ez dagoela oinarri sendodun teoriarik errore desberdinen zergatien inguruan (Dulay *et al.*, 1982; Martin del Rey, 2004). Aurrerago ikusiko dugun moduan (V. kapitulua), tesi-lan honetan aurkezten dugun euskarazko erroreen eta desbideratzeen sailkapena irizpide linguistiko eta deskribatzaileetan oinarrituta dago, eta ez ditu erroreen zergatiak barne hartzen.

- Xede-hizkuntzako egitura zailenak bakarrik identifikatzeko joera dago, maiz uste baitugu ikasleek gehien errepikatzen dituzten erroreak egitura zailenetan gertatu ohi direla. Hori, ordea, ez da beti egia, askotan, erroreak egiteko beldurrez, ikasleek zenbait egitura saihesten baitituzte. Eta hizkuntzaren i(r)a(k)askuntza prozesua osotasunean ezagutu ahal izateko, ikasleek zein errore egiten dituzten eta zergatik jakitea bezain garrantzitsua da zein errore ekidin ditzaketen eta zergatik ekiditen dituzten jakitea. Erroreen Analisisa ez da gai ikasleek baztertzen dituzten fenomeno linguistiko horiek guztiak atzemateko eta azaltzeko (Schachter eta Celcé-Murcia, 1977).
- Hizkuntza-ikasleen edota testuen inguruko informazioa askotan ez da kontuan hartzen: ikertzaile batzuk erroreak detektatzera, sailkatzera eta aztertzer mugatzen dira, ikasleen inguruko informazio psikolinguistikoa (ama-hizkuntza, adina, hizkuntza ikasten daraman denbora, etab.) eta corpusaren ezaugarri nagusiak (testuaren generoa, gaia, ikasleari agindu zaion ariketa mota, etab.) kontuan hartu gabe. Erroreen eta desbideratzeen analisirako garatu dugun lan-inguruneak (III. kapitulua) errore-adibideekin batera desbideratzeak eta egitura zuzenak aztertzeko aukera ematen digu, baita erroreen eta hizkuntza-ikasleen inguruko informazioa gorde eta baliatzeko aukera ere. Izan

ere, informazio hori guztia beharrezkoa da hizkuntzaren ikasketa prozesua ezagutzeko. Are gehiago, kasu batzuetan azterketaren emaitzak edota ondorioak ere baldintza ditzake horrek. Testu motaren arabera, esaterako, mota bateko zein besteko erroreak egin daitezke, erabiltzen diren egiturak askotarikoak izan daitezke eta errore moten maiztasuna ere alda daiteke. Era berean, testu motaren arabera, egitura bat erroretzat hartu ahal izango da, edo ez (Schachter eta Celcé-Murcia, 1977); izan ere, testu motak eragina du egitura bat erroretzat hartzeko edo ez hartzeko garaian.

- Bell-ek (1974) ere kritika egiten dio Erroreen Analisari ondorio estatistiko urriak eskaintzen dituelako, errorearen interpretazioa subjektiboa izan ohi delako eta ikasleen benetako erroreak aurreikusteko ahalmen handiegirik ez duelako. Gure ustez, ordea, corpus-lagin adierazgarriak bilduz gero, posible da ikasleek egindako errore edota desbideratze mota ohikoenen aurreikuspenak egitea; horixe da, hain zuzen, corpusak erabiltzearen abantaila bat.

Laburbilduz, hauek dira Erroreen Analisiak jaso izan dituen kritika batzuk: erroreak bakarrik aztertzen dira, desbideratzeak, egitura zuzenak edota testuingurua kontuan hartu gabe; errorearen sailkapenak edo kategoriak definitzen direnean, zehaztasuna falta ohi da; errore sistematikoen zergatia(k) azaltzen direnean, desadostasunak agertu ohi dira; errorearen deskripzioa eta zergatia askotan nahastu egiten dira; xede-hizkuntzako egitura zailenak bakarrik identifikatzeko joera dago; hizkuntza-ikasleen edota testuen inguruko informazioa ez da kontuan hartzen; eta errorearen interpretazioa subjektiboa izan daiteke.

Egia da Erroreen Analisiaren literaturan oraindik ere oinarritzko irizpide batzuk zehazteke daudela, oro har; alderdi batzuk ez daudela guztiz argi. Baina, aldi berean, ekarpen interesgarriak egin dira azken urteotan ikerketa-alor horretan. Erroreen Analisiak, Analisi Kontrastiboak ez bezala, erroreak soilik xede-hizkuntzan oinarrituta deskribatzen eta tratatzen ditu, ikaslearen ama-hizkuntza kontuan izan gabe (James, 1998), eta errore-adibideez gain, ikasleen tarte-hizkuntza eta ikasleei buruzko informazio psikolinguistikoa aztertzeko beharra ere kontuan hartzen du. Izan ere, errorearen balio negatiboa desagerrarazteaz gain, Corder-ek (1974) kontzeptu berri bat asmatu zuen: dialekto iragankorra, ama-hizkuntzatik eta xede-hizkuntzatik desberdina den sistema, baina era berean, ama-hizkuntzaren edota

xede-hizkuntzaren ezaugarriak eta ezaugarri idiosinkratikoak dituen. Termino horretan oinarrituta sortu zuen Selinker-ek tarte-hizkuntzaren kontzeptua (II.2.3. atala). Erroreen Analisia, beraz, Analisi Kontrastiboaren eta Tarte-hizkuntzaren arteko zubia izan zen.

Erroreen Analisiaren teoria hizkuntzen i(ra)kaskuntzaren alorrean sortu zen arren, gaur egun ez da hizkuntza-ikasleen ekoizpenetara mugatzen; aitzitik, Hizkuntzalaritza Aplikatuaren barruko hainbat alorretan ezinbestean dira aztergai erroreak edota desbideratzeak. Izan ere, hiztun guztiok egiten ditugu mota bateko zein besteko hanka-sartzeak noiznahi, nonahi (dela azterketetan, dela gutunetan, dela argitalpenetan, dela hizketa-egoera jakinetan...). Horregatik, HAPen alorrean ere, erroreen eta desbideratzeen tratamendu automatikoa bideratuz, ekarpen garrantzitsuak egin daitezke bai hizkuntza-ikasleentzat (xede-hizkuntzaren inguruan egiten dituzten hipotesiak egiaztatu edota ezeztatu ahal izateko), bai hizkuntza-irakasleentzat (ikasleen beharrak zein diren modu errazean ezagutu ahal izateko) eta bai ikertzaileentzat ere (hizkuntzaren ikasketa-prozesua bideratzen laguntzeko baliabideak prestatu ahal izateko).

II.2.3. Tarte-hizkuntza

Tarte-hizkuntza terminoa Selinker-ek (1969 eta 1974) proposatu zuen *Language Transfer* argitalpenean, nahiz eta kontzeptu bera lehenagotik sortua zen: Corder-ek dialekto iragankorra (*dialecto transitorio*), dialekto idiosinkratikoa (*dialecto idiosincrásico*) edota gaitasun iragankorra (*competencia transitoria*) etiketak erabili zituen termino horretarako, Nemser-ek (1971) hurbilketazko sistema (*sistema aproximativo*) eta Porquier-ek (1975) bitarteko sistema (*sistema intermediario*).

Tarte-hizkuntza ikasle batek bigarren hizkuntza ikasten duen heinean sortzen doan sistema edo kode linguistikoa da, etengabe aldatzen doana ikasleak xede-hizkuntza 'lortzen' duen arte. Selinker eta haren jarraitzaileak arduratu ziren hizkuntza-ikasleen tarte-hizkuntza aztertzeaz: ikasleek une bakoitzean duten hizkuntzaren ezaugarriak eta hizkuntza horrek duen bilakaera zuten aztergai nagusi. Tarte-hizkuntza konplexua da eta konplexutasun hori areagotuz joan ohi da ikasketa-prozesua aurrera doan heinean. Ikasleak bereganatzen dituen egitura berriak eta lexikoak markatzen dute ikasle bakoitzaren sormen-prozesuari esker urratsez

urrats garatuz doan tarte-hizkuntza (Martín del Rey, 2004). Esaldi zuzenez eta erroredunez osatutako sistema den arren, erroreak izango dira tarte-hizkuntzaren adierazle nagusiak, ikaslea ikasten ari den seinale. Bestalde, tarte-hizkuntza horretan ikasleak estrategia komunikatibo batzuk asmatu eta baliatu ohi ditu, xede-hizkuntzan sortzen zaizkion komunikazio-arazoak gainditzeko. Estrategia horiek ikertzea ere ezinbestekoa da hizkuntzaren ikasketa-prozesua osotasunean ezagutu ahal izateko. Tarte-hizkuntzak hiru ezaugarri nagusi ditu: ikasle baten diskurtsoa ama-hizkuntzarekiko eta bigarren hizkuntzarekiko independentea da eta barnetik egituratua den sistema linguistikoa baten emaitza da; sistema linguistikoa hori aldatzen joaten da ikaskuntza-prozesuan zehar; eta egoera jakin batean maila bereko ikasleen sistema linguistikoa oso antzekoa da, ikaskuntza-esperientziaren ondorioz sortutako diferentzia batzuk kenduta (Maritxalar, 1999; Selinker, 1974; Nemser, 1971; Corder, 1971).

Ikasle bakoitzak une jakin batean duen sistema ezagutuz, ikasleak xede-hizkuntza ikasteko jarraitzen duen prozesua aztertzea da, beraz, tarte-hizkuntzaren azterketaren helburu nagusia.

II.3. Errore kontzeptua

Errore kontzeptuari dagokion gaia zabala da eta faktore asko barne hartzen ditu. Horregatik, erroretzat hartzen dena, askotan, bakoitzaren helburu edota ikerketa-lerroaren arabera izan ohi da. Gure xedea euskarazko erroreen tratamendu automatikoa bideratzeko eta gure hizkuntzaren i(r)a(k)askuntza prozesuaren alorrean ekarpenak egiteko baliabideak prestatzea da. Baina *zer hartzen dugu erroretzat?*

II.3.1. Euskararen egoera soziolinguistikoa

Euskarak egoera soziolinguistikoa berezia bizi du: hizkuntza-komunitate elebidun batean bizi den hizkuntza gutxitua eta eremu urrikoa da. Gutxitua izan da, urte askotan, euskararen erabilera; eta kasu batzuetan baita debekatua ere hainbat leku,

egoera edota mementotan. Eremu urriko hizkuntza da, bestalde, hiztun kopuru txiki batena baino ez delako. Horren adierazgarri dira 2006an bildutako datuetan oinarritutako IV. Inkesta Soziolinguistikoaren emaitzak⁵.

Eremu urriko hizkuntza gutxitua den arren, euskara indartuz joan da hizkuntzaren alde abian jarri diren ekimenei esker. 1968an euskara idatziaren batasuna bideratzeko ahalegin sendo eta formalak hasi zirenetik eta azken 25-30 urteotan finkatzen joan den politika linguistikoari esker, nabarmen berreskuratu da euskara eta hazi da euskaldunon kopurua. Berreskurapen horretan, eragin handia izan du 1982 urteko Euskararen Erabilpenaren Normalizazio Linguistikorako Oinarrizko Legeak. Lege horrek “*Euskal Herriaren berezko hizkuntza eta hizkuntza ofiziala gaztelaniarekin batera*” definizioa emanez, euskararen ofizialtasuna ezarri zuen (Alonso, 2003). Legeak, gainera, hizkuntzaren erabileraren berreskurapena eta normalizazioa lortzeko hainbat neurri artikulatzen zituen, eta legea indarrean sartzearekin batera, areagotu egin ziren hizkuntzaren hedapenerako programak, ekintza sozial eta kulturalak, ikastolen eta beste erakunde ofizial batzuen jarduerak, etab. Beraz, euskararen estandarizazioa mamitzeko eta planifikatzeko ahaleginak lehenagokoak diren arren (1918an sortu zen Euskaltzaindia), 1968an ekin zitzaion, zinez eta formalki, euskara idatziaren batasuna bideratzeko zereginari. Euskara batua diseinatzeko, erdialdeko euskalkietan (gipuzkera, lapurtera eta behe-nafarrera) oinarritutako eredu konposizionala hartu zen abiapuntutzat, euskalki guztien ekarpenak bilduko zituen eredua. Eginkizun horren alderdi arauemailea Euskaltzaindiak⁶ hartu zuen bere gain.

Orduz geroztik, euskara batuaren kodifikazioari, ezarpenari eta garapen lexiko-diskurtsiboari ekin zitzaion. Prozesu horiek fase modura ikusten dira hizkuntzen normalizazio-plangintzei buruzko bibliografian (Haugen, 1983; Bibiloni,

5

[http://www.euskara.euskadi.net/r59-738/eu/contenidos/libro/iv_inkesta_soziol/eu_ink/adjuntos/IVInkesta\(Eusk\).pdf](http://www.euskara.euskadi.net/r59-738/eu/contenidos/libro/iv_inkesta_soziol/eu_ink/adjuntos/IVInkesta(Eusk).pdf)

⁶ “Sortu eta lasterrera, hiru helburu nagusi ezarri zizkion bere buruari: a) ortografia arautzea, b) hiztegia zehaztea, eta c) Euskal Herri osorako literatura hizkera eraikitzea. 1920an argitaratu zuen bere arautegia, eta haxe zioen 6. artikulua: *Ekingo du euskalki oroen odolez azitako euskal-eredu bat gertutzen, itzez, itz-yoskeraz ta idazkeraz albait erabatetsuena, euskaldun guziak idatz-irakurketan erabiltzekoa*” (Zuazo, 2005).

1997). Nolanahi ere, euskararen kasuan kodifikazioa, ezarpena eta garapen lexiko-diskurtsiboa hasieratik gainezarri diren prozesuak izan dira. Horrek azkartu egin du estandarizazio-prozesua, baina egokitzea eta berregokitzea ere eskatu du hainbat mementotan (Elordui eta Zabala, 2009). Hizkuntzaren estandarizazio-prozesua bukatu gabe dagoenez eta euskara batuaren kodifikazioa, ezarpena eta garapen lexiko-diskurtsiboa gainezarri egin direnez, eredu linguistiko aldakorra jasotzen dugu euskaldunok.

Bestalde, gure hizkuntza-komunitatean hainbat dialekto hitz egiten dira eta oraindik ez dago guztiz ezarrita euskara batuaren eta dialektoen arteko banaketa funtzionala. Ondorioz, egoera soziolinguistiko askotarikoak gertatzen dira: hiztun mota ugari bizi gara elkarrekin eta hiztun bakoitzaren egoerak, inguruneak eta lanbideak, besteak beste, baldintzatu egiten dute hizkuntza-gaitasuna, hizkuntzaren erabilera, hizkuntza ikasteko helburuak edota motibazioa(k). Dibertsitate edota ezegonkortasun horrek zalantzak sorrarazi ahal dizkigu hiztunoi, eta erroreak egiteko aukerak ugaritu egiten dira. Euskara ikasten ari diren ikasleen kasuan, hizkuntzaren normalizazio ezaren eragina are larriagoa da: ikasle bakoitzak xede-hizkuntzara heldu artean jasotzen duen inputa ezegonkorra bada, zalantzez eta zailtasunez josita ikasiko du hizkuntza, ziurtasun murriztarekin eta errore askorekin.

Testuinguru soziolinguistiko horretan HAPen alorrean euskara ikasteko edota lantzeko laguntza-tresnak garatuz, gure hizkuntza-komunitateari ekarpen garrantzitsuak egin ahal dizkiogula aurreikusten dugu.

II.3.2. Euskara batua eta euskalkiak

Hizkuntzaren biziraupena bermatzeko sortu zen euskara batua, estandarra edo arautua. Geroztik, euskara batuaren eta euskalkien arteko banaketa funtzionalari buruzko eztabaida bizirik dago euskararen hizkuntza-komunitatean, eta gai horren inguruan hainbat gogoeta egin dira, askotariko iritziak plazaratu dira eta bibliografia oparoa argitaratu da (King, 1983; Larringan, 1993; San Martin, 1998; Villasante, 1998; Zuazo, 2000 eta 2005, etab.). Izan ere, euskara batua berriki sortutako aldaera da, lehenagotik existitzen ziren dialektoekin erabilera esparruak banatu behar dituen. Gu ez gara euskalkiek euskara batuari egin behar dioten ekarriaren auzian sakonki sartuko, baina oinarritzko aipamen edo azalpen batzuk ezinbestekoak irudit

zaizkigu, erroreen azterketa eta prozesamendua bideratzeko lanetan eredu linguistiko estandarraren eta ez-estandarraren arteko bereizketa garrantzitsua baita.

Euskara batua beharrezkoa da gure hizkuntzaren berreskurapenerako eta biziraupenerako, leku guztietako euskaldunok elkarri ondo ulertzeko hizkuntza-eredu bat izateko. Bibiloni-ren (1997) hitzetan, *“una llengua estàndard és una varietat, altament codificada, que serveix de model de referència a tots els membres d’una comunitat lingüística en uns àmbits determinats o en tots els àmbits. A més d’aquesta funció bàsica, però, l’estàndard fa altres funcions: és un element clau per a l’existència de la consciència de comunitat lingüística i per a l’existència d’un sistema equilibrat d’actituds positives dels parlants vers la pròpia llengua, indispensable perquè aquesta tingui garanties de plenitud i de continuïtat”*. Horrek ez du esan nahi, ordea, euskalkiak gutxietsi edota baztertu behar direnik, eredu estandarrarekin batera euskalkiak erabili behar ez direnik. Izan ere, eurei dagozkien une eta eremuetan euskalkiak erabiltzen jarraitzea ezinbestekoa da euskararen aberastasunerako: *“euskara herrikoia ez da euskara txarra, erregistro herrikoia dagozkion egoeretan erabiltzen denean. Alderantziz, orduan goi mailako erregistroko euskara gerta daiteke tokiz kanpokoia eta, neurri batean behintzat, herritarren artean desagokia”* (Oyharçabal, 1996). Hizkuntza komunikazio-tresnatzat hartzen dugun garai honetan, beraz, euskalkia(k) ezagutzea garrantzitsua da hainbat komunikazio-egoeratan moldatu ahal izateko. Bestalde, une honetan euskara batuaren kodifikazio prozesuan garrantzi handiena hartu duen prozesua lexikalizazioa da, Hiztegi Batuaren finkapenean gauzatzen dena; eta lexikalizazioan funtsezkoa da dialekto guztien ekarria.

Zuazoren (2000) arabera, euskara batuaren eta euskalkien banaketa hizkuntzaren erabilera funtzionalak eta eremu geografikoak baldintzatu behar dute: *“Ona da hizkuntza bera erabiltzen duten hiztun guztiek eredu bakarra izatea, baina eredu bakar hori ez dago une eta gune guztietan zertan erabili (...) eremuaren luze-zabalak aginduko du, alde batetik, hizkera estandarrari edo euskalkiari lehentasuna zeini eman, eta bestetik eta bereziki, jardunaren nolakoa izan beharko dugu kontuan. Eskualde jakin bateko ahozko jardunean, bertako euskalkia erabiltzea izan daiteke zentzuzkoena eta onuragarriena, eta ez euskara batua”*.

Oyharçabal (1993) harantzago doa tokian tokiko berezitasunak integratzen dituen euskara batua aldarrikatzen duenean: *“bateratzea beharrezkoa badu euskarak sendotzeko, denek dakigu ez dela aski euskaldun anitzen hizkuntza jokabide ezkorren funski aldarazteko (...) eta ez da aski euskara batuaren gomendioei jarraikiz idaztea, testu bat euskaldun guztientzat irakurterraza edo berehalakoan irakurgarri gerta dadin”*. Euskara batuari eta euskalkiei eskatzen zaizkien baliabideak desberdinak ez ezik, elkarren aurkakoak ere badira, bere ustetan: *“batak arautegi hertsia, homogeneousotasuna, kolorerik eta usainik eza behar du, argitasuna eta zehaztasuna beharrezkoak dituelarik; bestea, berriz, aniztasuna, malgutasuna, arauen mugetan ibiltzea eta espresibitatea eskatzen du oroz gainetik. Argitasuna eta zehaztasuna bezain beharrezkoak ditu, kasik, anbiguitatea eta lausotasuna”* (Oyharçabal, 1993). Hala, euskara batuak euskalki guztietako hitz guztiak onartu behar ditu *“euskal altxorrekoak diren ber, inongo hitzik baztertu gabe (...) eta euskararen joskera errespetatzen den puntutik, nork bere esku du adierazpideak eta esamoldeak euskalkitik hartzeko aukera”* (Oyharçabal, 1996).

Maiak eta Elorduik hainbat lan argitaratu dituzte (Maia, 2001a; Maia, 2001b; Elordui eta Maia, 2002; Elordui, 2006) hizkuntza estandarra eta dialektoak i(r)a(k)skuntzan, oro har, nola ulertu eta tratatu behar diren aztertzen dutenak. Bere aburuz, azken urteotan (70.eko edo 80.eko hamarkadatik honantz) asko aldatu da hizkuntzen i(r)a(k)skuntzaren alorrean eredu estandarraren eta euskalkien erabileraren inguruko ideia eta ikuspuntua. Hizkuntza-eredua modu zurrunean ulertzeko joeraren kontrako iritziak han eta hemen zabaldu dira: gaur egungo i(r)a(k)skuntzaren teorizazioaren arabera, eredu estandarra ez da nahikoa leku eta egoera guztietan komunikazio eraginkorra erdiesteko; eredu zurrun bakar baten aurka defendatzen da hizkuntzak funtzio ugari dituela eta horiei erantzuteko hainbat hizkuntza-eredu direla egokiak, kasuan kasu. Orain, beraz, euskalkiaren eta tokian tokiko hizkeren errealitateari garrantzi handiagoa ematen zaio; izan ere, bizitzan arrakastaz komunikatzen ikasteko, hizkuntzaren errealitate konplexua hartu behar da aintzat i(r)a(k)skuntzaren alorrean: eredu estandarrarekin batera komunikazio-egoera bakoitzak eskatzen duen erregistroan edota hizkera-mailan moldatzeko gaitasuna ahalbidetuko digun hizkuntzaren i(r)a(k)skuntza da behar duguna. Erregistro desberdinetako baliabideak beharrezkoak dira, beraz, hizkuntzaren kode ezberdinak menderatu eta edozein komunikazio-egoeratan edo eremu geografikotan ondo

moldatzeko. Halaber, euskararen i(ra)kaskuntzaren alorrean, indar gehiago izan behar lukete euskalkiek hizkuntza-maila baxuetan (dialektoa ikasleek kalean entzuten duten hizkeratik hurbilago baitago) eta hizkuntza menderatu ahala hartu behar lukete garrantzia euskara estandarrean oinarritzen diren erregistro formalek (Barrios *et al.*, 2008).

Trudgill (1983) ere eredu estandarrarekin batera ikaslearen jatorriko dialektoa onartzearen eta i(ra)kastearen aldekoa da. Azken finean, hizkuntzak aldaera desberdinak ditu, ez da zurruna, ez da finkoa. Hori abiapuntutzat hartuta, beraz, hizkuntza bat ongi jakiteko nahikoa al zaigu eredu estandar idatzia bakarrik menderatzea ala zerbait gehiago behar dugu? Cassany *et al.*-en (1994) hitzetan, “*el alumno más preparado lingüísticamente es el que puede hablar y escribir en su variedad dialectal propia y en el estándar correspondiente, con varios registros, y también el que puede comprender otras variedades distintas a la suya*”. Gaur egun “*euskara batua onartua da gizartean, baina ez da askiesten bazter guztietan aritzeko; euskalkiaren aldeko kezkak eta iritziak sumatzen dira jendartean; euskalkien berri hurbilagotik dakigu (xehetasun asko desberdinak dira, tokiaren arabera); irakaskuntzan, konstruktibismoak indarra hartu du, eta horrekin loturik datozen kontzeptu batzuk (ikasketa esanguratsua, garapen hurbileko eremua, motibazioa) euskalkiari jaramon egitera gonbidatzen edo behartzen dute (...) Euskara batua behar dugu (kulturaren dimentsio osoaz jabetzeko, eta euskaldun guztien arteko komunikazioa bermatzeko, euskaraz, erdarak indartsu diren gizartean) (...) Baina gizarte euskalkidun batean (herri batean, eskualde batean) euskara batua sarritan arrotza gerta daiteke, hotza, berotasunik gabekoa, urrutikoa. Horrek eskatzen du eskolak arreta jartzea beste eredu "funtzionalago" bati, tokian tokiko komunikazio moduei jaramon berezia eginen liokeena. Eskolako jarduna herri jakin batean garatu behar da, eta eskolako hizkuntz ereduak kontuan eta aintzat hartu behar ditu tokiko hizkuntz aldaerak eta berezitasunak*” (Maia, 2001a).

Maiak, Elorduik eta Barrios *et al.*-ek gogoeta edota planteamendu horiek eskolako i(ra)kaskuntzaren alorrera zuzenduta egiten badituzte ere, ideia berberak baliagarriak dira bigarren hizkuntzaren i(ra)kaskuntzaren alorrean, oro har, eta nola ez, baita HAPen alorrean ere; hau da, hizkuntza estandarraren eta ez-estandarraren arteko auziak eskolaren eremutik haratago hizkuntzaren erabilera-eremu guztietan du eragina: biak, euskara batua eta kode dialektalak, ezinbestekoak zaizkigu gure

hizkuntza-komunitatean, bai gure komunikazio-gaitasunak sendotzeko eta bai gure hizkuntzaren biziraupena bermatzeko ere. Euskaltzaindiak, euskalkien erabileraz egiten duen 137. adierazpenean (Euskaltzaindia, 2004), zera dio: *“euskalkia eta batua elkarrekin uztartu beharrak sortzen dituen arazoen konponbideak era askotakoak izan daitezke, eta konponbideok kasuan-kasuan egokitu behar zaizkio aldian aldiko komunikazio-egoerari; betiere batasunari eutsiz eta euskal hiztunen bazterketa gutxietsiz”*.

Gure ikerketa, oro har, idatzizko jardunera mugatzen da, euskararako gramatika-eta estilo-zuzentzaile bat garatzeko eta euskararen i(ra)kaskuntzarako baliagarriak izan daitezkeen laguntza-tresnak sortzeko asmoak gidatzen baitu. Idatzi askotan hizkuntza estandarra eskatzen da, batuaz idatzi beharra dugu; eta horretarako, ezinbestekoak zaizkigu oinarrian euskara estandarrari dagozkion arau linguistikoak betetzen laguntzen diguten baliabideak. Horrek ez du esan nahi, baina, euskara batuan soilik idazten dugunik, ez eta forma ez-estandarrak kasu guztietan errore direnik. Izan ere, hainbat testuingurutan eta egoeratan forma dialektalak erabiltzea onargarria bezain zuzena da. Eta ideia horren aldeko apustua gero eta nabarmenagoa da, azken aldian, gure hizkuntza-komunitatean: egun, gero eta leku gehiago dute euskalkiek, dela telebistan, dela aldizkarietan, dela ekitaldi ofizialetan, elkarriketetan, literaturan...

Bestalde, asko ugaritu dira, egun, ordenagailu bidezko hainbat komunikazio-egoera: posta elektronikoa, foroak, *chata*... Komunikazio mota hau, askotan, ahozko elkarriketaren ordezkotzat izan ohi da; hau da, gai jakin baten inguruan lagun artean hitz egitea da helburua eta erabili ohi den erregistroa nahiko informala izan ohi da, komunikazio-egoera ofizial edota instituzionaletan eskatzen den hizkuntza-estilo formalarekin alderatuta (Pano, 2008). Komunikazio-egoera horietan euskara estandarra oinarri duten bitartekoak baino baliagarriagoak izango litzaizkiguke aldaki ez-estandarrak ere ezagutzen dituzten laguntza-tresnak.

Euskara estandarraren eta ez-estandarraren auziak eragina dauka, beraz, gure ikerketetan: HAPen alorrean euskararen egoera soziolinguistiko erreala kontuan hartzea garrantzitsua da gure hizkuntza-komunitatearen beharrak ase nahi baditugu. Ondorioz, euskara batua oinarri hartzea abiapuntu egokia den arren, ez da nahikoa; eta hizkuntzaren eredu estandarrari dagozkion baliabideak eta dialekto

desberdinetako ezaugarriak bereizten eta tratatzen dituzten tresna informatikoak garatzeko premia aurreikusten dugu, euskararen garapen funtzionalari eta hiztunen komunikazio-gaitasunari lagundu ahal izateko.

II.3.3. Errore kontzeptua definitzeko hainbat ikuspuntu

Errorea, berez, hizkuntza-komunitate batean ezarritako arauetatik (gramatika-arauetatik edo arau soziopragmatikoetatik) at dagoen oro dela esan daiteke. Hala ere, errore kontzeptuaren inguruan termino eta definizio asko erabili izan dira literaturan (Corder, 1967; Norrish, 1981; Dulay *et al.*, 1982; Fernández, 1991 eta 1997; Larsen-Freeman eta Long, 1994; James, 1998; Maritxalar, 1999; D'Aquino eta Ribas, 2001; Torijano, 2004; Alexopoulou, 2005; Oronoz, 2009, besteak beste) eta hainbat errore mota bereizi izan dira (*errors, mistakes, deviations, flaws, faults...*).

Dulay *et al.*-ek (1982) eta Fernández-ek (1991), esaterako, errore esaten diote komunitate jakin batean ezarritako arauak betetzen ez dituen egitura desbideratu orori, eta ez diete erreparatzen erroreen ezaugarriei, ez eta erroreak egin izanaren zergatiei.

Corder-ek (1967) erroreak (*systematic errors*), hutsegiteak (*mistakes*) eta lapsusak (*lapsus*) bereizten ditu. *Erroreak* hizkuntza-ikasleak ez-jakintasunagatik (arauak edo forma zuzenak oraindik ikasi ez izanagatik) erabiltzen dituen egitura/forma okerrak dira, sistematikoak dira eta ikaslearen gaitasun linguistikoari lotuta daude (ez jardunari). Erroreak ikaslea ikasten ari den seinaleztat hartu behar dira, ikaslearen tarte-hizkuntzaren adierazgarri dira. Bestalde, normala da ikasleak guztiz menderatzen ez dituen egituretan tarteka hutsak egitea, egitura horiek batzuetan ongi eta beste batzuetan gaizki erabiltzea. Araua edo forma zuzena ezagutu arren ikasleak nahi gabe egiten dituen akats ez-sistematikoei *hutsegite* esaten die Corder-ek. *Lapsusak*, berriz, kontzentrazio ezagatik, arreta faltagatik, nekeagatik... edonork egiten ditugun okerrak dira (inola ere ez sistematikoak). Hizkuntza-ikasleak gai dira jardunari dagozkion hutsegiteak eta lapsusak zuzentzeko; ez, ordea, hizkuntza-gaitasunari lotutako erroreak atzemateko, ez eta zuzentzeko ere. Corder-en ekarpen horrek eragin handia izan zuen hizkuntzen i(r)akaskuntzaren alorrean, ordura arte erroreak gaitzesten zituen teoria konduktistaren aurka agertu eta erroreak

hizkuntzaren ikasketa prozesuaren adierazgarritzat hartu baitzituen. Norrish (1981), James (1998) eta Torijano (2004) ere bat datoz Corder-en bereizketa eta definizioekin.

Lee-k (1990) erroreak (*errors*) eta hutsegiteak (*mistakes*) bereizten ditu. Idaztean edo hitz egitean egoera fisiko edota mentalek (nekeak, estresak, arreta faltak...) eraginda hiztun arruntek edota hizkuntza-ikasleek egindako erroreei hutsegite esaten die. Hutsegiteek, beraz, performantziarekin dute zerikusia, jardunarekin, eta ez hizkuntza-gaitasunarekin. Hiztuna bera jabetzen da egindako hutsegiteaz eta zuzentzeko gai da. Hizkuntzaren i(r)a(k)askuntzaren ikuspuntutik, erroreak ikaslearen hizkuntza-gaitasunaren araberakoak dira bere ustez, hizkuntzaren ezagutza-mailaren eta ikasketa-prozesuaren isla. Ikaslearen tarte-hizkuntzaren edota ikasleak ikasi duenaren adierazgarri dira errore sistematikoak eta ez dute performantziarekin zerikusirik. Ezjakintasunaren eraginez egiten direnez, norbera ez da bere erroreak zuzentzeko gai. Gu ez gatoz bat bereizketa horrekin: ez dugu uste erroreak eta hutsegiteak hiztun arrunt edota hizkuntza-ikasleekin lotu behar direnik, hiztun arruntek ere egin baititzakete bai hutsegiteak eta baita erroreak ere. Horregatik, hain zuzen, ortografia-, gramatika- eta estilo-zuzentzaileak edota kontsulta-tresna informatikoak ez dira hizkuntza-ikasleei bakarrik zuzendutakoak.

Errore kontzeptuaren inguruan, beraz, hainbat termino eta definizio eman dira. Hala ere, autore batzuek (Sridhar, 1981; Jain, 1984 eta Bustos, 1998, esaterako) diote definizio gehienak sinpleegiak direla eta zenbait faktore garrantzitsu ez direla aintzat hartzen. Bestalde, erroreen inguruko definizioak murriztaileak direla ere kritikatu izan dute; alegia, askotan izaera formala (lexikoa, semantika, fonetika, morfologia, sintaxia eta morfosintaxia) baino ez dela kontuan hartzen. Horren aurrean, Bustos-ek (1998) erroreak bi ikuspegitatik definitu behar direla dio, ikuspegi formaletik eta ikuspegi funtzionaletik. Bere iritziz, errore bat gramatika-, lexiko-, fonetika- edota testu-mailakoa den esateaz gain, beharrezkoa da errorea komunikatiboa (komunikazio-egoeraren parametro jakin baten aurrean –gaia, solaskidea, xedea...– huts egiten duena) edo soziopragmatikoa (ingurune soziokulturalaren eskakizunen bati egokitzen ez zaiona) den zehaztea.

Gure ustetan ikuspegi funtzionala eta formala lotuta doaz. Egiten dugun erroreen analisia ikuspegi formalean kokatzen da: egitura erroredunen forma da mementoz

aztertzen duguna. Baina erroreen azterketa erabatekoa izan dadin, ezinbestekoa da ikuspegi funtzionala ere kontuan hartzea. Horren adibide garbia da forma estandar eta ez-estandarren inguruko gogoeta (II.3.2. atala). Izan ere, ikuspegi formaletik, testu idatzi batzuetan detektatzen diren forma dialektal asko edota arau akademikotik urruntzen diren baliabideak ez dira testuinguru horretan egokiak, ez baitagozkio euskara batu formalari; baina ikuspegi funtzionaletik, forma dialektalak erabiltzea onargarria eta egokia da beste komunikazio-egoera batzuetan. Eta HAPen alorrean garatzen ditugun tresnak hori kontuan hartzeko gai izatea ekarpen interesgarria litzateke gure hizkuntza-komunitaterako.

Sridhar-ek (1981), Jain-ek (1984) eta Bustos-ek (1998) azpimarratu zutenez, errorea definitzean errorearen ezaugarri esplizituei (azaleko egiturari), kategoria linguistikolari eta errorea egin izanaren arrazoiari soilik erreparatu beharrean, egitura bat erroretzat jotzeko, aintzat hartzekoak dira beste hainbat faktore. Bustos-ek, esaterako, gogoeta egiten du Corder-en errore-definizioaren inguruan eta esaten du egitura oker bat errore sistematikoa den ala ez erabakitzeko, ‘errore’ horren sistematikotasun-maila balioztatu behar dela. Horretarako, hizkuntza-ikasleen corpusean detektatu diren adibideak erroreak (*konpetentzia-erroreak*) edo hutsegiteak (*performantzia-erroreak*) diren neurtu beharko genuke, bi irizpidetan oinarrituta: irizpide kuantitatiboan (aztertutako corpusean egindako errore guztien agerpen-kopurua neurtuz) eta kualitatiboan (errore kopurua, egitura zuzenen kopurua eta saihestutako egituren kopurua alderatuz).

Bestalde, erroreen definizioetan ez da argi geratzen erroreak pertsona edo talde mailakoak diren. Corder-en eta Jain-en ustetan, erroreak hizkuntza i(r)a(k)asteko estrategien araberakoak dira eta, beraz, maila bakoitzean edota tarte-hizkuntza bakoitzean errore berdintsuak errepikatu ohi dira. Baina hala izanik ere (talde mailakoak edota i(r)a(k)asteko estrategien araberakoak), Bustos-ek azpimarratzen du erroreen definizio errealista bat lortu nahi bada, erroretzat zer hartu erabakitzeko orduan ikasle bakoitzaren hizkuntza-esperientziari ere garrantzia eman behar zaiola (zein duen ama-hizkuntza edota beste zein hizkuntza dakizkien bezalako datuei). Izan ere, hizkuntza arteko eta hizkuntza barneko erroreak bereizten badira, horrek esan nahi du ama-hizkuntza bera duten ikasleek errore mota berberak egingo dituztela edota tarte-hizkuntza berdintsua izango dutela. Adierazpen hori interesgarria izan daiteke ikuspuntu didaktikotik. Horregatik, hain zuzen,

ordenagailuz lagundutako hizkuntzen i(ra)kaskuntzaren alorrean garatu dugun datu-basean ikasle bakoitzaren historia gordetzeko atala sortu dugu (III.2. atala), hizkuntzaren ikasketa-prozesua osotasunean aztertzeko informazio hori guztia kontuan hartzea ezinbestekoa delako.

Era berean, ikasketa-maila edo i(ra)kas-estrategia bakoitzari tarte-hizkuntza bat badagokio, hizkuntza-maila bakoitzari errore mota jakin batzuk dagozkiola ondoriozta daiteke, markatutako adibide erroredunak aztertuz. Baina nola definitu maila bakoitza, aldi berean maila horien bereizgarri diren erroreak antzeman ahal izateko? Gure ustez, hizkuntza-ikasleen corpusaren azterketa sakon batek lagun dezake hori argitzen; alegia, ikasleen testuak eta bertan etiketatutako erroreak hizkuntza-mailaren arabera sailkatuz eta aztertuz gero, hel gaitezke urrats bakoitzean egiten diren errore eta desbideratze usukoenak zein diren jakitera, euskararen ikasketa-prozesuaren ikuspegi orokor bat osatzera. Horretarako aukera eskaintzen digu, hain zuzen, III. kapituluan aurkeztuko dugun lan-inguruneak.

Bestalde, komunikatzeko garaian erroreak izan dezakeen larritasun-maila aintzat hartzekoa dela dio Sridhar-ek (1998); hau da, erroreak ez direla desbideratze formalak soilik eta, beraz, ikuspuntu komunikatibotik erroreak izan dezakeen eragina ere kontuan hartu beharrekoa dela. Errorearen larritasun-maila aintzat hartzekoa bada, eta maila bakoitzean errore mota jakinak egiten badira, maila jakin batean erroretzat hartzen dugunak beste maila batean ez du zertan errore izan. Adierazpen hori garrantzitsua da bai errore kontzeptuaren definizioari dagokionean, baita zuzenketa-estrategiei dagokienean ere; izan ere, egitura edo eduki bat maila jakin bati ez badagokio, egitura hori oker erabiltzea erroretzat hartu behar al da beti?; zuzendu egin behar al da beti? Horren aurrean, bat egiten dugu Bustos-ek (1998) eta Jain-ek (1984) esandakoarekin: hizkuntzen i(ra)kaskuntzaren alorrean, errore bat ezin dugu erroretzat hartu bere horretan, ikaslearen ezaugarriak, hizkuntza-maila edota baliatu dituen ikas-estrategiak bezalako faktore batzuk kontuan hartu gabe, faktore horiek guztiek erabakitzen baitute egitura oker bat errorea (*konpetentzia-errorea*) ala hutsegitea (*performantzia-errorea*) den. Egitura erroredun guztiak, beraz, ezin dira beti erroretzat hartu eta, ondorioz, erroreak zuzentzean kontuan hartu behar dira hizkuntza-maila edota idatziaren helburua bezalako faktoreak; horien arabera, errore edota desbideratze batzuk zuzendu egin beharko dira eta beste batzuk ez.

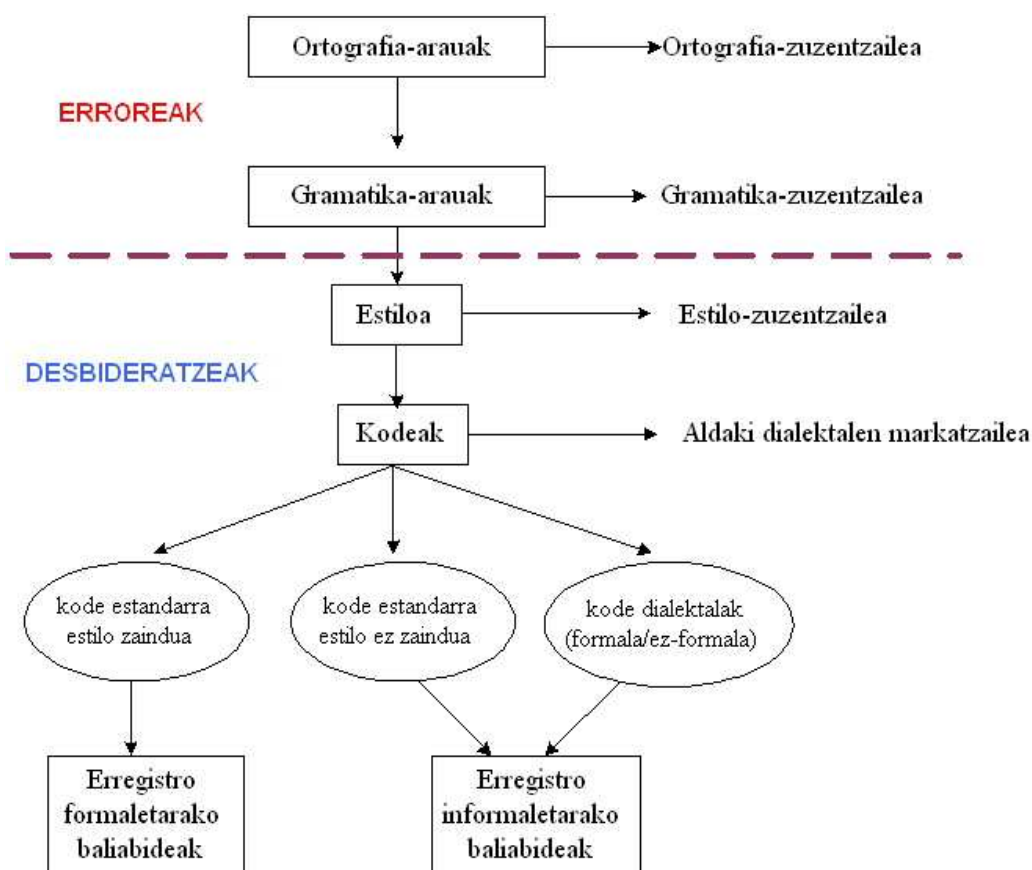
II.3.4. Gure ikuspuntua: erroreak eta desbideratzeak

Erroreen eta desbideratzeen analisisa da tesi-lan honen aztergai nagusia. Corpusetan detektatutako erroreak eta desbideratzeak dira gure azterketen oinarria, abiapuntua, bi ikerketa-alorretan ekarpenak egiteko: erroreen tratamendu automatikoaren alorrean, euskararako ortografia-, gramatika- eta estilo-zuzentzailea edota aldaki dialektalen markatzailea bezalako tresnak garatzea dugu helburu; hizkuntzen i(ra)kaskuntzaren alorrean, euskararen ikasketa-prozesua ezagutzea eta euskara ikasteko, lantzeko edota hobetzeko laguntza-tresna sendoak prestatzea da gure xede nagusia. Testuinguru honetan, ‘errore’ eta ‘desbideratze’ kontzeptuak bereizten ditugu:

- Erroretzat hartzen ditugu euskal gramatiketan zehaztutako eta Euskaltzaindiak ezarritako arauak (ortografikoak, lexikoak, gramatikalak...) betetzen ez dituzten egiturak, edozein kasutan gaizki egongo direnak. Ortografia- eta gramatika-zuzentzaileak detektatu beharreko adibideak dira, beraz, erroreak.
- Desbideratze esaten diegu gramatikaren ikuspuntutik zuzenak izan arren, testuinguru jakin batean egokiak ez diren egiturei. Desbideratzeen artean multzo hauek bereizten ditugu:
 - Sinonimoetara jo beharrean behin eta berriz errepikatzen diren hitzak edota egiturak (juntagailu baliokideak erabili ordez beti *eta* juntagailua idaztea, adibidez) eta erroreak egiteko beldurrez-edo ekidin ohi direnak. Horrelako desbideratzeak detektatzea estilo-zuzentzaileari dagokio.
 - Komunikazio-egoeraren edota eremu geografikoaren arabera gaizki erabili diren kodeak, hau da, hizkuntza estandarra dagokien idatzietan kode ez-estandarra erabiltzea edota kode dialektalei dagokien testuinguruetan eredu estandarra erabiltzea. Horrelako egituren tratamendua aldaki dialektalen markatzaileari dagokio.
 - Testuinguru komunikatibo jakin baterako desagokiak diren baliabideak, hau da, formaltasun mailak, hartzailearen hizkuntza-profilak edota komunikazio-egoeraren xedeak eskatzen duen erregistroari ez dagozkionak. Beraz,

erregistroaren egokitasuna kontuan hartzen duten baliabideak ere garatu beharko lirateke⁷, IXA taldean oraindik eremu honi ekin ez diogun arren.

Garrantzitsua da desbideratzetzat ditugun egiturak desbideratze gisa tratatzea, erroreetatik bereiz. Erroreak eta desbideratzeak bereiztuta, horiek tratatzeko tresnak ere bereizi behar dira: erroreen detekzioa ortografia- eta gramatika-zuzentzaileei dagokie, estilo kontu orokorrak estilo-zuzentzaileari, kodearen egokitasunari dagozkionak aldaki dialektalen markatzaileari eta erregistroen egokitasuna etorkizunean garatu beharko litzatekeen erregistro-markatzaileari (3. irudia).

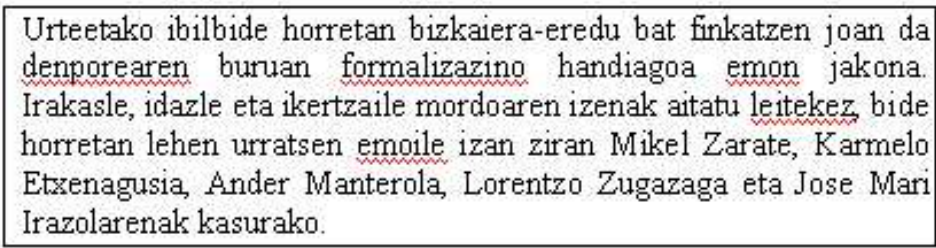


3. irudia. Erroreak eta desbideratzeak tratatzeko baliabideak.

⁷ Erregistroen markatzailea gara liteke erregistro zehatzen deskripzio linguistikoa izango bagenu eta informazio hori tresna egokietan inplementatuko bagenu; adibidez, ‘gutun komertzialen erregistro-markatzailea’, ‘txosten teknikoen erregistro-markatzailea’ edota ‘hizkuntzalaritzako artikulua espezializatuen erregistro-markatzailea’ bezalako aplikazioak diseina litezke.

Gure hizkuntza-komunitatean bizi dugun egoera kontuan izanik, erroreak eta desbideratzeak bereiz tratatzen dituzten tresnak garatzea ekarpen garrantzitsua iruditzen zaigu, HAPen alorrean eman beharreko aurrerapausoa, erabiltzaileentzat bitarteko eraginkorrak sortu nahi baldin baditugu. Esaterako, forma ez-estandarrek aldaki dialektal bezala azpimarratzen badira (eta ez errore bezala), erabiltzaileek jakingo dute forma horiek euskalkiren batean existitzen direla eta, beraz, komunikazio-egoera edota eremu geografiko batzuetan erabili ahal direla.

Idea hori du oinarrian, hain zuzen, garatze bidean den XUXENb bizkaierarako zuzentzaile ortografikoak (Garcia, 2008). Mendebaldeko euskarari dagozkion baliabide bereizgarriak hartzen ditu kontuan. XUXENb-ek hiztegi-elementuak (*berba*, *ugaraxo*, *laga...*), forma-aldakiak (*emon/eman*, *barri/berri*, *gitxi/gutxi...*), deklinabide- eta erlazio-atzizkiak (*lagunAGAZ*, *dakiALA...*) edota adizkiak (*dabe*, *zatozie*, *dakarre...*). Izan ere, XUXEN estandarra (Agirre *et al.*, 1992) mendebaldeko euskaraz idatzita dagoen testu bati aplikatzen zaionean, mendebaldeko euskaraz ongi dauden hitzak okertzat jotzen ditu, hau da, mendebaldeko euskaran zuzenak diren hitzak txartzat hartzen ditu, forma ez-estandar horiek hainbat komunikazio-egoeratan erabiltzea egokia den arren (4. irudia).



Urteetako ibilbide horretan bizkaiera-eredu bat finkatzen joan da denporearen buruan formalizazino handiagoa emon jakona. Irakasle, idazle eta ikertzaile mordoaren izenak aitatu leitekez bide horretan lehen urratsen emoile izan ziran Mikel Zarate, Karmelo Etxenagusia, Ander Manterola, Lorentzo Zugazaga eta Jose Mari Irazolarenak kasurako.

4. irudia. XUXEN-ek okertzat hartu arren, mendebaldeko euskaran zuzenak diren hitzak.

XUXENb-ren helburua, beraz, mendebaldeko euskarak dituen berezitasunak ezagutzea eta zuzentzea da, forma ez-estandarrek testuinguru jakinetan ontzat ematea. Horrela, XUXENb aplikatuz, 4. irudian erroretzat hartzen diren mendebaldeko euskarari dagozkion hitzak ez lirateke errore gisa markatuko. Bizkaierarako markatzaileak XUXEN ortografia-zuzentzaile estandarrek dituen baliabide berberak ditu, XUXEN-en bizkaierazko baliokidea baita. Horrela, testuaren izaeraren arabera, mendebaldeko euskarari edo euskara batuari dagokion

zuzentzailea erabiltzeko aukera emango dio erabiltzaileari (hots, XUXENb-n integratuta egongo da batuaz ere zuzentzeko aukera).

Adibide horrek erakusten digu aldaki dialektalen markatzaileek gure hizkuntza-komunitatean egin dezaketen ekarpena. Orain artean mendebaldeko euskararako ortografia-zuzentzailea bakarrik bideratu den arren, ekarpen eta erronka garrantzitsua litzateke gure hizkuntza-komunitatearentzat beste euskalkiei dagozkien ezaugarriak ezagutuko dituzten markatzaileak ere garatzea. Horretarako, funtsezko informazio-iturri izango dira euskal dialektologiaren alorrean egindako lanak, bai eta abian den *TSABL: Towards a Syntactic Atlas of the Basque Language*⁸ proiektuaren barruan sortutako “Euskararen Atlas Sintaktikoa” aplikazioaren bidez jasoko den informazioa ere.

Tesi-lan honetan, baina, gramatika-erroreen azterketa eta prozesamendura mugatuko gara: determinatzaile-erroreak ditugu hemen aztergai, gainerako erroreen eta desbideratzeen azterketan sartu gabe. Hala ere, prestatu ditugun bitartekoak errore eta desbideratze mota desberdinen analisia(k) egiteko diseinatuta daude.

II.4. Analisi tradizionaletik ordenagailuz lagundutako analisira

Teknologia berriek azken urteotan hartu duten indarra dela medio, Erroreen Analisi tradizionala Ordenagailuz Lagundutako Erroreen Analisiak ordezkatu du gaur egun. Hala, erroreak eta desbideratzeak aztertzeko ordenagailuak eskaintzen dizkigun abantailak eta bitartekoak baliatzen ditugu: euskarri elektronikoan jaso eta antolatutako corpusak, erroreak edota desbideratzeak etiketatzeko editoreak eta sailkapenak, etiketatutako adibideak eta horien inguruko informazioa gordetzeko datu-baseak... Tresna informatikoen ikerketak azkarrago, errazago eta zehatzago egiteko aukerak eskaintzen dizkigute (Dagneaux *et al.*, 1998).

Erroreen analisi tradizionalaren eta ordenagailuz lagundutako erroreen analisiaren helburuak edota aztergaiak berdinak dira: ikasle bakoitzak ikasten ari den

⁸ <http://www.iker.cnrs.fr/-tsabl-towards-a-syntactic-atlas-of-.html?lang=fr>

hizkuntza zenbateraino menderatzen duen ikustea, hizkuntza berri bat nola ikasten den ezagutzea, hizkuntza bat ikasteen ikasleek aurkitzen dituzten zailtasun nagusiak zein diren ikertzea, etab. Horretarako, testuetan detektatutako erroreak edota desbideratzeak aztertzen dira. Eta ordenagailuak erroreen analisiaren alorrean ikertzeko bide eta aukera berriak eskaini dizkigu: ikerketak bizkortu, erraztu eta hobeto kontrolatu ahal dira; ikasleen tarte-hizkuntzari dagozkion datuak behar bezala antolatzeko edota gordetzeko bitartekoak eskura ditugu; egitura jakin baten erabilera okerrak eta zuzenak modu errazean alderatzeko aukera daukagu; interesatzen zaizkigun aldagaiak kontrolatzeko ahalmena daukagu; ikasle bakoitzaren maila eta beharren arabera materialak edota laguntza-tresnak garatzeko eta egokitze bideak ditugu. Datuak ondo antolatuta izanik, lortuko ditugun emaitzak ere zehatzagoak izango dira, eta gordetako informazioa motibazio aplikatu desberdinekin baliatu ahalko da hainbat ikerketa-alorretan.

Bestalde, analisi tradizionalari egotzitako hainbat muga gainditzea lortzen dute ordenagailuz lagundutako azterketek (Díaz-Negrillo eta Fernández-Domínguez, 2006). Esaterako, aditu askok kritikatu izan dute Erroreen Analisisia ikasleen erroreak bakarrik aztertzea mugatzen dela, desbideratzeak edota egitura zuzenak kontuan hartu gabe. Ordenagailuaren bidez, erroreekin batera ekoizpen zuzenak ere aztertzeko aukera daukagu, baita egitura gramatikalak eta ez-gramatikalak modu errazean alderatzeko aukera ere: testuak euskarri elektronikoan izanik, egitura jakin bat hautatu eta aldera ditzakegu bere erabilera zuzenak eta okerrak (KWIC⁹ sistemaren bidez, esaterako, erraza da horrelako zerrendak automatikoki eskuratzea). Horrela, egitura horren inguruan ikasleak ikasi duena eta ikasi ez duena hobeto ezagut daiteke.

Erroreen Analisari kritika egin zaio, halaber, bigarren hizkuntzaren i(r)a(k)askuntzaren ikuspegi estatikoa eskaintzen duelako. Ordenagailuak, ordea, aukera ematen digu modu erraz eta bizkorrean maila desberdinetako hizkuntza-ikasleen testuak konparatzeko, maila bereko hainbat ikasleren testuak alderatzeko, ikasle bakar batek maila desberdinetan idatzi dituen testuak erkatzeko,

⁹ KWIC: Key Word In Context. Hitz-zerrendak testuinguruan ikusteko aplikazioa.

etab. Izan ere, testuak euskarri elektronikoan izanik, hainbat software balia daitezke hitz, esaldi edota egitura jakin batzuk aztertzeko: adibidez, *nahi* aditzaren erabilera, zuzena eta okerra, hizkuntza-maila desberdinetan eta ikasleka azter daiteke modu errazean; edota atzizki jakin baten erabilera (*-ra*, esaterako) ere aztertu ahalko genuke atzizki hori duten hitzen zerrendak erauzi eta maiztasunak neurtuz, atzizkiaren erabilera testuinguruan ikusiz, etab. Era berean, erroreekin batera desbideratzeen edota egitura zuzenen azterketa modu errazean egitea ere posible da ordenagailuaren laguntzaz; eta ikasleei dagokien informazio psikolinguistikoa edota testuen inguruko datuak ondo antolatuta biltzeko aukerak eskaintzen dizkigu, gainera, ordenagailuak.

Erroreak sailkatzeko baliatzen diren kategoria-sistemak, oro har, ondo definitu gabeak, nahasgarriak eta subjektiboak izan ohi direla egotzi izan zaio, halaber, analisi tradizionalari, errore mota batzuei zein kategoria dagokion erabakitzea ez dela erraza gertatzen, errore batzuk kategoria batean baino gehiagotan sailka daitezkeela, etab. Ordenagailuz lagundutako analisisetan ere zailtasun berberak aurkitzen ditugu erroreen sailkapen bat garatzean. Baina bada abantaila edo desberdintasun bat: ordenagailuari esker, datu guztiak euskarri informatikoan ditugu gordeta (datu-base batean, hain zuzen) eta ondorioz, sailkapena osatzen duten kategoriak eta azpikategoriak aldatzea, gehitzea edota kentzea ez da zaila gertatzen. Izan ere, litekeena da sailkapena sortzean aurreikusitako kategoria edota azpikategoria guztiak corpusetan detektatutako adibide errealekin bat ez etortzea. Horregatik, abantaila handia da sailkapenak modu errazean aldatzeko aukera izatea.

Analisi tradizionalaren aldean, beraz, ordenagailuz lagundutako analisisiek aurrerapen garrantzitsuak ekarri dizkigute ikusmiran ditugun ikerketa-alorretan ekarpenak egiten jarraitu ahal izateko. Azken urteotan garatu diren tresna informatiko edota softwareek erraztu egin dizkigute, nabariki, eskuz nekez egin ahalko genituzkeen azterketak. Egun, bada, *erroreen analisisiaz* hitz egiten dugunean, ordenagailuz lagundutako erroreen analisisiaz ari garela ulertu behar da.

Tesi-lan honen ekarpenetako bat corpusetan detektatzen diren erroreen eta desbideratzeen analisirako lan-ingurune bat garatzea izan da, bi ikerketa-alor uztartuz: tratamendu automatikoa eta ordenagailuz lagundutako hizkuntzen i(r)a(kaskuntza).

II.5. Laburbilpena

Bigarren hizkuntzen glotodidaktika edo i(ra)kaskuntza-prozesua aztergai izan dute hiru ikerketa-lerro nagusik: Analisi Kontrastiboa izan zen lehena (1945-1967), Fries eta Lado aitzindari zituela. Analisi Kontrastiboaren helburua ama-hizkuntzaren eta xede-hizkuntzaren arteko alderaketa sistematikoa egitea zen, hizkuntza-ikasleen ikasketa prozesua ezagutu eta prozesu horretan esku hartu ahal izateko. 1967an Corder-ek *The Significance of Learners' Errors* argitaratu zuenean, Erroreen Analisia Analisi Kontrastiboari lekua hartzen hasi zitzaion, hizkuntza-ikasleek xede-hizkuntzan egindako erroreen azterketa helburu zuela eta erroreak hizkuntzaren ikasketa-prozesuaren funtsezko adierazle zirela aldarrikatuz. Hasieran kritika ugari jaso zituen arren, indarra eta sinesgarritasuna irabazten hasi zen teoria hori erroreen inguruan egindako ikerketetan lortutako emaitza positiboek esker. 1972tik aurrera, berriz, Tarte-hizkuntzaren kontzeptua indarra hartzen hasi zen Selinker-i esker: hizkuntza-ikaslearen ama-hizkuntzaren eta xede-hizkuntzaren arteko sistema linguistiko autonomo eta aldakorra da tarte-hizkuntza, eta egiten diren urrats horien azterketa ezinbestekoa da hizkuntza baten ikasketa-prozesua osotasunean ezagutu ahal izateko. Oro har, hiru teoria edo ikuspuntuak izan dira kritikatuak; baina hirurek utzi dizkigute, era berean, ekarpen interesgarriak.

Erroreen Analisia hainbat alorretan jarraitzen den ikerlerroa da. Gaur egun, jada, ordenagailuz lagundutako analisiak egiten dira, ordenagailuak eskaintzen dizkigun abantailak etekina atereaz: azterketa bizkorragoak, eraginkorragoak eta kontrolatuagoak egiteko aukera ematen digu ordenagailuak.

Erroreen analisiak errore kontzeptua definitzea eskatzen du, erroretzat zer hartuko den eta erroreen aurrean nola jokatu den erabakitzea. Guk 'erroreak' eta 'desbideratzeak' bereizten ditugu. Errore esaten diegu euskara estandarrari dagozkion gramatika-arauak betetzen ez dituzten egitura erroredunei; desbideratzetzat hartzen ditugu, berriz, gramatikaren ikuspegitik zuzenak izan arren testuinguru jakin batean desegokiak diren egiturak: behin eta berriz errepikatzen direnak, eremu geografiko bati ez dagozkionak, komunikazio-egoera baterako desegokiak direnak...

HAPen alorrean, erroreak eta desbideratzeak bereiz tratatzen dituzten tresnak garatzea ekarpen garrantzitsua iruditzen zaigu, gure hizkuntza-komunitatean bizi

dugun egoera soziolinguistikoa dela eta. Izan ere, euskara normalizazio-prozesuan dago 1968an hizkuntza idatziaren batasuna bideratzeko ahaleginak hasi zirenetik eta gure hizkuntza-komunitatean hainbat dialekto hitz egiten dira. Egoera horren aurrean, euskalkien banaketa hizkuntzaren erabilera funtzionalak eta eremu geografikoak baldintzatu behar dutenez, hizkuntza ez-estandarrari dagozkion hitzak edota egiturak erabiltzea onargarria da testuinguru batzuetan, baina ez beste batzuetan: forma ez-estandarrak edota estilo zaindugabeari dagozkionak erroretzat hartu behar dira erregistro formala eskatzen duten komunikazio-egoeretan; ez, ordea, erregistro informala eskatzen dutenetan; eta alderantziz, erregistro informala erabili behar den kasuetan kode estandarrari edota estilo zainduari dagozkion zenbait baliabide erabiltzea ez da egokia, baina komunikazio-egoera formalean ezinbestekoak dira baliabide estandarrak. Beraz, erroreak eta desbideratzeak bereizi eta desberdin tratatzeko gai diren sistemak garatzea interesgarria bezain baliagarria izango da Euskal Herriko egoera soziolinguistiko errealekin bat egin eta hizkuntza-komunitatearen beharrak asetzeko ekarpen bat egingo badugu.

Erroreen edota desbideratzeen ikerketa sistematikoa bi ikerketa-alorretan baliatzen dugu guk. Erroreen tratamendu automatikoaren alorrean, euskararako ortografia-, gramatika- eta estilo-zuzentzailea edota aldaki dialektalen markatzailea bezalako tresnak garatzea dugu helburu. Horretarako, erroreen eta desbideratzeen deskripzio formalean oinarritzen gara: informazio linguistikoa eta teknikoa zehazten dugu. Halaber, erroreak eta desbideratzeak bereiz tratatzea garrantzitsua dela esan dugu: ortografia- eta gramatika-zuzentzaileak detektatuko dituzte erroreak; eta estilo-zuzentzaileak edota aldaki dialektalen markatzaileak, berriz, desbideratzeak. Tresna horiei esker, erabiltzaileak erroreak eta desbideratzeak modu diferenteetan markatuta ikusiko ditu: ortografia- eta gramatika-zuzentzaileak detektatutako erroreak errore dira testuinguru guztietan; estilo-zuzentzaileak markatutako desbideratzeak gramatikaren ikuspegitik zuzenak diren arren, ez dira euskara estandarrean egokitzen hartzen; eta aldaki dialektalen markatzaileak detektatutako adibideak testuinguru jakin batzuetan egokiak ez diren arren, onargarriak dira beste komunikazio-egoera edota eremu geografiko batzuetan.

Hizkuntzen i(r)a)kaskuntzaren alorrean, berriz, hizkuntza-erroreen inguruan ikusmolde berri bat garatu da: gaur egun erroreak hizkuntzen ikaskuntza-prozesuaren erakusle nagusitzat ditugu, ikasle bakoitzak ikaskuntza-prozesu horretan jarraitzen

dituen urrats desberdinen adierazle dira. Testuetan detektatutako erroreen edota desbideratzeen diagnosi linguistikoa eginez, ikasleek xede-hizkuntza zenbateraino menderatzen duten jakitera irits gaitezke, hizkuntza berria nola ikasten ari diren ezagutzera, hizkuntza ikastea aurkitzen dituzten zailtasun batzuk atzematera. Eta horren arabera, xede-hizkuntza lantzeko tresnak presta daitezke.

Euskarazko erroreen eta desbideratzeen analisiaren motibazio aplikatua, beraz, bikoitza da. Eta ikusmiran ditugun bi ikerketa-lerroak uztartzeko aukera ematen digu, hain zuzen, euskarazko erroreen eta desbideratzeen analisirako garatu dugun lan-inguruneak (III. kapitulua), batean egiten den lana bestean aprobetxatuz, eta alderantziz.

Euskarazko erroreak eta desbideratzeak aztertzeke garatu dugun lan-ingurunea ekarpen garrantzitsua da; errorearen tratamendu automatikoaren eta ordenagailuz lagundutako hizkuntzaren i(ra)kaskuntzaren alorretan azterketak egiteko funtsezkoak zaizkigun baliabideak bildu ditugu bertan.

III. Erroreak eta desbideratzeak aztertzeke lan-ingurunea

III.1. Sarrera

Erroreen eta desbideratzeen tratamendu automatikoa Hizkuntzaren Azterketa eta Prozesamendua (HAP) esaten zaion ikerketa-alorraren barruan kokatzen da. Hizkuntzaren tratamendu automatikoz arduratzen den adarra da HAP, hizkuntza aztertu eta ordenagailuak ulertzeke moduan formalizatzen duena, ikuspegi konputazionaletik giza hizkuntza modu logikoan modelatzen saiatzen dena. HAPek “*hizkuntza hartzen du gai eta ordenadorea lanabes*” (Aduriz *et al.*, 1994). Eta horretarako ezinbestekoa da hainbat ikerketa-alorretako adituen elkarlana, diziplinartekotasuna.

50.eko hamarkadan Amerikako Estatu Batuetan sortu zen HAP helburu jakin bat ikusmiran zuela: atzerriko hizkuntzetatik ingelesera itzultzeke programak egitea. HAPen oinarrian, beraz, itzulpen automatikoaren ideia zegoen. Gai horietan aurrendari izan ziren adituak berehala ohartu ziren, ordea, testuak automatikoki itzuli ahal izateke ezinbestekoa zela hizkuntza ulertzea eta prozesatzea, perpaus- eta morfologia-mailan behinik behin; eta sintaxia ulertzeke, berriz, diskurtsoaren semantika eta hizkuntzaren pragmatika ulertu behar zirela. Oinarritzko aztergai nagusi asko definitu beharra zegoen, beraz: ezagutza linguistikoa konputagailuan nola adierazi, ezagutza hori nola erabili, ezagutza linguistikoa eta hizketa-gaiari dagokion ezagutza programa informatikoetan nola uztartu, azterketa eta tratamendu linguistikoa nola banatu modulu sinple eta independenteetan (horietako bakoitza eraginkorra izan zedin), etab. Urte batzuk beranduago, 60.eko hamarkadan, Adimen Artifiziala sortu eta HAP beronen atal edo

aztertze-eremu bilakatu zen. Testuak itzultzeko asmo bezala hasi zen ikerketa-alor hori, beraz, giza hizkuntza aztertze eta prozesatzeko arlo bat izatera pasa zen. Helburu horrekin, HAPen barruan askotariko ikerketak egin dira eta aplikazio ugari sortu dira hainbat hizkuntzatan: analizatzaile morfologiko, sintaktiko eta semantikoak, corpus eta datu-base linguistikoak, ortografia-, gramatika- eta estilo-zuzentzaileak, itzultzaile automatikoak, etab. Horretarako, hizkuntzaren funtzionamendua maila guztietan aztertu eta uztartu beharra dago: lexikoa, morfologia, sintaxia, semantika, pragmatika, fonetika...

Bestalde, azken urteotan teknologia berriek erakutsi duten hazkunde geldiezinak eta zabalkuntzak ezinbesteko egin dute ordenagailuen eta bestelako gailu teknologikoen erabilera herrialde aurreratuetan. Horren ondorioz, HAPek berebiziko garrantzia hartu du, aurrerapen handiak ekarri ditu eta alor askotaraino iritsi da: irakaskuntza, itzulpengintza, medikuntza, psikologia, soziologia, ekonomia, etab.

Orain dela hogeitaz bat urte ekin zion IXA taldeak euskararen azterketa eta prozesamenduari. Giza hizkuntzaren azterketa, ulermena, sormena, galderei erantzutea, informazioa erauztea, erroreen tratamendua, hizkuntzen i(ra)kaskuntza edota itzulpengintza automatikoa dira, besteak beste, IXA taldearen helburu nagusiak. Hala, euskararen azterketa eta prozesamenduaren alorrean ikerketa ugari bideratu eta hainbat ekarpen egin ditugu orain artean; eta lanean jarraitzen dugu beste hizkuntza batzuetan garatu dituzten baliabideak euskararako ere garatuko ditugun itxaropenarekin.

Erroreen eta desbideratzeen tratamendu automatikoa ezinbesteko urratsa da HAP osotasunean egin nahi baldin bada. Izan ere, testu idatzietan, gutxi edo gehiago, beti egin ohi ditugu erroreak, hiztunok eta hizkuntza-ikasleek, konturatu gabe edota ezjakintasunaren eraginez. Euskaraz idaztean hanka sartzea, gainera, ez da hain zaila, euskarak bizi duen egoera soziolinguistikoa kontuan hartuta: gutxi batzuen hizkuntza da, hizkuntza-komunitate elebidun batean bizi da indartsuagoak diren erdaren artean eta estandarizazio-prozesuan murgilduta dago. Erroreen eta desbideratzeen azterketa eta prozesamenduak testuinguru horretan egin ditzakeen ekarpenak garrantzitsuak izango direla uste dugu.

Kapitulu honetan euskarazko erroreak eta desbideratzeak aztertze garatu dugun lan-ingurunea aurkeztuko dugu (III.2. atala) eta lan-ingurunea osatzen duten bitarteko batzuk oinarri hartuta euskara ikasteko edota lantzeko sortu ditugun bi aplikazioen

berri emango dugu (III.3. atala), aztergai ditugun bi ikerketa-alorretan egin daitezkeen ekarpenen adibide gisa.

III.2. Euskarazko erroreak eta desbideratzeak aztertzeke lan-ingurunea

Euskarazko erroreen eta desbideratzeen azterketa eta prozesamendua bideratzeko lan-ingurune oso bat garatu dugu, bi ikerketa-alorretan ekarpenak egin ahal izateko:

- erroreen tratamendu automatikoaren alorrean, euskararako ortografia-, gramatika- eta estilo-zuzentzailea edota aldaki dialektalen markatzailea bezalako tresnak garatzea da gure xede nagusietako bat;
- ordenagailuz lagundutako euskararen i(ra)kaskuntzaren alorrean, hizkuntzaren i(ra)kaskuntza prozesua ezagutzen edota bideratzen laguntzeko hainbat tresna prestatu nahi ditugu.

Alor batean zein bestean ezinbestekoa zaigu euskarazko erroreen eta desbideratzeen analisi sakonak egitea, eta horretarako aukera eskaintzen digun lan-ingurune bat osatu dugu, baliabide nagusi hauekin: corpusak, EtikErro erre-editorea, euskarazko erroreen eta desbideratzeen sailkapena eta *Erroreak* eta *Ikasleak* datu-baseak (Aldabe *et al.*, 2006; Aldabe *et al.*, 2007a; Aldabe *et al.*, 2007c).

Erroreen eta desbideratzeen analisiaren abiapuntua hizkuntza-ikasleen corpusa izan ohi da, testu-bilduma horietan egoten baita, normalean, adibide gehien. Hizkuntza-ikasleek egindako erroreen analisiaren alorrean ikerketa eta ekarpen ugari egin dira azken urteotan hainbat hizkuntzatan, denetarik helburuekin (Milton eta Chowdhury, 1994; Granger, 1998a; Yang eta Akahori, 1998; Tono, 2003; Martín del Rey, 2004; Yang Huizhong eta Gui Shichun, 2005, besteak beste). Tono-k (2003) hizkuntza-ikasleen corpusak baliatu ohi dituzten ikerketa-alor hauek bereizten ditu: erroreak etiketatzeke editoreen garapena, hizkuntza-ikasleen testuetan egiten diren erroreen analisia, jatorrizko hitzen eta hizkuntza-ikasleen arteko desberdintasunak alderatzeko azterketak, hainbat ama-hizkuntzatako ikasleen tarte-hizkuntza(k) aztertzeke eta konparatzeko ikerketak eta hizkuntza-ikasleen testuetan oinarrituta material didaktikoak diseinatzeko egiten diren analisiak. Baina corpusak baliagarriak izango badira, ezinbestekoa da bertan aurkitzen diren adibideak etiketatzea.

Etiketatzeko-lana garestia eta astuna denez, lan hori errazten duten softwareak, editoreak edota laguntza-tresnak baliatu ohi dira gaur egun.

Erroreak etiketatzeko editorea

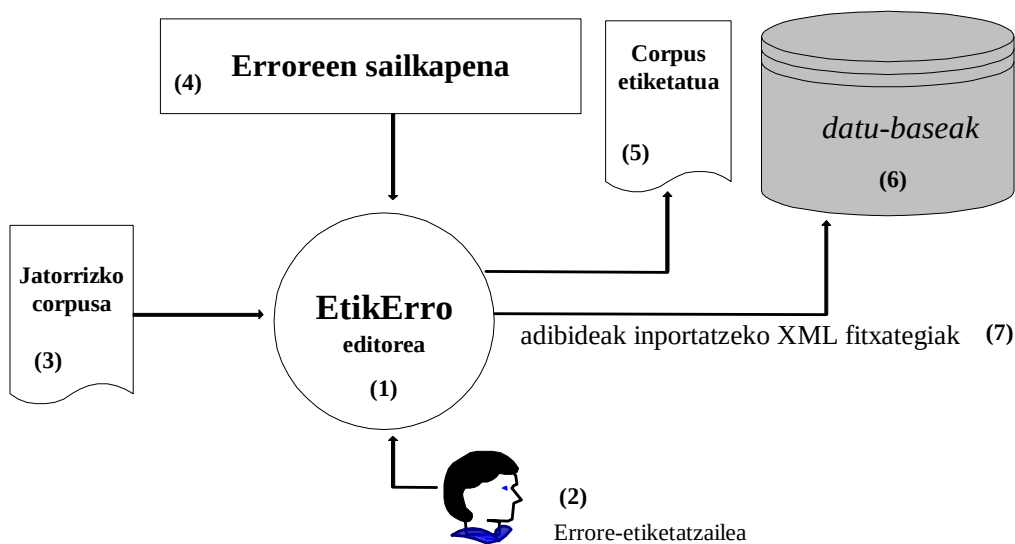
Granger-en ikertaldeak, esaterako, azterketa ugari egin ditu hizkuntza-ikasleen corpusen analisiaren alorrean. Beste lan askoren artean, testuetan erroreak etiketatzeko editore bat garatu dute: UCLEE (the *Université Catholique de Louvain Error Editor*) (Hutchinson, 1996; Dagneaux *et al.*, 1996; Dagneaux *et al.*, 1998). Editore horrek erroreen sailkapen bat du oinarrian eta sailkapenaosatzen duten kategoriei edota azpikategoriei dagozkien kodeekin etiketatzen dituzte erroreak, errore-kode bakoitza `< >` ikurren barruan zehaztuz (5. irudia). Erroreen sailkapenean hiru maila bereizten dituzte: errore mota (`<G>`, *grammar*), errore-kategoria (`<NBR>`, *number*) eta gramatika-kategoria (`<VSC>`, *finite simple*):

Ces gens <code><G><NBR><VSC></code> #pensent\$ pense <code></VSC></NBR></G></code> aussi que les femmes ne sont pas aptes à prendre des responsabilités.

5. irudia. UCLEE errore-editorearekin etiketatutako adibidea.

Behin errore-kodeak corpusean txertatuta, WordSmith softwarea (Scott, 1996) baliatzen dute erroreen inguruko informazioa eratzeko. Datu kuantitatiboak eskuratzeko, estatistikak ateratzeko edota errore mota jakinak testuinguruan ikusteko aukerak eskaintzen ditu tresna horrek: esaterako, `<G>` edota `<NBR>` edota `<VSC>` etiketa duten agerpen guztien datuak eta adibideak eskuratzeko aukera; eta etiketatutako errore batzuen eta besteen konbinazioei dagozkien datuak lortzea ere posible da. Testu etiketatuak, beraz, informazio-iturri aberatsak dira eta WordSmith bezalako tresnek errore-adibideen inguruko emaitza zehatzak eskuratzeko eta ondorioak ateratzeko aukera eskaintzen dute.

UCLEE editoreaz aparte, antzeko ezaugarriak dituzten errore-editore gehiago ere garatu dira: the TOSCA-ICLE tagger (de Haan, 2000), the TagEditor (Izumi *et al.*, 2003), Yang Huizhong eta Gui Shichun-en editorea (2005); Episimiotis testu-etiketatzaila (Koutsis *et al.*, 2007) edota Jaengo Unibertsitateko errore-editorea (Díaz-Negrillo eta García-Cumbreras, 2007), besteak beste. Editore horiek ere errore-sailkapenak dituzte oinarrian eta WordSmith bezalako tresnen bidez aztertzen



dago(z)kion zuzenketa(k); eta (5) etiketatutako adibideari euskarazko erroreen eta desbideratzeen sailkapenaren arabera dagokion kategoria edo azpikategoria (7. irudia). Informazio hori ezinbestekoa da testuetan aurkitutako adibideen diagnosi linguistikoa egiteko.

- (1) Testuaren erref.: il95_9_t_e
- (2) Testuaren inguruko informazioa:
 - Hitz-kopurua: 245
 - Testu-mota: gutuna
 - Erregistroa: formala
- (3) Adibidea: **autoa bat* erosi *dugu*
- (4) Zuzenketa(k): *auto **bat** erosi *dugu**
*autoa erosi *dugu**
- (5) **Kategoria/azpikategoria:**
DETerminatzaile Errepikatzea <DETE>
 1. kategoria: gramatika-errorea
 2. kategoria: determinatzaile-errorea
 3. kategoria: determinatzailea errepikatua

7. irudia. EtikErro bidez datu-basera inportatzen den informazio linguistikoa.

Adibide erroredunei edota desbideratzeei etiketak esleituz corpusak etiketatzeko tresna da, beraz, EtikErro. Baina era berean, lan-ingurune batean integratuta egoteak aukera ematen digu testuetan etiketatutako adibide horiei informazio osagarria (datu teknikoak edota psikolinguistikoak) gehitzeko. Horrela, editorearekin etiketatutako adibideen inguruko hainbat datu gordetzen ditugu, baliagarriak izango zaizkigunak bai euskarazko errorean eta desbideratzeen tratamendu automatikoa bideratzeko eta bai euskararen i(ra)kaskuntzarako laguntza-tresnak garatzeko. Eta horixe da, hain zuzen, EtikErro editoreak erroreak etiketatzeko garatu diren beste editoreekiko eskaintzen digun desberdintasun nagusi bat.

Erroreak eta Ikasleak datu-baseak

Testuetan etiketatutako erroreak eta desbideratzeak bi ikerketa-alorretan ekarpenak egiteko baliatzen ditugu guk: errorearen tratamendu automatikoan eta ordenagailuz lagundutako hizkuntzen i(ra)kaskuntzan. Alor batean eta bestean helburuak desberdinak direnez, bi datu-base sortu ditugu: *Erroreak* eta *Ikasleak*. Datu-base horietan testuetan etiketatutako adibideei dagokien informazio linguistiko bera gordetzen dugu; hau da, bi datu-baseek errorearen diagnosi linguistikoa egiteko behar den

informazioa partekatzen dute. Informazio linguistikoaz gain, baina, datu-base bakoitzean bestelako datu batzuk gordetzen ditugu:

- *Erroreak* datu-basearen helburua errearen eta desbideratzeen inguruko informazio linguistikoa eta teknikoa gordetzea da; hau da, datu-base honetan 7. irudian zehaztutako informazio linguistikoarekin batera, errore edo desbideratze bakoitza automatikoki detektatzeko balia daitezkeen tresnen edota tekniken inguruko informazioa jasotzen da: determinatzaile-erroreak, esaterako, Murritzapen Gramatika (MG) formalismoan oinarritutako erregela bidez detektatzen dira (8. irudia).

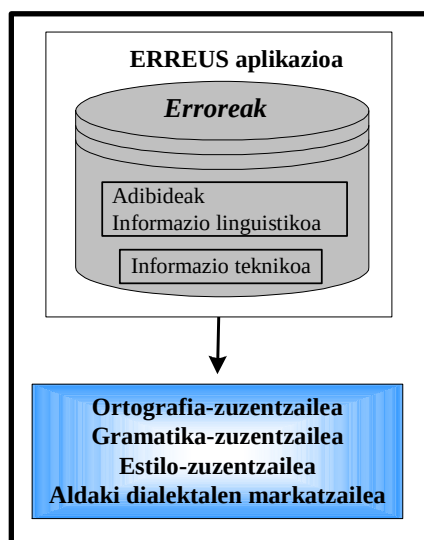
Adibidea: **autoa bat* erosi dugu

Informazio teknikoa:

- Errearen deskribapena: determinatzailea errepikatu da sintagman (**DETE**)
- Detektatzen da? Bai
- Detektatu duen tresna: **MG**
- Zuzendu da? Ez
- Zuzendu duen tresna: -

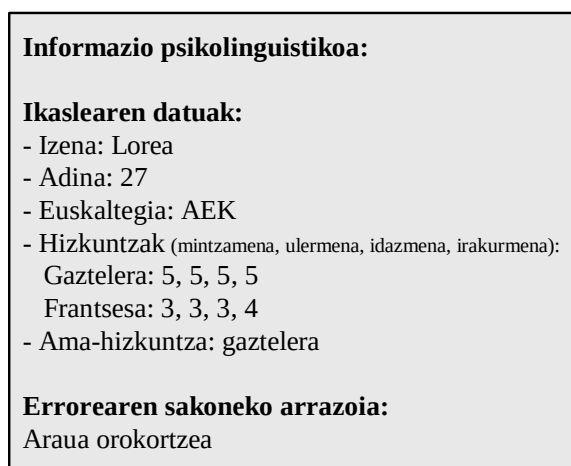
8. irudia. *Erroreak* datu-basean gordetzen den informazio teknikoa.

Erroreak datu-basean, beraz, erreari eta desbideratzei dagokien informazio linguistikoa eta teknikoa gordetzen dugu, ortografia-, gramatika- eta estilo-zuzentzailea edota aldaki dialektalen markatzailea bezalako tresnak garatu ahal izateko. *Erroreak* datu-baseko informazioa ERREUS (Aldabe *et al.*, 2005a) web aplikazioaren bidez eskura daiteke (9. irudia):



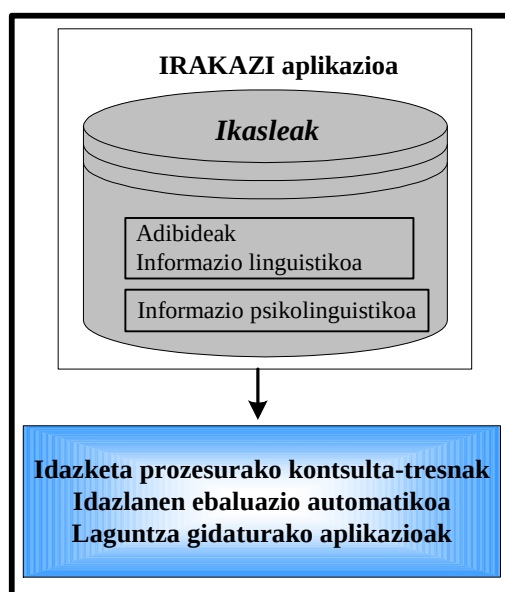
9. irudia. *Erroreak* datu-basearen ingurunea.

- *Ikasleak* datu-basearen helburua da hizkuntzaren ikasketa prozesua ezagutzeko eta ordenagailuz lagundutako hizkuntzen i(r)a kaskuntza bideratzen laguntzeko tresnak prestatzeko behar den informazioa biltegitratzea. Horretarako, datu-basean erroreei eta desbideratzei dagokien informazio linguistikoarekin batera, datu psikolinguistikoak gordetzen ditugu (10. irudia), bi ataletan:
 - o lehenengo atalean biltzen ditugu ikasle bakoitzaren datu pertsonalak (izena, adina, ama-hizkuntza, hizkuntza-eskola, maila, taldea, irakaslea, sarrera-data...), ikasleak hizkuntzen ikasketa-prozesuan egin duen ibilbidea, ikasleak hitz egiten dituen hizkuntzak eta hizkuntza bakoitzean duen maila (idazketa, ulermena, irakurketa eta hizketa kontuan hartuz);
 - o bigarren atalean errorea edota desbideratzea eragin ahal izan duten sakoneko arrazoiak jasotzen ditugu, hau da, ikasleek egitura linguistikoak modu batean edo bestean erabiltzeko egon daitezkeen arrazoiak. Arrazoi edo kausa posibleen artean, bi atal nagusi bereizten dira (Maritxalar, 1999): barruko faktoreei edo faktore psikologikoei dagokiena (ikas-estrategiek, komunikazio-estrategiek edota faktore fisiologikoek eragindakoa) eta inguruneari lotutako faktoreak (irakaskuntza-transferentziak, komunikazio-egoerak, faktore soziokulturalak, etab.).



10. irudia. *Ikasleak* datu-basean gordetzen den informazio psikolinguistikoa.

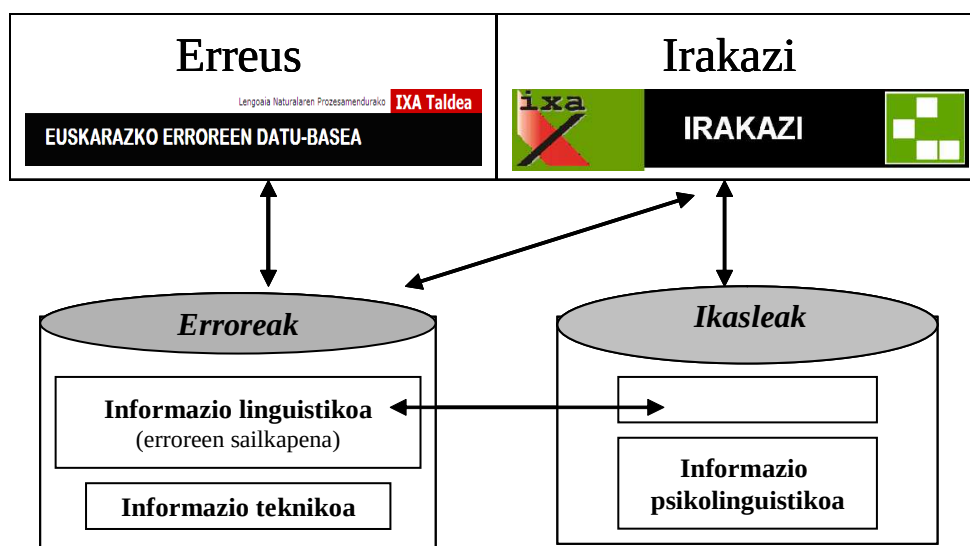
Hizkuntza-ikasleen testuak aztertuz (zein errore edota desbideratze mota egiten dituzten gehien, zein egitura ekiditen dituzten, dituzten zailtasun ohikoenak zein diren, etab.), ikertzaileok hizkuntzen i(r)a(k)askuntza prozesuan parte har dezakegu ikasleen beharren eta zailtasunen araberrako laguntza-tresnak garatuz, ikas-estrategia egokiak proposatuz, erroreak zuzentzeko estrategiak finkatuz, etab. Hizkuntza i(r)a(k)asteko, lantzeko edota hobetzeko laguntza-tresna informatikoak garatzeko, ordea, ezinbestekoa da etiketatutako adibideei dagokien informazio linguistikoarekin batera, datu psikolinguistikoak ere kontuan hartzea. Informazio hori guztia IRAKAZI (Aldabe *et al.*, 2005b) web aplikazioaren bidez eskura daiteke (11. irudia):



11. irudia. *Ikasleak* datu-basearen ingurunea.

Datu-base bakoitzari, beraz, informazio mota jakin bat dagokio: *Erroreak* datu-basean informazio teknikoak gordetzen da eta *Ikasleak* datu-basean informazio psikolinguistikoa. Era berean, bi datu-baseek informazio linguistiko berdina partekatzen dute, erroreen eta desbideratzeen diagnosi linguistikoa egiteko oinarritzat hartzen dena.

Datu-base horietan adibideak gordetzeko bi bide daude: EtikErro editorearen bitartez¹⁰ corpusak etiketatzen direnean gordetzea, edota eskuz, zuzenean, corpusak etiketatu gabe gordetzea. Datuak eskuz sartzeko, bi web aplikazio baliatzen ditugu (Arrieta *et al.*, 2003): ERREUS web aplikazioak *Erroreak* datu-basea du oinarrian eta IRAKAZI aplikazioak, berriz, *Ikasleak* datu-basea. Baina IRAKAZI aplikazioak *Erroreak* datu-basea ere baliatzen du, detektatutako adibideen diagnosi linguistikoa egiteko oinarri hartzen den sailkapena *Erroreak* datu-basearen diseinuan integratuta baitago. Hau da, IRAKAZI aplikazioaren interfazetik *Ikasleak* datu-basean adibide berri bat sartzean, adibide hori eta bere deskripzio linguistikoari dagokion informazioa *Erroreak* datu-basean gordetzen da berez; eta *Ikasleak* datu-basean ikasleari buruzko informazio psikolinguistikoa bakarrik jasotzen da (12. irudia):



12. irudia. *Erroreak* eta *Ikasleak* datu-baseen lotura.

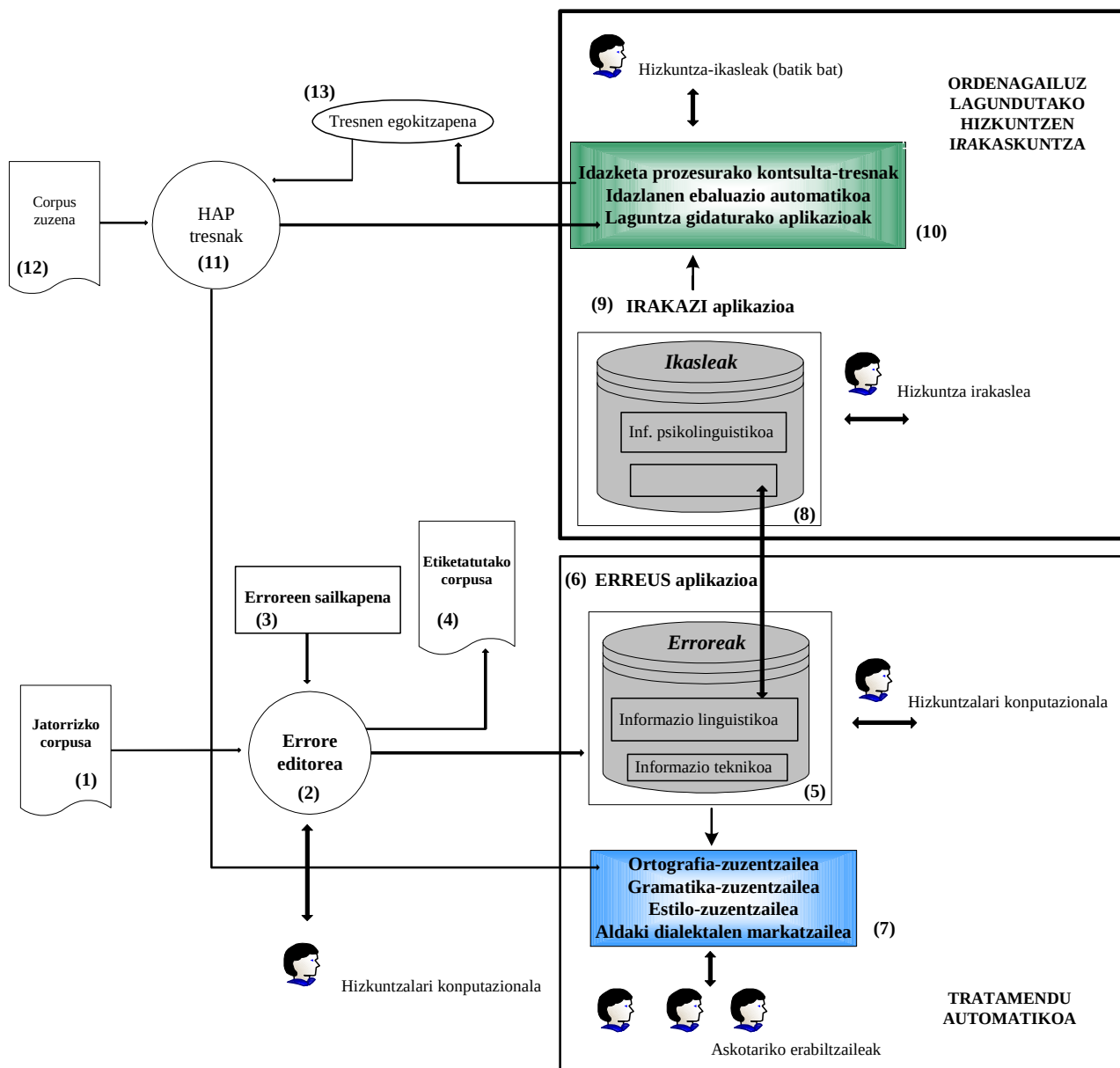
¹⁰ Errore-editorearen bidez erroreen edota desbideratzeen inguruko informazio linguistikoa bakarrik inportatzen da datu-basera; adibide horiei dagozkien gainerako datuak (datu teknikoak edota psikolinguistikoak) eskuz gehitu behar dira.

Bi datu-baseen arteko loturari esker, beraz, informazio linguistikoari dagokion atala partekatzen dute: ERREUS aplikaziotik sartzen diren adibideak eta horiei dagozkien datu linguistikoak *Erroreak* datu-basean gordetzen dira, informazio teknikoarekin batera. IRAKAZItik sartzen diren adibideak eta horien inguruko datu linguistikoak ere *Erroreak* datu-basean gordetzen dira, eta adibide horiei dagokien informazio psikolinguistikoa, aldiz, *Ikasleak* datu-basean.

Beste aukera bat izan zitekeen datu-base bakoitzak erroreen eta desbideratzeen diagnosi linguistikorako bere sailkapena izatea, hau da, bi datu-baseetan sailkapen berbera kokatzea. Sailkapen berdina bi tokitan gorde eta mantentzea baino egokiagoa iruditu zaigu, ordea, bi datu-baseek baliatuko duten sailkapena datu-base bakarrean gordetzeko aukera izatea. Horrela, biltegitratutako adibideak eta horien inguruko datu linguistikoak bi leku desberdinetan gorde beharra aurrezten dugu, eta lana errazten dugu. Izan ere, datu-base bakoitzean sailkapen bat izango bagenu, bietako bat aldatu beharko litzatekeen bakoitzean eguneratu beharko genuke bestea ere; eta horrek lan bikoitza eskatuko liguke. Bada, testuetan detektatutako adibideak eta horien inguruko informazio linguistikoa *Erroreak* datu-basearen tauletan gordetzen ditugu, nahiz eta *Ikasleak* datu-baseari ere badagozkion datu horiek.

Datu-baseek, bestalde, erabiltzaile ugari izan ditzakete: euskara-irakasleak, hizkuntza-ikasleak, hizkuntzalariak, ikertzaileak... Erabiltzaile bakoitzak errore-analisiaren alorrean izango duen ikuspuntua, jakintza edota esperientzia askotarikoa izango da. Ondorioz, garrantzitsua iruditu zaigu datu-baseek gune publikoa eta pribatua edukitzea, bertan gordetzen den informazioa kontrolatuta eduki ahal izateko. Beraz, gune publikoak alde pribatuan landu diren adibideak kontsultatzeko aukera edota datu-basean adibide berriak sartzen joateko aukera eskaintzen ditu (datu-basean informazioa sartzeko, erabiltzaileak erregistratu egin beharko du). Gune pribatua, berriz, hizkuntzalari konputazional batek kontrolatzen duen gunea da. Bertan, gune publikoan erregistratutako erabiltzaileek sartutako adibideak gainbegiratzen dira; eta beharra ikusiz gero, adibideak zuzendu egiten dira edota baztertu, adibideak egokiak ez direla ondorioztatzen bada. Horrez gain, etiketatutako adibideen tratamendu automatikorako beharrezkoa den informazio tekniko ere gehitzen zaie gune publikotik sartu diren adibideei, erabiltzaile arruntek ez baitute informazio hori ezagutuko. Eta gune pribatutik bakarrik egunera daiteke, halaber, erroreen eta desbideratzeen sailkapena, horretarako beharra ikusten denean.

13. irudiak erakusten digu euskarazko errorearen eta desbideratzearen analisirako garatu dugun lan-ingurunearen eskema orokorra. Errorearen eta desbideratzearen azterketaren abiapuntua hizkuntza-ikasleen corpora da (1); izan ere, erabiltzaileek dituzten zailtasunen iturburu nagusietako bat da corpora. EtikErro editorearen bidez (2), testuetan dauden adibide erroredunak etiketatzen ditugu, euskarazko errorearen eta desbideratzearen sailkapena (3) osatzen duten kategoriak oinarri hartuta. Horrela, euskara-ikasleen corporak etiketatzeaz gain (4), erroreak eta desbideratzeak biltegiratzeko garatu dugun *Erroreak* datu-basea (5) aberasten dugu, etiketatutako adibideak datu-base horretara inportatzen baititu editoreak. Datu-base horretan ERREUS (6) izeneko aplikazioaren bidez ere sar daitezke adibideak. Adibideekin batera, errorearen tratamendu automatikorako beharrezkoa den informazio linguistikoa eta teknikoa gordetzen dira bertan, euskararako ortografia-, gramatika- eta estilo-zuzentzailea edota aldaki dialektalen markatzailea bezalako tresnen (7) garapenerako baliatzen den informazioa. Era berean, *Erroreak* datu-basea *Ikasleak* (8) datu-baseari lotzen zaio, integrazio prozesu bati esker. Horrela, errorearen edota desbideratzearen inguruko informazio linguistikoa partekatzen dute bi datu-baseek. IRAKAZI (9) aplikazioaren bidez, *Ikasleak* datu-basean hizkuntza-ikasleek egindako erroreak edota desbideratzeak gordetzen dira, datu psikolinguistikoekin batera. Informazio hori funtsezkoa da ordenagailu bidezko hizkuntzen i(ra)kaskuntzaren alorrean, erabiltzaileen beharretara egokitutako laguntza-tresnak edota bitartekoak (10) sortu ahal izateko (idazketa prozesurako kontsulta-tresnak, idazlanen ebaluazio automatikorako aplikazioa, laguntza gidatua eskaintzen duen aplikazioa...). Aipatu ditugun baliabideez gain, errorearen eta desbideratzearen analisirako ezinbestekoak zaizkigu, halaber, HAPen alorrean (11) garatu diren beste hainbat tresna ere (analizatzaile sintaktikoa, esaterako); corpus zuzenak (12) oinarri hartuta sortu dira tresna horiek eta kasu gehienetan egokitu egin behar izaten dira (13) hizkuntzen i(ra)kaskuntzaren alorrean erabiltzeko.



13. irudia. Euskarazko erroreen eta desbideratzeen analisirako lan-ingurunea.

Lan-ingurune honek, beraz, erroreak eta desbideratzeak bi euskarritan etiketatu, gorde eta kontsultatzeko aukera eskaintzen digu: testuetan bertan eta datu-baseetan. Corpusetan etiketatutako adibideak testuinguruan ikus eta azter daitezke; datu-baseetan, aldiz, adibide solteak bakarrik ikusten dira. Ortografia- edota gramatika-erroreen tratamendu automatikoa bideratzeko, nahikoa da adibide solteak (errorea dagoela ikusteko behar den testuinguru minimoa) aztertzea. Desbideratzeen tratamendu automatikoa bideratzeko, ordea, ezinbestekoa da testuak osotasunean aztertzea: estilo kontu orokorrak tratatzeko, aldaki dialektalak markatu behar diren ala ez erabakitzeke

edota erregistroaren egokitasuna neurtzeko, adibidez, testuingurua kontuan hartu beharra dago. Eta hizkuntzen i(ra)kaskuntzaren alorrean ere, ikasleen tarte-hizkuntza edota hizkuntzaren ikasketa prozesua osotasunean ezagutzera iristeko, erroreak testuinguruan aztertu behar dira, egiten diren beste errore motak eta desbideratzeak, egitura zuzenak eta hizkuntza-ikasleen inguruko informazioa kontuan hartuz. Horregatik, azaldu ditugun datu-baseetan gordetako erroreak eta desbideratzeak testuinguruan ikusi eta aztertu ahal izateko aukera prestatu dugu, adibide bakoitza dagokion jatorrizko testuari lotuz.

Esan bezala, garatu ditugun datu-baseetan gordetako adibideak eta horiei dagokien informazioa oinarri hartuta, euskarazko errorearen eta desbideratzearen diagnosi linguistikoa, teknikoa eta psikolinguistikoa bidera ditzakegu. Diagnosi horiek funtsezkoak dira bai ortografia-, gramatika- eta estilo-zuzentzailea edota aldaki dialektalen markatzailea bezalako tresnak garatzeko, eta bai ordenagailu bidezko hizkuntzen i(ra)kaskuntzaren alorrean erabiltzaileen beharretara egokitutako laguntza-tresnak prestatzeko. Lan-ingurune honek, beraz, aukera eskaintzen digu euskarazko errorearen eta desbideratzearen analisi berbera bi ikerketa-alarretan baliatzeko, eta hau ekarpen garrantzitsua da HAPen alorrean aurrerapausoak ematen jarraitzeko. Bestalde, lan-ingurune hau euskarazko errorearen analisirako garatu dugun arren, edozein hizkuntzatara egokitu daiteke.

Euskarazko errorearen eta desbideratzearen analisirako lan-ingurunea askotariko erroreak eta desbideratzeak aztertzeko prest dagoen arren, tesi-lan honetan determinatzaile-erroreak ditugu aztergai, eskura ditugun bitartekoekin errore mota guztien analisisia eta prozesamendua aldi berean bideratzea ezinezkoa baitzaigu.

III.3. Ordenagailuz lagundutako euskararen i(ra)kaskuntzarako aplikazioak

Euskarazko errorearen eta desbideratzearen analisisia bi ikerketa-alarretan baliatzeko aukera eskaintzen digun lan-ingurunea aurkeztu dugu aurreko atalean. Hemen, ordenagailuz lagundutako hizkuntzen i(ra)kaskuntzaren alorrean euskara ikasteko, lantzeko edota hobetzeko garatu ditugun bi aplikazio azalduko ditugu (Aldabe *et al.*, 2008).

Azken urteotan proiektu ugari eraman dira aurrera sare bidezko hizkuntzaren i(ra)kaskuntzaren alorrean, hala nola, hainbat kultura eta hizkuntzetako ikasleen artean ikas-estrategiak partekatzeko xedearekin sortutako ELNET (European Learning NETwork) proiektua; hizkuntza-ikasleen arteko idatzizko komunikazioa lantzeko Sperling-ek (1995) kafetegi batean girotutako foro birtuala; ikasleak eta irakasleak denbora errealean elkarrekintzan jartzen zituen DIWE sistema (Beauvois, 1997); sistemaren eta ikaslearen arteko elkarriketaren bidez ahoskera lantzeko multimedia programak biltzen zituen FLUENT proiektua (1997), eta abar luze bat.

Beste hizkuntza batzuetan egin diren lanekin aldera ezina den arren, ordenagailu bidezko euskararen i(ra)kaskuntzaren alorrean ere hainbat aplikazio sortu dira 90.eko hamarkadaz geroztik: euskara ordenagailu bidez ikasteko *Ikas-Txip* (Zubiri, 1990), *Bai&By* eta *Boga* sistemak, HEUSKLEARNING ingurunea (Armendariz *et al.*, 2004) edota irakasleentzako materiala biltzeko HABEk garatutako *Ikasbil* ingurunea, esate baterako. Izan ere, ordenagailuaren ekarpenak handiak dira hizkuntzen i(ra)kaskuntzaren alorrean: hipermediak abantaila ugari eskaintzen ditu, edozein motatako materialak (testua, bideoa, ahotsa...) barne hartzen baititu; eta Internetek ikasleari nahi duen moduan nabigatzeko eta interesatzen zaion informazioa aukeratzeko askatasuna ematen dio. Ikaslea informazio-sarean galtzeko arriskua ukaezina den arren, teknologia berriak hizkuntzen i(ra)kaskuntzaren alorrean erabiltzen hastea aurrerapauso garrantzitsua izan da: hasiera bateko CD-ROMek eskaintzen zituzten tresna arruntetatik (hiztegiak, entziklopediak, jokoak...) haratago, erabiltzaileen beharrak kontuan hartzeko eta erabiltzaileen arteko elkarrekintzarako aukera ematen du hipermediak.

Ordenagailuz lagunduta hizkuntzak i(ra)kasteko garatutako sistemek ariketa *itxiak* eta *irekiak* eskaini ahal dituzte. Ariketa itxietan (testak, ordenazio-ariketak, aukera anitzeko ariketak...) emaitza zuzena gordeta izan ohi du ordenagailuak; eta emaitza posibleak aurreikusita daudenez, erantzunak automatikoki trata daitezke. Hala ere, ariketa mota horiek bi desabantaila dituzte: erantzun posibleak mugatuak dira eta zoriz erantzun eta asmatzeko aukera izaten da. Ariketa irekiek, aldiz, ariketa itxiek ez dituzten abantailak eta onurak eskaintzen dizkigute, ikasleak nahi duena idazteko aukera ematen baitute. Ondorioz, ikasleak hizkuntzaz duen ezagutzari buruzko informazioa zehatzagoa da, eta posible da ikasle bakoitzak duen hizkuntza-mailaren araberrako laguntza eskaintzea. Ariketa irekiek, baina, badute eragozpen bat: ezinezkoa da ikaslearen emaitzak aurreikustea eta, beraz, ikasketa-prozesua automatikoki

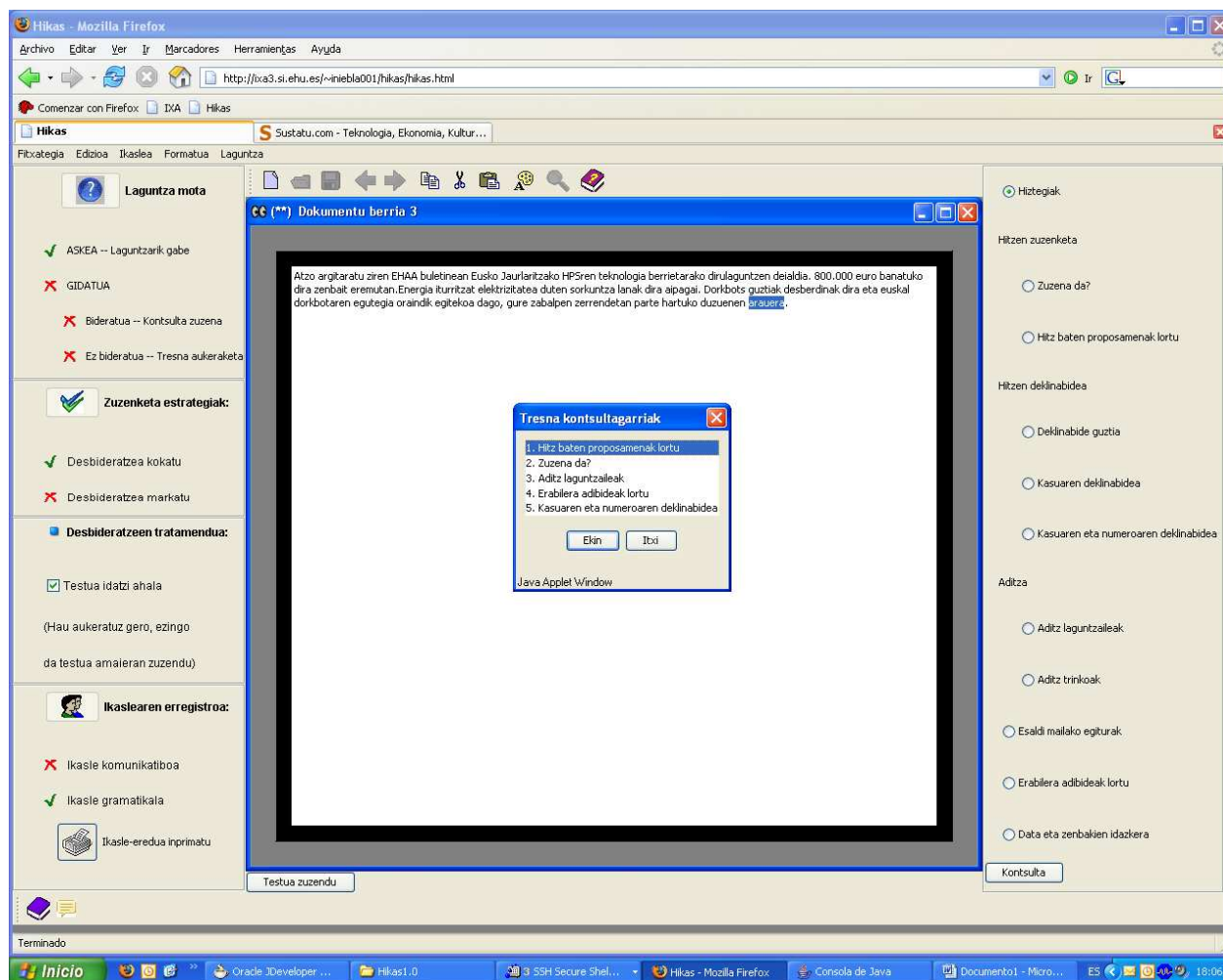
aztertzea eta tratatzea zaila gertatzen da. Ariketa irekiak prestatzeko ezinbestekoa da HAPen alorrean egindako lana, hau da, hizkuntza i(r)a(k)asteko aplikazio irekien oinarrian HAPen alorrean garatutako sistemak daude.

Euskara ikasi, landu edota hobetu nahi duten ikasleentzako tresna linguistikoak prestatzen direnean, garrantzitsua da, halaber, erabiltzaileen hizkuntza-ezagutza kontuan hartzea. Horretarako, ezinbestekoa da diseinatzen ditugun tresnetan ikasle-eredua (ikasleari eta bere tarte-hizkuntzari dagokion informazioa) integratzea. Eta III.2. atalean aurkeztu dugun *Ikasleak* datu-baseak, hain zuzen, hizkuntza-ikasleen datuak eta egiten dituzten errorearen edota desbideratzearen inguruko informazioa gordetzeko, aztertzeako eta prozesatzeko aukera eskaintzen digu. Bada, datu-base horretan bildutako informazioa beharrezkoa da ikasle bakoitzari edota ikastalde oso bati dagozkion datuak eskuratzeko, hizkuntza-maila bakoitzean egin ohi diren errore tipikoenak zein diren jakiteko, maila desberdinetan egiten diren errore motak eta kopuruak alderatzeko, mota bakoitzeko errore-adibideak eskuratzeko, adibide horietan oinarrituta ikas-estrategiak prestatzeko, etab.

IXA taldean, HAP alorrean garatu diren hainbat tresna eta *Ikasleak* datu-basean bildutako informazioa (erroreei, desbideratzei eta hizkuntza-ikasleei dagokiena) baliatuz, ordenagailuz lagundutako hizkuntzen i(r)a(k)askuntzaren alorrean bi aplikazio prestatu ditugu: HIKAS (Aldabe *et al.*, 2007b) eta idazlanen ebaluazio automatikorako aplikazioa (Castro-Castro *et al.*, 2008).

Ikasle-eredua kontuan hartuz ikaslearen beharrak asetzen dituen web aplikazioa da HIKAS. Aplikazioaren helburua hizkuntza-ikasleari idazketa prozesuan laguntzeko tresnak eskaintzea da, testu bat idatzi ahala hizkuntza-ikasleari sor dakizkiokeen zalantzak argitzeko behar duen laguntza eskaintzea, ikasle-eredua kontuan hartuta. Horretarako, hainbat tresna linguistiko integratu dira aplikazioan: hiztegiak; hitz okerrak zuzentzeko proposamenak ematen dizkigun tresna; hitzak nola deklinatu behar diren ikasteko balio duen tresna; aditzaren erabileraren inguruko informazioa ematen duena; datak, orduak eta zenbakiak nola idatzi behar diren erakusten diguna; erabileraren adibideak erakusten dizkiguna... Horrez gain, ikaslearen ikasketa prozesuaren jarraipena egiten du aplikazioak, hizkuntza-ikasleak dituen zalantzak eta zailtasunak aztertu, egindako aurrerapenak ezagutu eta ikasle-ereduaren araberrako laguntza eskaini ahal izateko. HIKASen interfazearen eskuinaldean ageri dira kontsulta daitezkeen

laguntza-tresnak eta ezker aldean, berriz, ikasle-ereduari dagozkion aukerak (14. irudia). Ikasle-eredua definitzeko, *Ikasleak* datu-basean gordetako informazioa baliatzen da.

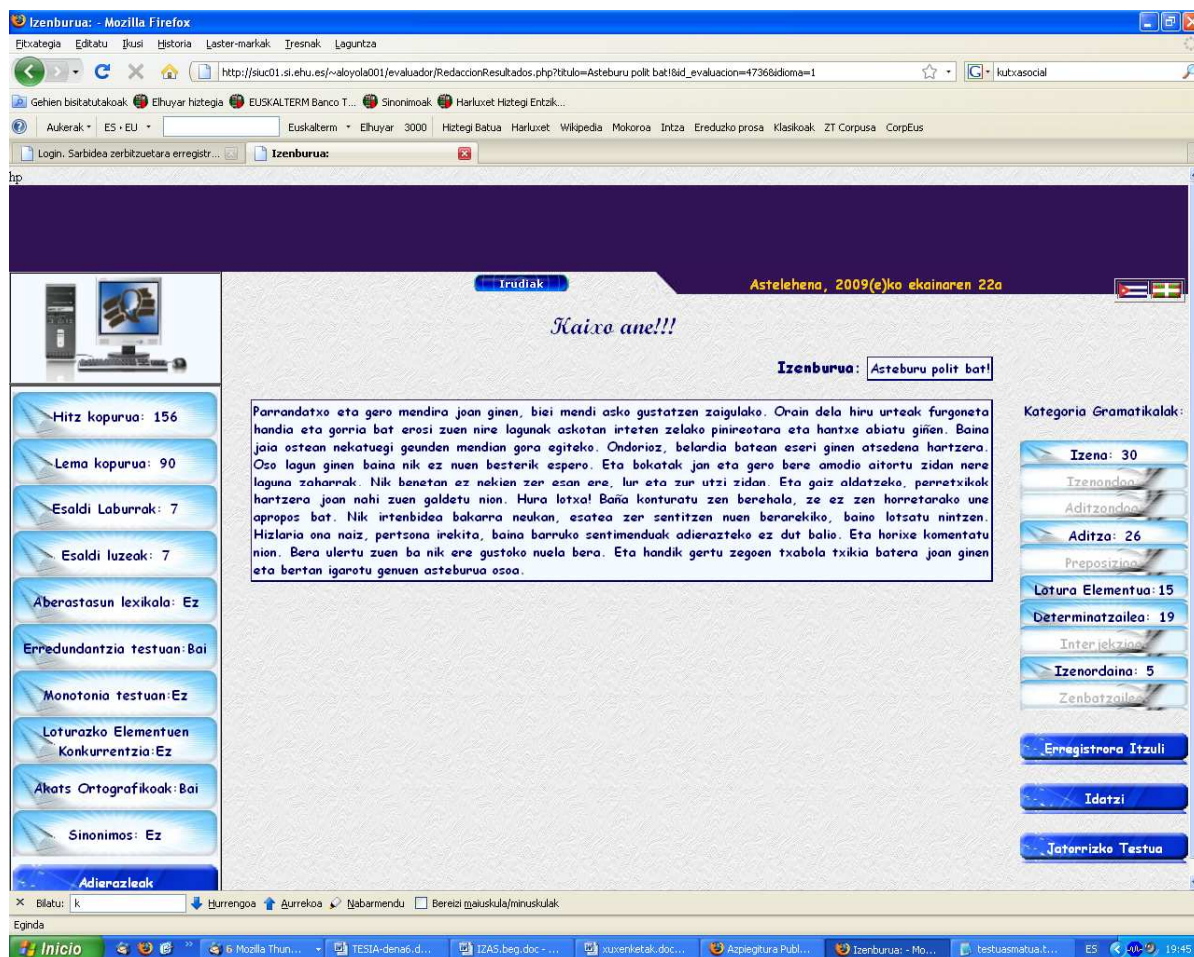


14. irudia. HIKAS aplikazioaren interfazea.

Halaber, errore edota desbideratze jakinen inguruko gramatika-oharrak, arauak, gomendioak edota proposamenak kontsultatzeko edota horien inguruko zalantzak argitzeko tresnak prestatu eta integratu ahalko dira HIKAS aplikazioan, III.2. atalean azaldu dugun lan-ingurunean jasotzen diren datu linguistikoak eta teknikoak baliatuta.

Idazlanen ebaluazio automatikorako aplikazioari dagokionez, sistema eleaniztuna da, euskara eta gaztelania barne hartzen dituena. Aplikazioa erabilerraza da: testu bat idatzi eta “analizatu” botoia sakatuz, testuaren ebaluazio orokor bat ematen zaio erabiltzaileari (irakasleari edo ikasleari) (15. irudia). Horretarako, analizatutako testuan

dauden hitz eta lema kopuruak kontatzen dira; esaldi laburrak eta luzeak identifikatzen dira; aberastasun lexikoa ebaluatzen da; testuetan erredundantziarik badagoen esaten da, bai eta testua monotonia egiten den ala ez; errore ortografikorik ba ote dagoen adierazten da; eta kontuan hartzen da sinonimorik edota klase semantiko berekoak diren hitzik erabili den. Hitzen gramatika-kategoriak (izenak, aditzak, loturazko elementuak, determinatzaileak...) eta kopuruak ere zehazten dira. Datu horietan guztietan oinarrituta, aplikazioak testuaren ebaluazio orokor bat egin eta hainbat gomendio ematen ditu. Ebaluatutako testuak gorde egiten dira, ikasle eta irakasle bakoitzak edozein unetan egindako lanak kontsultatzeko aukera izan dezan.



15. irudia. Idazlanen ebaluazio automatikorako sistemaren interfazea.

Testuen ebaluazio automatikoa bideratzeko ezinbestekoa da ikasle-eredua (ikasle bakoitzari dagokion informazio psikolinguistikoa, *Ikasleak* datu-basean gordetzen dena), ikasleak duen ezagutza-mailaren arabera kontuan hartu behar zaizkion erroreak

edota desbideratzeak zein diren jakiteko. Eta beste faktore batzuk ere aintzat hartzekoak dira: hitz berdinen eta desberdinen kopuruak, paragrafoen antolaketa, esaldien ordena edota egitura linguistikoen egokitasuna, jarioa, edukia, koherentzia... Izan ere, idazlan oso baten ebaluazioa egiteko ez da nahikoa erroreen detekzio automatikoa soilik.

Bestalde, epe laburrean integratuko da XUXENg-ren lehen bertsioa aurkeztu ditugun aplikazioetan. Horrela, orain artean tratatu diren gramatika-erroreak detektatuko dira: komunztadura-erroreak, data-egituretan egindakoak, postposizio-lokuzioetan egindakoak edota determinatzaile-erroreak, besteak beste. Gainera, erroreak detektatzeko idatzi ditugun erregeletan zehazten diren datuak (errore bakoitzari dagokion kategoria linguistikoa, definizioa edota deskripzioa, adibidea eta zuzenketa(k)) ere baliatu ahalko dira hizkuntza-ikasleei oharrak edota aholkuak eman ahal izateko.

III.4. Laburbilpena

Tesi-lan honen ekarpen nagusietako bat da kapitulu honetan aurkeztu dugun euskarazko erroreen eta desbideratzeen analisirako lan-ingurunea. Ingurune hau hainbat baliabidek osatzen dute: corpusek (jatorrizkoek eta etiketatuek), adibide erroredunak etiketatzeko errore-editoreak, euskarazko erroreen eta desbideratzeen sailkapenak eta etiketatutako adibideak eta horiei dagokien informazioa (linguistikoa, teknikoa edota psikolinguistikoa) gordetzeko datu-baseek. Baliabide horiek guztiak funtsezkoak dira bai erroreen eta desbideratzeen tratamendu automatikoa bideratzeko eta bai ordenagailuz lagundutako hizkuntzen i(ra)kaskuntzaren alorrean hizkuntza ikasteko, lantzeko edota hobetzeko lagungarriak diren tresnak prestatzeko. Bada, lan-ingurune honetan bi ikuspuntu edo ikerlerro uztartzen ditugu: alor batean egindako lana baliagarria zaigu besterako, eta alderantziz. Bi ikerketa-alorrak bateratzea ekarpen garrantzitsua da HAPen alorrean.

EtikErro editorea eta bi datu-baseak (*Erroreak* eta *Ikasleak*) aurkeztu ditugu hemen. Errore-editoreak, oro har, testuetako adibideak etiketatzeko baliatzen dira, errore-sailkapenak oinarri hartuta; ondoren, informazioa erauzteko tresnak baliatu ohi dira etiketatutako adibideen inguruko datuak eskuratzeko eta aztertzeako. Errore-editore gehienek ez bezala, EtikErro-k aukera ematen digu erroreak eta desbideratzeak corpusen gainean etiketatu eta adibide horiek, automatikoki, datu-base batera

inportatzeko. Ondoren, datu-basean gordetako adibideei informazio osagarria gehitzen diegu.

Testuetan etiketatutako adibideak eta horien inguruko datuak biltegitratzeko prestatu ditugu, hain zuzen, bi datu-baseak. *Erroreak* datu-basean jasotako adibideak erroreen tratamendu automatikoaren alorrean euskararako ortografia-, gramatika- eta estilo-zuzentzailea edota aldaki dialektalen markatzailea bezalako tresnak garatzeko baliatzen ditugu eta, beraz, adibide bakoitzari dagokion informazio linguistikoa eta teknikoa gordetzen ditugu bertan. *Ikasleak* datu-basean, berriz, etiketatutako adibideei dagokien informazio linguistikoarekin batera, datu psikolinguistikoak jasotzen ditugu, ordenagailuz lagundutako hizkuntzen i(r)a kaskuntzaren alorrean erabiltzaileen beharrak kontuan hartzen dituzten laguntza-tresnak garatu ahal izateko, hala nola, HIKAS aplikazioa edota idazlanen ebaluazio automatikorako aplikazioa. Datu-base biak loturik daudenez, alor bateko lana baliagarria zaigu besterako, eta alderantziz. Hala, erroreen eta desbideratzeen tratamendu automatikoaren alorrean egindako lana, esaterako, aipatu aplikazioetan ere balia dezakegu: landutako gramatika-erroreak detektatuko dira eta erroreei dagokien informazioa (kategoria linguistikoa, deskripzioa, adibideak edota zuzenketak) hizkuntza-ikasleei oharrak edo aholkuak emateko erabili ahalko da.

Aurkeztutako errore-editoreaz eta bi datu-baseez gain, euskara-ikasleen corpusa eta euskarazko erroreen eta desbideratzeen sailkapena dira lan-ingurune honi dagozkion oinarritzko beste bi baliabide. Horietaz hitz egingo dugu, hain zuzen, ondorengo kapituluetan (IV. eta V. kapituluak).

Hizkuntza-ikasleen corpora errorean eta desbideratzeen informazio-iturri aberatsa da, gure azterketen abiapuntua.

IV. Corpusgintza eta hizkuntza-ikasleen corpora

IV.1. Sarrera

Hizkuntzalaritzaren alorrean azken urteotan asko ugaritu dira corpusak oinarri hartzen dituzten ikerlanak. Izan ere, corpusak hizkuntza lantzeko baliabide ezin hobeak dira, hizkuntza-ekoizpenak aztertzeko eta gramatika-teoriek diotena egiaztatzeko bitarteko aproposak. Euskarri elektronikoan jasotako testuen gainean ordenagailuaren laguntzaz azterketa linguistikoak egiteari corpus-hizkuntzalaritza esaten diogu (Granger, 1998a; Ties, 2006) eta corpus-hizkuntzalaritzaren metodologiari, berriz, corpusgintza (Areta *et al.*, 2008). Corpusgintza diziplina arteko alorra da, hizkuntzalarien eta informatikarien ezagutzak eta ekarpenak behar dituena.

Lehenengo corpus elektronikoa (Brown corpora) argitaratu zen urtea (1964) hartzen da corpusgintza modernoaren eta corpus-hizkuntzalaritzaren hasiera-urtetzat. Chomsky-k eta teorialari sortzaileek corpus-hizkuntzalariei kritikatzeko zieten corpora ez zela hizkuntza aztertzeko baliabide egokia, corpus bat ezin zitekeela izan mugagabea (infinitua) den hizkuntza baten eredu. Euren iritziz, corpusek hizkuntzaren erabilera errealaren (*performance*) inguruko informazioa eskaintzen digute, baina ez dira ekoizlearen hizkuntza-gaitasuna (*linguistic competence*)

islatzeko bitarteko egokiak; corpusetan oinarritutako ikerketek, gainera, zerrenda simple batzuetara murrizten dituzte deskripzio linguistikoak (Teubert, 2006). Hizkuntzaren Azterketa eta Prozesamenduaren (HAP) alorrean egindako aurrerapenei esker, ordea, 1980 urte ingurutik aurrera benetako indarra eta oparotasuna hartzen hasi zen corpusgintza, eta gaur egun, jada, inork ez du zalantzan jartzen, edozein teoria linguistikoren aldekoa izanda ere, corpusak beharrezkoak direla hizkuntza ‘den bezala’ aztertu nahi baldin bada (Rojo, 2002).

Corpusa dokumentu askok osatutako testu-multzoa da; diseinu bat edota irizpide batzuk oinarri hartuta, hizkuntza aztertzeko, hizkuntzari buruzko datuak eskuratzeko eta prozesatzeko baliatzen den testu-bilduma sistematikoa¹¹. Ikertzaile bakoitzak ikusmiran dituen aztergai edota helburuen arabera, irizpide jakin batzuk finkatzen ditu (Atkins *et al.*, 1992; Sinclair, 1996; McEnery *et al.*, 2005; Evans¹²): gaiaren, erregistroaren, dialektoaren, ekoizlearen edota jatorriaren araberrako corpusak eratu ohi dira.

Hizkuntza edo hizkera baten lagin adierazgarria da corpusa, ahozkoarena edota idatzizkoarena. Egun, jada, euskarri elektronikoan biltegituta egon ohi dira testuak, sistema informatikoen laguntzaz azterketa linguistikoak egiteko prestatuak (Francis, 1982; Sinclair, 1991; Engwall, 1992; Leech eta Fliegelstone, 1992; EAGLES, 1996; Bowker eta Pearson, 2002; Meyer, 2002; Oyharçabal, 2002; Saurí, 2004; Brady *et al.*, 2005; McEnery *et al.*, 2005; Ties, 2006; Areta *et al.*, 2008). Ordenagailuaren laguntzari esker, ikerketak azkarrago egiten dira eta emaitzak fidagarriagoak izan ohi dira. Corpusak, gainera, hainbat arlotan baliatzen dira: hiztegigintzan, hizkuntzen i(r)a-kaskuntzan, hizkuntza-teknologietan, etab. Baliabide emankorrek eta eraginkorrek dira corpusak.

Kapitulu honetan corpusgintzaz, oro har, eta zehazkiago, euskara-ikasleen corpusaz mintzatuko gara. Hasteko, IV.2. atalean, corpusgintzaren mugak eta abantailak aipatuko ditugu. IV.3. atalean corpusgintzari dagokion metodologia

¹¹ *Sistematiko* diogunean esan nahi dugu corpusaren egitura eta edukiak antolatzeke garaian hainbat irizpide linguistiko eta estralinguistiko –testuaren jatorria, testu kopurua, hautatutako testu motak, testuak aukeratzeko arrazoiak, etab.– kontuan hartu behar direla.

¹² <http://www.corpus.bham.ac.uk/Unit1.pdf>

izango dugu hizpide. Eratu ohi diren corpus motak eta hainbat hizkuntzatan osatu diren testu-bildumak aurkeztuko ditugu ondoren, IV.4. atalean. Jarraian, IV.5. puntuan, euskarazko erroreak eta desbideratzeak aztertzeko ezinbestekoa zaigun hizkuntza-ikasleen corpusaz hitz egingo dugu: euskara-ikasleen testuak identifikatzeko asmatu dugun kode-sistema azalduko dugu (IV.5.1.) eta euskarazko erroreen eta desbideratzeen analisirako baliatzen dugun corpusaren inguruko datuak zehaztuko ditugu (IV.5.2.). Azkenik, IV.6. atalean, kapitulu honetako ideia garrantzitsuenak laburbilduko ditugu.

IV.2. Corpusgintza: mugak eta abantailak

Azken hamarkadetan HAPen eta teknologia berrien alorretan egin diren aurrerapenei esker, corpusgintzak indarra hartu du (Granger, 1998a; Tono, 2003; Chujo *et al.*, 2004; Liu Fang, 2004; McEnery *et al.*, 2005) eta azterketa linguistikoak egiteko oinarri sendoa, aberatsa eta eraginkorra bilakatu da corpusa: “*corpuseko datuak aztertuz, hizkuntzaren aztertzailerek (hizkuntzalariek, irakasleek, euskara-teknikariek...) ondorioak atera ditzakete eta proposamenak egin ere bai, dagokion alorreko adituek hizkuntza-ereduari buruzko argibideak edo ‘gidalerroak’ izan ditzaten*” (Alegria *et al.*, 2005).

Corpusgintzak, ordea, baditu hainbat muga edo zailtasun. Corpusak eratzea, esaterako, lan handia eta konplexua da: testuak eskuratzeko jatorri emankorrak eta baliagarriak bilatu behar dira lehenik eta behin, eta corpusa ondo antolatzeak metodologia bat definitu. Garrantzitsua da, halaber, egingo den ikerketaren helburuak eta irizpideak hasieratik finkatzea, horrek baldintzatuko baitu, hein handi batean, azterketaren arrakasta edo porrota (Demuth, 2008): ikerketaren kalitatea corpusaren kalitateari lotuta egon ohi da eta biltzen den corpus motak eragin zuzena izango du lortzen diren emaitzetan (Granger, 1998a). Gerta liteke, ordea, finkatutako helburu horiek lortzeko osatzen den corpusa baliagarria ez gertatzea. Izan ere, corpusa diseinatu, bildu eta antolatzeak bere denbora eskatzen du, corpusa prestatzen denbora gehiegi pasatzen da batzuetan eta bitarte horretan hizkuntzak aldaketak jasan baldin baditu, baliteke testuak erabilgarri jartzen direnerako corpusa zaharkituta geratzea (Granger, 1998a). Arazo hori areagotu egiten da euskara bezalako hizkuntza ez-normalizatuetan. Bestalde, corpusen ezaugarri nagusietako bat adierazgarritasuna

da, eta oso zaila da bildutako lagina adierazgarria izango den ala ez hasieratik ziurtatzea.

Tresna informatikoak ere ezinbestekoak dira corpusgintzaren alorrean, bildutako testuak aberastu eta aztertu ahal izateko. HAPen alorrean garatutako tresnak eskuratzea, ordea, zaila da ikerketa-talde baten babesik gabe. Tresna informatikoak eskura izanda ere, fenomeno linguistiko askoren azterketak eskuzko lana eskatzen du oraindik, lan neketsua eta garestia. Gainera, hizkuntzaren prozesamendurako baliatzen ditugun teknikak ere beren mugak izan ohi dituzte. Corpus-bilketaren beste mugetako bat egile-eskubideen eta baimenen kontua izan ohi da, kasu batzuetan eragozpen asko jartzen baitira. Eta corpusaren kostua ere desabantaila izan daiteke.

Aipatu mugak hor dauden arren, ordea, ez dugu zalantzarik corpusgintzak desabantaila baino abantaila gehiago eskaintzen dituela: esate baterako, gaur egun, jada, testu gehienak euskarri elektronikoan jasotzen dira eta hain astuna eta garestia den transkripzio-lana aurrezten zaio, horrela, ikertzaileari. Paperezko formatuak ez bezala, euskarri informatikoak testu-masa handiak gorde eta eskuragarri jartzeko aukera eskaintzen digu, corpusak informazio linguistikoarekin aberasten dira eta interesatzen zaizkigun azterketa linguistikoak paper gaineko formatuan baino bizkorrago egiten dira horrela. Ordenagailu bidezko azterketetan lortzen diren datuak edota emaitzak, bestalde, eskuz lortzen direnak baino zehatzagoak eta fidagarriagoak izan ohi dira (Leech, 1993). Eta euskarri elektronikoan gordetako corpusak, gainera, dinamikotasuna eskaintzen du, paperezko testuen aldean (Behrens, 2008).

Egun, azterketa linguistikoak egiteko eta hizkuntza konputazionalki tratatzeko baliabide gero eta sendoagoak garatzen dira corpusgintzaren alorrean (Areta *et al.*, 2008): corpusetan fenomeno linguistiko jakinen kontaketa eta konparazio azkarrak egiteko sortu diren metodo sinpleetatik haratago, jatorrizko corpusak informazio linguistikoarekin aberasteko aukera gehiago dauzkagu HAPen alorrean garatu diren tresnei esker (lematizatzaileak, analizatzaile morfologikoak, morfosintaktikoak eta sintaktikoak, *treebankak*, etab.). Testuen gainean azterketa linguistikoak egiteko baliabide horiek ezinbestekoak zaizkigu corpus-hizkuntzalaritzaren alorrean aurrera egin ahal izateko, bitarteko horietatik eskuratzen dugun informazioa baita, besteak beste, egiten ditugun ikerketa askoren oinarria.

Corpus-hizkuntzalaritza hizkuntzaren erabilera errealetan oinarritzen da eta, beraz, datu kuantitatibo edota kualitatibo errealak islatzen ditu. Horrek hizkuntzaren azterketa praktikoagoak, aplikagarriagoak, egiteko aukera eskaintzen digu: erabiltzaileek dituzten benetako zailtasunak aztertuz, bakoitzaren beharretara egokitzen den laguntza pertsonalizatua bidera daiteke. Izan ere, ordenagailuak erabiltzaileari eskainiko dion laguntza adibide espekulatiboetan oinarrituta egotea baino egokiagoa da hizkuntza errealetik ahalik eta hurbilen egotea. Hizkuntzalarien eta hiztunen intuizioan oinarritutako arau abstraktuak egiaztatzeko ere, gero eta gehiago baliatzen dira corpusak hizkuntzalaritzaren alorrean.

Bestalde, corpusak eratzeko metodologia finkatuz joan da (McEnery eta Wilson, 1996; Granger, 1998a; Tono, 2003; Granger, 2004) eta corpusak egiteko eta kudeatzeko tresnak, corpusak etiketatzeko bitartekoak edota testuak analizatzeko eta ustiatzeko baliabideak ere ugarituz doaz. Testuinguru honetan, aipatzekoa da informatikarien eta hizkuntzalarien arteko elkarlana; izan ere, osagarriak dira batzuen eta besteen ekarpenak; beharrezkoak dira bi alorretako jakintzak corpusen azterketa oso bat egin ahal izateko.

Asko dira, beraz, corpusgintzaren abantailak: euskarri elektronikoan gordetako corpusak eta HAPen alorrean garatutako tresnak eskura izanik, testuen azterketa-lana erraztu eta bizkortu egiten da eta datu errealetan oinarritutako emaitzak lortzen dira. Corpusa funtsezko bitartekoa da, bai HAPen alorrean eta bai errorean edota desbideratzeen analisiaren alorrean ere; erabiltzaileen beharrak edota zailtasunak kontuan hartzen dituzten tresnak garatzeko aukera ematen digun funtsezko bitartekoa da corpusa.

IV.3. Corpusgintzaren metodologia orokorra

Corpusgintza hizkuntza aztertzeko testuak baliatzen dituen metodologia da; ez da hizkuntzalaritzaren barruko alor bat fonetika, sintaxia, semantika edota pragmatika diren moduan. Aipatu alor horietan hizkuntza ikuspegi desberdinetatik deskribatzen edo azaltzen saiatzen gara, baina corpusgintza ez da hizkuntzaren alor bakar batera mugatzen; aitzitik, beste alor horietan ikertzeko balia dezakegun metodologia da. Lau urrats nagusi bereizi ohi dira corpusgintzan: corpusaren diseinua, corpus gordinaren eraketa (testuen bilketa eta antolaketa), corpusaren etiketatzea eta

corpusak analizatzeko edota ustiatzeko tresnen garapena (McEnery eta Wilson, 1996; Granger, 1998a; Meyer, 2002; Alegria *et al.*, 2005; IXA taldea eta Elhuyar Fundazioa, 2007).

Corpusa ez da testu multzo handi baten bilketa hutsa. Corpusak datu errealean eta adierazgarrietan oinarritutako deskripzio linguistikoak egiteko, ondorioak ateratzeko edota hizkuntza-tresna lagungarriak garatzeko bitarteko eraginkorra eta ahaltsua behar du izan. Horregatik, corpus-bilketaren diseinua funtsezko urratsa da. Baliabide eraginkor bat sortu nahi badugu, ezinbestekoa da corpora osatuko duten testuak biltzeko irizpideak hasieratik ongi finkatzea: corpora zertarako erabiliko den, testuak zeren adierazgarri izatea nahi dugun, zer nolako testu motak interesatzen zaizkigun eta zertarako. “*Corpusa populazio oso baten lagin-multzo bat da eta ikuspegi estatistikotik, lagin hori adierazgarria izango da baldin eta hura aztertuz ateratzen ditugun ondorioek populazio osorako balio badute (...) Diseinuaren xedea da helburu horretarako adierazgarria izango den lagin-bilketa bat egitea lortzea*” (Alegria *et al.*, 2005). Corpusaren diseinuan ondoko parametro hauek hartu behar dira kontuan:

- Corpusaren tamaina: corpora osatuko duten testuen kopurua. Maiz agertzen den fenomeno linguistikoa aztertzeko, normalean ez da corpus handiegirik behar izaten; aztergaia usukoa ez bada, aldiz, corpusak handia beharko du. Tamaina handiko corpusak osatzea lortzen badugu, aukera daukagu gero corpus osoa edo corpusaren lagin txikiago bat baliatzeko, aztergai dugun fenomeno linguistikoaren arabera.
- Corpusaren oreka: corpus orekatuak eratzea komeni izaten da; horretarako, hainbat jatorriz, gaiz, erregistroz, diskurtsoz, helburuz, etabarrez sortutako ekoizpenak bildu behar dira, eta azpiatal bakoitzeko testu kopuru berdintsua gorde behar da. Corpus orekatu esaten zaie, hain zuzen, “*argitaratutako testuen artean metodo estatistikoak erabiliz adierazgarritasuna eta aniztasuna bilatzen duten corpusei*” (IXA taldea eta Elhuyar Fundazioa, 2007).
- Corpusaren izaera: corpora itxia edo irekia izatea. Corpus itxietan ez dago testu berriak sartzeko aukerarik eta, beraz, informazioa egonkor mantentzen da. Corpus irekietan, aldiz, testu berriak gehitzen direnez, datuak aldatzen doaz.

- Corpusaren adierazgarritasuna: aztergai dugun fenomeno linguistikoa edo hizkuntza-eredua ahalik eta ondoen islatzen duen testu-bilduma osatuari esaten zaio corpus adierazgarria. Corpusaren adierazgarritasuna neurriaren eta orekaren arabera neurtu ohi da: zenbat eta corpus handiagoa osatu eta zenbat eta dialekto, erregistro, gai edota genero gehiago bildu, orduan eta aberatsagoa izango da osatutako hizkuntza-eredua; aztergai diren hizkuntza-fenomenoen aniztasuna handiagoa izango da. Bestalde, corpusa ikergai dugun hizkeraren erakusgarri erreala izatea nahi badugu, biltzen ditugun testuek ez dute bereziki corpusa osatzeko idatziak izan behar.

Corpusa diseinatu eta eratu ostean, etiketatze-faseari ekingo zaio. Izan ere, corpusak hizkuntza-azterketak egiteko baliatu nahi badira, ezinbestekoa da testuak informazio linguistikoarekin etiketatu eta hornitzea. Bi etiketatze-fase daude: corpusaren oinarritzko egituraren etiketatzea (izenburu, azpiatal, paragrafo, taula, oin-ohar eta testuen ezaugarri tipografiko esanguratsuen etiketatzea¹³) eta etiketatze linguistikoa (hitzei informazio linguistikoa gehitzea). Testuak etiketatzea eta prozesatzea ezinbestekoa da corpusa bitarteko linguistiko gisa baliatu nahi baldin bada eta, horretarako, tresna ugari garatu dira dagoeneko: lematizatzaileak, analizatzaile morfologikoak, morfosintaktikoak eta sintaktikoak, *treebankak*, etab. Corpusaren etiketatze linguistikoa automatikoki, erdi-automatikoki edota eskuz egin ohi da, kasuan kasu. Emankorra izan dadin, Leech-ek (1993) corpusen etiketatze-prozesuan derrigorrean bete beharreko irizpide hauek proposatzen ditu:

- Etiketatutako corpusetik etiketak ezabatzeko aukera egongo da, hau da, testu gordina, jatorrizkoa, berreskuratzeko aukera.
- Corpusean txertatutako etiketak bere horretan erauzi ahalko dira.
- Corpusa etiketatzeko irizpide batzuk finkatu edo jarraibide batzuk prestatu behar dira eta etiketatzaileek eskura izango dute etiketatze-prozesuari dagokion eskuliburua, denek ahalik eta modu koherenteenean etiketa dezaten.

¹³ Dokumentuak kodetzeko erabiltzen den eredu ezagunenetakoa bat TEI (*Text Encoding Initiative*) da.

- Etiketatzeko eskemek aho batez onarturiko irizpideetan oinarrituta egon behar dute.
- Corpora nork eta nola etiketatu duen zehaztu behar da beti: etiketatzailerik bakar bat edo gehiago aritu diren; etiketak automatikoki, erdi-automatikoki edota eskuz esleitu diren; etiketatze-prozesua gainbegiratu den, etab.
- Etiketatzearen fidagarritasuna eta sendotasuna bermatu behar da, etiketatutako corpusean beti errore-tasa bat egongo den arren.

Irizpide horiek kontuan hartzea garrantzitsua da, beraz. Hala ere, ikertzaile bakoitzak etiketatze-lanak egiteko finkatzen dituen jarraibideak, oro har, bere helburuen arabera izan ohi dira eta, beraz, hizkuntzaren azterketak egiteko biltzen diren corpusen diseinua aldatu egiten da ikerketa-proiektu batetik bestera (Tono, 2003). Testu erroredunak biltzea lan zaila da eta batzuetan finkatutako hainbat irizpide zorrotz betetzea ez da erraza izaten. Baina guk Leech-ek proposatutako irizpideak kontuan hartu ditugu tesi-lan honetan aurkezten dugun euskara-ikasleen corpora bildu eta etiketatzeko.

IV.4. Corpus motak

Eratzen diren corpus motak askotarikoak dira. Lehenengo bereizketa nagusia ahozko eta idatzizko corpusen artekoa da, euskarriaren aldetik desberdinak baitira. HAPen alorrean, ahozko corpusak transkribatu egin behar izaten dira testuak aztertu ahal izateko. Idatzizko corpusak ere, paperean baldin badaude, transkribatu egiten dira ondoren ordenagailu bidez tratatu ahal izateko. Hala ere, gaur egun, jada, testuak euskarri informatikoan jasotzea da normalena.

Corpus moten aukera zabala da: corpora izan daiteke orotarikoa edo berezia, orokorra edo espezializatua, eleaniztuna edo elebakarra, diakronikoa edo sinkronikoa, irekia edo itxia, etiketatua edo etiketatu gabea, e.a. Orotariko corpora edo erreferentzia-corpora da mota guztietako testuak (testu orokorrak, bereziak, espezializatuak) jasotzen dituen bilduma, hau da, *“hizkuntzaren erabilera-eremu guztietarako baliagarria edo adierazgarria izatea helburu duena”* (Aranzabe, 2008); corpus berezia, berriz, dialekto, genero, ekoizle, garai... jakin bateko testuez osaturiko testu-multzoa da. Corpus orokorra komunitate batek partekatzen duen

hizkuntza islatzen duen bilduma da, hizkuntzaren ezaugarri nagusiak eta aldaera guztiak aztertzeko prestatzen dena; eta corpus espezializatua komunikazio-arlo espezializatu batean erabiltzen diren testuen bildumek osatzen dute (alor teknikoak, zientziarenak, giza eta gizarte-zientziarenak, etab.); hain zuzen ere, arlo bakoitzari dagozkion ezaugarriak aztertzeko erabiltzen dira corpus espezializatuak. Corpusak, gainera, izan daitezke elebakarrak (hizkuntza bakar batean eginiko ekoizpenen bildumak) edo eleaniztunak (hainbat hizkuntzatan sortutako testu-multzoak, bi hizkuntza-sistemaren edo gehiagoren artean azterketa konparatiboak egiteko baliagarriak direnak); diakronikoak (historian zehar hizkuntzak izan duen bilakaera islatzen duten corpusak, une jakin batetik aurrera hizkuntza nola aldatu den erakusten dutenak) edo sinkronikoak (une jakin batean hizkuntzaren egoera zein den aztertzeko baliatzen direnak); irekiak (testu berriak gehituz etengabe osatuz doazenak) edo itxiak (memento jakinetan bildutako testu-multzo finkoak); etiketatuak (eskuz edota automatikoki informazio linguistikoz hornitutako corpusak) edo etiketatu gabeak (jatorrizko corpusak, inongo etiketa edo informazio gehigarririk ez dutenak).

Erroreen analisirako, oro har, hizkuntza-ikasleen corpusak baliatu ohi dira gehienbat, testu horietan baitago aztertu beharreko adibideen kopururik handiena. Euskarazko erroreak eta desbideratzeak aztertzeko ere euskara-ikasleen testuak erabiltzen ditugu. Euskara-ikasleen corpusa corpus berezia da, ekoizle mota jakin baten testuek osatzen dutelako; elebakarra da, euskaraz eginiko idazlanen bilduma delako; corpus irekia da, testu berriekin handitzen eta osatzen joango dena, nahiz eta analisi jakinetarako corpus-lagin itxiak baliatzen ditugun. Gure helburua jatorrizko testuak etiketatzea da, informazio linguistikoz hornitutako testuak bitarteko aberatsak baitira, bai erroreen tratamendu automatikorako baliagarriak diren adibideak eskuratzeko eta bai hizkuntza-ikasleen i(ra)kaskuntza prozesuari dagozkion ezaugarriak edota urratsak aztertzeko. Euskara-ikasleen corpusak, bestalde, azterketa sinkronikoak edota diakronikoak egiteko aukera eskaintzen digu, hau da, hizkuntza-maila jakin bati dagozkion ezaugarrien analisia egin daiteke edota euskara-ikasleen i(ra)kaskuntzaren ibilbidea ikertu, maila desberdinetako testuak aztertuz.

HAPen alorrean eratu diren testu-bildumei dagokienez, denetariko corpusak osatu dira gaur artean, hainbat hizkuntzatan. *British National Corpus (BNC)* da mundu mailan erreferentzia-corpus nagusienetakoa, ingelesarekin loturiko azterketak

egiteko gehien baliatzen dena; askotariko testuek osatzen duten 100 milioi hitzeko bilduma da, elebarrak eta sinkronikoa, ahozko hizkuntzaren eta idatziaren laginak biltzen ditu, eta TEI (*Text Encoding Initiative*) testu-kodeketarako gidalerroen arabera etiketatuta dago. AEBetako ingelesaren ereduak, berriz, *American National Corpus* (ANC) deritzon 22 milioi hitzeko testu-bildumak osatzen du. *FRANTEXT* izena du frantsesaren erreferentzia-corpusak eta 150 milioi hitzez osatuta dago. *Corpus de Referencia del Español Actual de la Real Academia Española* (CREA) da, bestalde, gaztelaniazko erreferentzia-corpus nagusia, gaztelaniaz hitz egiten den hainbat zonaldean azken hogeita bost urteetan ekoizitako testuen bilduma da, 125 milioi bat hitz etiketatuk osatzen dutena. Hizkuntza gutxitu batzuetan ere eratu dira erreferentzia-corpusak. Txekieraren kasuan, esaterako, 100 milioi hitzeko corpus zabala bildu dute (*SYN2000*). Corpus gaelikoa, berriz, txikia da, 15 milioi hitzek osatzen dute. Espainia mailan, *Corpus de Referencia do Galego Actual* (CORGA) garatu da Galizian, 13,3 milioi hitz etiketarekin; eta *Corpus Textual Informatitzat de la Llengua Catalana* (CTILC) deritzon 52 milioi hitzeko testu-bilduma daukate etiketatuta Katalunian. Beste hainbat hizkuntzatarako (poloniera, hungariera, errusiera, greziera, eslovakiera, txinera, kroaziera...) eta askotariko helburuekin erreferentzia-corpus gehiago ere eratu dira.

Gure hizkuntza-komunitatean XX. mendeko euskararen corpus estatistikoa (*XXMECE*)¹⁴ da euskarazko erreferentzia-corpusa izan zitekeenetik gertuen dagoen testu-bilduma; 6.351 idazlanetatik hartutako 4.658.036 hitzek osatzen dute eta lematizatuta dago. Beste hizkuntzekin alderatuta, euskararen erreferentzia-corpusa txikia da, “neurriarengatik eta, batez ere, eskuragarritasun ezagatik, ez da nahikoa” (IXA taldea eta Elhuyar Fundazioa, 2007). Ondorioz, hainbat eragilek egin dute jada 50 milioi bat hitzeko testu-bilduma jaso eta euskarazko erreferentzia-corpus sendo bat garatzeko aldarrikapena. Halaber, *Ereduzko Prosa Gaur* (EPG)¹⁵ 2000-2006 urte bitartean 235 literatur liburutatik eskuratutako 10,2 milioi hitzek eta prentsarik (*Berria* egunkaritik eta *Herria* aldizkaritik) eskuratutako 9,6 milioi hitzek osatzen duten corpus orokorra da, lema eta kategoria mailan etiketatuta dagoena.

¹⁴ <http://www.euskaracorpora.net/XXmendea/>

¹⁵ <http://www.ehu.es/euskara-orria/euskara/ereduzkoa/>

*Zientzia eta Teknologiaren Corpora-k (ZTC)*¹⁶, berriz, 1990-2002 urte bitartean zientzia eta teknologiaren alorrean euskaraz argitaratu diren obren bilduma adierazgarria izatea du helburu, 7,6 milioi hitzez osatuta dago eta lema, kategoria eta azpikategoria mailan etiketatuta. Era berean, 300 mila hitzeko *Euskararen Prozesamendurako Erreferentzia Corpora (EPEC)* (Aduriz *et al.*, 2006a), maila morfologikoan (Aldezabal *et al.*, 2007a), sintaktikoan (Aldezabal *et al.*, 2007b) eta semantikoan (Agirre *et al.*, 2005) etiketatuta dagoena, baliatzen da IXA taldean euskararen tratamendu automatikorako. Horiez gain, beste corpus batzuk ere eratu dira euskaraz, askotariko helburuak eta ezaugarriak dituztenak (Areta *et al.*, 2008).

Tesi-lan honetan, baina, erredeen eta desbideratzeen analisia dugu mintzagai nagusi. Analisi horretarako, hizkuntza-ikasleen corpusak baliatu ohi dira batik bat. Corpus hauek bereziak dira, ekoizle mota jakin batek idatzitako testuez osatzen baitira. Azken urteotan, ikerketa ugari egin dira hizkuntza-ikasleen corpusen inguruan (Meunier, 1998; Pravec, 2002; Tono, 2003; Nesselhauf, 2004; Myles, 2005; Díaz-Negrillo eta Fernández-Domínguez, 2006; besteak beste) eta baliabide sendoak garatu dira hainbat hizkuntzatan:

- *International Corpus of Learner English (ICLE)* da alor honetan Europa mailan dagoen proiektu nagusienetako bat (Granger, 1993, 1994, 1996, 2002). Hainbat unibertsitatek parte hartzen dute eta bi milioi hitzetik gorako corpora osatu da ingelesa ikasten ari diren hamabost herrialde desberdinetako ikasleen testuekin. Batez ere goi-mailako ikasleen testuak bildu dira, bi xede nagusirekin: ikasleek egiten dituzten errore-adibideak aztertuz hizkuntza-ikasleen tarte-hizkuntza(k) ezagutzea eta ama-hizkuntza desberdina duten ikasleen erroreak alderatuz erroreak unibertsalak ala hizkuntza jakin bati dagozkionak diren aztertzea.
- *Polish-English Language Corpus Research and Applications (PELCRA)* ama-hizkuntza poloniera duten ingeleseko ikasleen testuekin osatutako corpora da, 500 mila hitz ingurukoa (Uzar, 1997; Mason eta Uzar, 2000). Hizkuntzaren i(r)a kaskuntza prozesua ikertzea eta horren arabera material egokiak garatzea da corpus horren helburu aplikatu nagusia.

¹⁶ <http://www.ztcorpusa.net/cgi-bin/kontsulta.py>

- *Cambridge Learner Corpus (CLC)* ama-hizkuntza desberdina duten ingeleseko ikasleek idatzitako testuekin osatutako 20 bat milioi hitzeko corpora da (http://www.cambridge.org/elt/corpus/learner_corpus.htm), eta erabilera komertziala dauka: hizkuntza-ikasleen adibideak kontuan hartzen dituzte ingelesa i(ra)kasteko hiztegiak, testu-liburuak edota bestelako baliabideak garatzeko.
- *Longman Learners' Corpus (LLC)* 10 milioi hitz inguru dituen ingeleseko ikasleen testu-bilduma da. Maila guztietako eta hainbat ama-hizkuntzako ikasleen ekoizpenak biltzen ditu hiztegiak sortzeko helburu nagusiarekin (<http://longmanusahome.com/dictionaries/learner.php>).
- Ama-hizkuntza txinera duten ingeleseko ikasleen testuekin, berriz, *HKUST Corpus of Learner English* (Flowerdew, 1996; Flowerdew, 1997; Milton, 1998; Milton eta Tsang, 1993) corpora eratu da, 25 milioi bat hitzekoa. Hori da ama-hizkuntza bakar bateko ikasleen testu-bildumarik handiena eta erabiltzaileen beharrak kontuan hartzen dituzten baliabide didaktikoak edota laguntza-tresnak sortzeko helburuarekin eratu da.
- *Chinese Learner English Corpus (CLEC)* milioi bat hitzez osaturiko corpora da, ama-hizkuntza txinera duten maila desberdinetako ingeleseko ikasleek idatzitako testuekin osatua. 61 errore mota etiketatu dira testu-bilduma horretan hainbat ikerketa egiteko.
- *JEFL* (Tono eta Aoki, 1998; Tono, 2000; Tono, 2002) ama-hizkuntza japoniera duten ingeleseko ikasleen testu-bilduma da, 700 mila hitz ingurukoa. Corpus horretan hizkuntza-maila guztietako ikasleen testuak biltzen dira, ikasleen erroreak eta tarte-hizkuntza aztertzeko helburuarekin.

Ingeleseko ikasleen testu-bildumez gain, Freetext proiektuaren barruan ama-hizkuntza desberdinetako frantseseko ikasleek idatzitako testu libreekin FRIDA corpora osatu da (Granger, 2003), 450 mila hitzekoa. Corpus horren zati batean (300 mila hitz inguru) 46.241 errore-adibide etiketatu dira, eskuz, erroreengintza inguruko datu estatistikoak eskuratzeko, errore mota jakinen inguruko azterketak egiteko edota ikasleen zailtasunak kontuan hartzen dituzten laguntza-tresnak garatzeko. FALCO (*Fehlerannotiertes Lernerkorpus* ‘error-annotated learner corpus’), berriz, alemaneko ikasleen testuak biltzen dituen corpora da, 36 mila bat hitz ingurukoa eta

errore-adibideekin etiketatua. Gaztelaniako ikasleen corpusari dagokionez, WOSLAC proiektuaren barruan osatze-bidean dago *Corpus Escrito del Español L2 (CEDEL2)*, ama-hizkuntza ingelesa duten gaztelaniako ikasleek idatzitako testuen bilduma, 400 mila hitz ingurukoa (Lozano, 2008). Eta ICLE proiektuaren barruan (Granger *et al.*, 2002) gaztelania ikasten ari diren ikasleen 125 mila hitzeko corpusa dago.

Asko dira hainbat hizkuntzatan osatu diren hizkuntza-ikasleen corpusak (Pravec, 2002; Tono, 2003; Díaz-Negrillo eta Fernández-Domínguez, 2006, besteak beste). Corpus horien guztien ezaugarriak (tamaina, ikasleen ama-hizkuntza eta xede-hizkuntza, hizkuntza-maila, etab.) desberdinak diren arren, denek dute helburu bera: hizkuntza-ikasleen testuen analisia; izan ere, corpus horiek informazio-iturri aberatsak dira hizkuntzaren i(ra)kaskuntza prozesuan lagungarriak izan daitezkeen tresnak garatzeko, ikasleen zailtasunak eta beharrak kontuan hartuta.

IV.5. Euskarazko erroreen eta desbideratzeen analisirako corpusak

Euskal komunitatean bizi dugun kasuistika soziolinguistikoa zabala da: hainbat hiztun mota ditugu (euskal hiztun arruntak, euskara-ikasleak, hizkera espezializatuetakoa ikasleak...) eta multzo bakoitzean historia soziolinguistiko ugari aurki daitezke (euskaldun goiztiarrak eta berantiarrak, euskalki jakin batean euskara batuan baino hobeto moldatzen direnak, batuaz euskalkian baino hobeto hitz egiten dutenak, hizkuntza ikasten urteak daramatzatenak, hizkuntza ikasten hasi berriak direnak, etab.). Batzuek gehiago, beste batzuek gutxiago, mota batekoak zein bestekoak, arrazoi batzuegatik edo beste batzuegatik, baina guztiok egiten ditugu erroreak, akatsak edota desbideratzeak. Izan ere, hizkuntzek askotariko kategoria, erregistro, jakintza-alor eta dialekto dauzkate eta zaila da hizkuntza bat alor guztietan menderatzea. Euskararen kasuan, gainera, hizkuntza normalizazio-prozesuan murgilduta egoteak erroreak egiteko aukerak areagotu egiten dizkigu.

Garatzen ditugun tresnak eraginkorrak izatea nahi badugu, ezinbestekoa da gure hizkuntza-komunitateak bizi duen egoera soziolinguistikoa kontuan hartzea, erabiltzaile desberdinek egindako erroreak edota desbideratzeak tratatzeko gai diren tresnak prestatzea. Horregatik, hain zuzen, euskarazko erroreen eta desbideratzeen analisirako ekoizleen arabera antolatutako corpusak baliatzen ditugu;

euskara-ikasleen, espezialitate-euskarako ikasleen eta euskal hiztun arrunten testuak, hain zuzen.

Euskara-ikasleen corpora da erroreen eta desbideratzeen azterketarako gure abiapuntu nagusia, bertan baitago adibide kopuru handiena. Hainbat euskaltegi, ikasturte, hizkuntza-maila, ikasle eta idazlan motak (laburpenak, eskutitzak, asmatutako ipuinak, etab.) osatutako testu-bilduma interesatzen zaigu. Hala, corpus hori funtsezko baliabidea da hizkuntza-ikasleen tarte-hizkuntza(k) aztertu edota erabiltzaileen benetako zailtasunak eta beharrak kontuan hartuta hizkuntzaren i(ra)kaskuntza prozesua bideratzeko laguntza-tresnak garatu ahal izateko. Izan ere, testu horiek aukera ematen digute ikasleek maila bakoitzean egiten dituzten errore edota desbideratze ohikoenak ezagutzeko, dituzten zailtasun nagusiak zein diren jakiteko, urrats batetik bestera gertatzen diren aldaketak alderatzeko, maila bakoitzari dagozkion ezaugarriak aztertzeke eta konparatzeko, ikasleek saihestu ohi dituzten egiturak antzemateko, etab. Ikasleen testuetan etiketatutako adibideak baliagarriak zaizkigu, halaber, erroreen tratamendu automatikoa bideratzeko, alor horretan ikuspuntu tekniko batetik tratatzen ditugun arren.

Espezialitate-euskarako ikasleei dagokienez, hiztun horiek duten euskararen ezagutza-maila askotarikoa da: batzuentzat ama-hizkuntza da euskara, beste batzuentzat beranduago eta arrazoi desberdinengatik ikasitako hizkuntza, batzuek eguneroko bizitzan erabiltzen dute, beste batzuek unibertsitateko eskoletan baino ez, etab. Hala ere, euskararen erabilera arrunta ongi ezagutu edo ez ezagutu arren, Euskara Teknikoa ikasten ari dira, menderatzen ez duten hizkera berezitu bat. Ondorioz, testu horietan bi errore edota desbideratze mota detektatu ahal dira: euskal hiztunak diren aldetik, edozein hiztunek egin ditzakeen hanka-sartze arruntak; eta komunikazio espezializatuarekin lotutako hizkuntza-berezitasunak ikasten ari diren aldetik, alor horretan guztiz trebatuta ez egoteagatik egin ditzaketenak. Testu horiek adierazgarriak dira espezialitate-alor bakoitzean egin ohi diren errore eta desbideratze mota usukoenak zein diren aztertzeke, arlo espezializatu bakoitzean sortzen diren arazo edota zailtasun nagusiak zein diren jakiteko, terminologia edota egitura bereziak ongi erabiltzen ote diren ikusteko, etab. Hizkuntzen i(ra)kaskuntzaren alorrean adar interesgarria da hizkera berezituen i(ra)kaskuntza aztertzea. Tesi-lan honetan, ordea, ez gara horretan sartuko, eta

espezialitate-euskarako ikasleen testuetan etiketatutako adibideak errorean tratamendu automatikorako baliatu ditugu soilik.

Zuzentzat hartzen diren euskal hiztun arrunten testu batzuk aztertzea ere komeni izaten da, hiztunok ere tarteka egiten baititugu hanka-sartzeak. Testu horietan, baina, errorean kopurua txikia izan ohi da. Horregatik, errorean eta desbideratzean tratamendu automatikoaren alorrean batez ere alarma faltsuen¹⁷ kopurua ebaluatzeko zaizkigu baliagarriak testu horiek, hau da, zuzenak diren elementuak erroretzat hartzen ez direla ziurtatzeko.

Beraz, euskara-ikasleen testuak errorean tratamendu automatikoaren alorrean eta ordenagailuz lagundutako hizkuntzen i(ra)kaskuntzarenean baliatzen ditugu; espezialitate-euskarako ikasleen idatziak eta zuzentzat hartzen diren hiztun arrunten testuak, berriz, errorean tratamendu automatikoaren alorrean bakarrik. Izan ere, alor horretan euskararako ortografia-, gramatika- eta estilo-zuzentzailea edota aldaki dialektalen markatzailea bezalako tresnak garatzea da xede nagusia. Horretarako, edozein hiztunek egindako errorean ahalik eta adibide gehien biltzea interesatzen zaigu, adibide bakoitzari dagokion informazio linguistikoarekin eta teknikoarekin batera. Ordenagailuz lagundutako hizkuntzen i(ra)kaskuntzaren alorrean, berriz, hizkuntzaren ikasketa-prozesua ezagutzea da helburua: zein errore edota desbideratze mota egiten dituzten ikasleek, zein egitura motarekin izan ohi dituzten zailtasun gehien, zein fenomeno linguistiko errepikatzen dituzten behin eta berriz, zein egitura saihestu ohi dituzten, etab. Alor horretan, beraz, euskara-ikasleen testuak aztertzen ditugu, datu psikolinguistikoak ere kontuan hartuz.

Ekoizleen arabera antolatutako corpusek hiztun mota bakoitzaren arabera azterketak egiteko eta hiztun mota bakoitzaren arabera baliabideak sortzeko aukera eskaintzen digute: euskara-ikasleei, esaterako, errorearen detekzioa, diagnosia (linguistikoa edota psikolinguistikoa) eta zuzenketa interesatuko zaizkie, erroreak ezjakintasunagatik egiten baitituzte gehienetan; hizkera berezitetako ikasleei, berriz, errorean detekzioa, zuzenketa eta ikasten ari diren espezialitate-hizkeraren berezitasunei lotutako fenomenoak interesatuko zaizkie batik bat; eta euskal hiztunei,

¹⁷ Alarma faltsu esaten zaie automatikoki errore gisa markatu diren egitura zuzenei.

azkenik, diagnosia baino gehiago interesatuko zaie errorearen edota desbideratzearen detekzioa eta zuzenketa, erroreak ezjakintasuna ez den beste faktore batzuen egiten baitituzte normalean.

IV.5.1. Euskara-ikasleen testuak identifikatzeko kode-sistema

III.2. atalean azaldu dugun moduan, *Ikasleak* datu-basea IRAKAZI web aplikazioaren bitartez atzitzen da. Aplikazio hori testuetan dauden adibideak eta adibide horiei dagokien informazioa (linguistikoa eta psikolinguistikoa) jasotzeko prestatuta egoteaz gain, hizkuntza-ikasleen testuak biltegitratzeko bitarteko aproposa ere bada: IRAKAZIren bidez, hizkuntza-irakasleek ikasleen testuak gordetzeko aukera daukate. Garrantzitsua da, baina, corpora bildu eta antolatzeko irizpide batzuk finkatzea. Guri, esaterako, egokia iruditu zaigu oinarritzko datu batzuk (euskaltegia, urtea, hizkuntza-maila, ikaslea eta ariketa mota) testu-fitxategiaren izenean esplizituki adieraztea, horrek testu-bilaketak edota kontsultak modu errazean egiteko aukerak eskaintzen baitizkigu.

Horrela, bada, euskara-ikasleen testuak izendatzeko kode-sistema bat asmatu dugu, Maritxalar (1999) tesi-lanean proposatutako kodeketa abiapuntutzat hartuta. Kodea lau eremuk osatzen dute:

- 1) testuaren jatorria (euskaltegia) eta urtea
- 2) hizkuntza-maila
- 3) ikaslearen identifikazio-hizkia(k)
- 4) ariketa mota

Jatorriari eta urteari dagokion lehen eremuan, euskaltegien izenen laburdurak (lehenengo bi hizkiak) eta urteari dagozkion azken bi zenbakiak erabiltzen ditugu. Ikasleen hizkuntza-maila zehazten dugu jarraian, ikasleak zein urratsetan dauden. Hirugarren eremua ikasleari dagokio, eta identifikazio-hizki batek osatzen du, ikasleen anonimotasuna gordetze aldera. Azkenik, ariketari dagokion eremuan ere identifikazio-hizki bat sartzen dugu. Edozein eremutan datuen bat falta zaigunean, *or* (*orokorra*) letrak eta zenbaki bat (kasuan kasu dagokiona) erabiltzen ditugu (1. taula):

Kodea	Azalpena
ae98_2_a_a	aek 1998_2. maila_a gisa identifikatutako ikaslea_a ariketa
il92_7_g_a	ilazki 1992_7. maila_g gisa identifikatutako ikaslea_a ariketa
il97_10_b_m	ilazki 1997_10. maila_b gisa identifikatutako ikaslea_m ariketa
tr05_2_or5_e	trintxerpe 2005_2. maila_or5 gisa identifikatutako ikaslea_e ariketa

1. taula. Euskara-ikasleen testu-fitxategiak identifikatzeko kodeen adibideak.

Kode-sistema horretan oinarrituta izendatu ditugu, beraz, euskara-ikasleen testu guztiak, fitxategika. Fitxategiaren izenak berak esplizituki adierazten digun informazioari esker, berehala jakin dezakegu, esaterako, zein euskaltegitako testuak dauzkagun, euskaltegi eta hizkuntza-maila bakoitzeko zenbat testu ditugun, ikasle beraren zenbat lan dauzkagun edota ariketa bera zenbat ikaslek egin duten. Aplikazio horrek, beraz, hizkuntza-ikasleen lanak edota ikasleen inguruko informazioa biltzeko, antolatzeko eta corpus gaineko azterketak egiteko aukerak eskaintzen dizkigu. Hau ekarpen garrantzitsua da ordenagailuz lagundutako euskararen i(r)a kaskuntzaren alorrean, bai ikertzaileontzat eta bai hizkuntzen i(r)a kaskuntzan diharduten adituentzat ere (irakasleentzat, material prestatzaileentzat...).

Euskara-ikasleen testuez gain, espezialitate-euskarako ikasleek eta hiztun arruntek idatzitako testu batzuk ere baliatzen ditugu gure azterketetan. Testu mota horiek, baina, ez ditugu aipatu kode-sistemaren arabera antolatzen; testuen jatorriari dagokion izenarekin izendatzen ditugu, informazio hori nahikoa baitzaigu.

IV.5.2. Baliatutako corpusen inguruko datuak

Euskarazko erroreen edota desbideratzeen analisia egiteko, garrantzitsua da ahalik eta adibide gehien aztertzea. Adibideak eskuratzeko bi bide edo aukera daude: gramatika- eta estilo-liburuetan oinarritutako arauetatik abiatzea edo corpus errealetan detektatutako adibideetatik abiatzea. Guk iturburu biak baliatzen ditugu, baina batez ere testu errealetan detektatutako adibideak interesatzen zaizkigu; izan ere, garatuko ditugun tresnen erabiltzaileen benetako zailtasunak ezagutzeko aukera ematen digute corpusek, baita liburuetan esaten dena egiaztatzeko eta esaten ez denaren berri edukitzeko aukera ere. Corpusa da, beraz, gure azterketen abiapuntu nagusia eta, bereziki, euskara-ikasleen corpusa.

IXA taldean euskara-ikasleen corpusaren bilketari Maritxalarrek (1999) ekin zion. Hainbat euskaltegitako ikasleek idatzitako testuak biltzen hasi zen, paperean, eta banan-banan transkribatzen. Ordudanik, corpora osatzen jarraitu dugu. Baina gaur egun, jada, euskarri informatikoan jasotzen ditugu idazlanak, eta aurreko atalean azaldutako kode-sistemaren arabera izendatzen eta gordetzen ditugu. Bestalde, idazlan berriez hornitzen gaituzten lankide gehiago dauzkagu eta euskara-ikasleen corpusaren aniztasuna ere ugartuz doa. Ahalik eta corpus handiena eratzea da gure helburua, aztertu nahi ditugun fenomenoak aurkitzeko probabilitate handiagoa izango baitugu horrela.

Euskarazko erroreen eta desbideratzeen analisia egiteko euskara-ikasleen corpora interesatzen zaigu bereziki, testu horietan egon ohi baita aztertu beharreko adibideen kopururik handiena. Euskara-ikasleen testuekin batera, baina, espezialitate-euskarako ikasleen lanak eta zuzentzat hartzen diren euskal hiztun arruntentzako testu batzuk ere baliatzen ditugu.

Euskara-ikasleen testuak erroreen tratamendu automatikoaren eta ordenagailuz lagundutako hizkuntzen i(ra)kaskuntzaren alorrerako zaizkigu baliagarriak. 1990ean hasita, euskara-ikasleen 113.290 hitzeko corpora eratu dugu gaur artean, sei euskaltegiren parte-hartzeari esker (2. taula).

Euskaltegia	Hitz kopurua
A	40.325
B	28.092
C	23.449
D	10.311
E	7.956
F	3.157
Guztira	113.290

2. taula. Orain artean bildutako euskara-ikasleen corpora.

Euskara-ikasleen testuak *Helduen Euskalduntzearen Oinarrizko Kurrikuluan* (HEOK) (HABE, 1999) definitutako hizkuntza-mailen arabera antolatuta dauzkagu. HEOKek lau maila bereizten ditu, baliokideak direnak Europako Kontseiluaren ekimenez hizkuntzak ikasteko, irakasteko eta ebaluatzeko sortu den Europako Erreferentzia Marko Bateratuan (EEMB) definitzen diren hiru ezagutza-maila nagusiekin: oinarrizko erabiltzaileari dagokion A maila, erabiltzaile independenteari

dagokion B maila eta erabiltzaile gaituari dagokion C maila (maila bakoitzak, era berean, bi azpi-atal hartzen ditu: A1, A2, B1, B2, C1 eta C2). 3. taulan ikus dezakegu zehazki HABEn garai batean definituta zeuden 12 urratsen, HEOKek bereizten dituen lau hizkuntza-mailen eta Erreferentzia Marko Bateratuari dagozkion mailen arteko parekotasuna (Camacho eta Lonbide, 2005):

Urratsak (HABE, 1989)	HEOK (HABE, 1999)	Erreferentzia mailak (EEMB)
---	4. maila	C2
10-11-12	3. maila	C1
7-8-9	2. maila	B2
5-6	1. maila	B1
3-4		A2
1-2		A1

3. taula. Definitu izan diren hizkuntza-mailen arteko parekotasunak.

Guk jaso ditugun euskara-ikasleen testuak HEOKen definitutako maileri dagozkie; HEOKeko lehenengo hiru mailetakoa ikasleek idatzitako testuak dira, hain zuzen, euskarazko erroreen eta desbideratzeen analisirako baliatu ditugunak. Eratu dugun euskara-ikasleen corpusaren neurria zein den erakusten digu 4. taulak.

Hizkuntza-maila	Hitz kopurua
1. maila (behe-maila)	39.117
2. maila (erdi-maila)	42.219
3. maila (goi-maila)	31.954
Guztira	113.290

4. taula. Euskara-ikasleen corpusa, HEOKeko hizkuntza-maila bakoitzean.

Espezialitate-euskarako ikasleen corpusari dagokionez, alor espezializatuetakoa hizkerek badituzte ezaugarri propio batzuk (terminologia zehatza, egitura bereziak...) eta, beraz, testu horiek baliagarriak dira hizkera espezializatuari dagozkien berezitasunak edota ezaugarri propioak aztertzeko. Bi iturritatik jaso ditugu testuak: alde batetik, EHUko hainbat ikasketatan eskaintzen den Euskara Teknikoko ikasgaian matrikulatutako ikasleen idazlanak dauzkagu (19.391 hitzeko corpusa) eta, bestetik, Informatika Fakultateko ikasle batzuen Karrera Bukaerako Proiektuen (KBP) lehen bertsioak (243.782 hitz). Guztira, beraz, 263.173 hitzeko corpusa osatu dugu espezialitate-euskarako ikasleen testuekin. Tesi-lan honetan,

ordea, Euskara Teknikoko ikasleen testuak bakarrik etiketatu ditugu, aztergai dugun errore motaren kopurua oso txikia baita KBPetan.

Azkenik, zuzentzat hartzen diren euskal hiztun arrunten testu batzuk ere aztertu ditugu errorearen analisia egiteko; izan ere, errorearen tratamendu automatikoaren alorrean garatzen diren tresnak askotariko erabiltzaileak dituztenez, garrantzitsua da, hizkuntza-ikasleen erroreekin batera, hiztunek egin ditzaketen hanka-sartzeak ere aztertzea. Multzo horretan bildu ditugun testuen artean, mota edota hizkera desberdinetan idatzitako testuak daude, hala nola, posta elektronikoko mezu batzuk (533.728 hitz), ikastetxeetako ikasleen idazlanak (8.102 hitz), argitalpen pare baten lehen bertsioak (47.967 hitz) edota argitaratutako prentsa-ohar batzuk (1.821 hitz). IXA taldean garatutako hainbat tresnaren oinarri izan den *Euskaldunon Egunkaria* corpusaren zati bat ere baliatzen dugu euskarazko errorearen edota desbideratzearen analisia egiteko. Zuzentzat hartzen diren euskal hiztun arrunten testuetan, oro har, errore-adibide gutxi egon ohi da. Horregatik, testu horiek tratamendu automatikoaren alorrean baliatzen ditugu batik bat, egon daitezkeen erroreak detektatzeko eta, bereziki, alarma faltsuen kopurua zein den ikusteko. Tesi-lan honetan, hain zuzen, *Euskaldunon Egunkaria* corpuseko 53.658 hitzeko lagin bat erabili dugu determinatzaile-errorearen inguruan egin dugun azterketa eta prozesamendua ebaluatzeko.

5. taulak erakusten digu euskarazko errorearen eta desbideratzearen analisiak egiteko eskura ditugun testu-motak eta kopuruak:

Ekoizleak	Testu mota	Hitz kopurua
Euskara-ikasleak	Behe-mailako testuak	39.117
	Erdi-mailako testuak	42.219
	Goi-mailako testuak	31.954
	Maila guztiak	113.290
Especialitate-euskarako ikasleak	Euskara Teknikoko testuak	19.391
	KBPak	243.782
Hiztun arruntak	Posta elektronikoko mezuak	533.728
	Ikastetxeetako idazlanak	8.102
	Argitalpenen lehen bertsioak	47.967
	Prensa-oharrak	1.821
	Euskaldunon Egunkaria	53.658

5. taula. Erroreen eta desbideratzeen analisirako baliatzen ditugun corpusen inguruko datuak¹⁸.

Hainbat testu mota baliatzen ditugun arren, euskara-ikasleen corpora da euskarazko erroreen eta desbideratzeen analisirako gehien baliatzen duguna, bertan baitago aztergai ditugun adibideen agerpen kopuru handiena. Corpus hau irekia da eta handitzen, osatzen eta aberasten joango da testu berriak jaso ahala, euskara-ikasleen corpus sendo eta adierazgarri bat eskura izan arte. Hori da, behintzat, gure helburua.

IV.6. Laburbilpena

Hizkuntzaren Azterketa eta Prozesamendua (HAP) bideratzeko gaur egun egiten diren ikerketa askoren oinarri dira corpusak. Ekoizpen errealak biltzen dituen informazio-iturri aberatsa da corpora, denetariko fenomeno linguistikoen agertokia eta azterketa estatistikoak egiteko bitarteko aproposa. Erroreen edota desbideratzeen analisiak egiteko ere corpusak baliatu ohi dira, batik bat hizkuntza-ikasleen corpusak, testu horietan egon ohi baita aztertu beharreko adibideen kopururik handiena.

Corpusgintzak muga batzuk badituen arren, eskaintzen dizkigun abantailak handiak dira. Batetik, euskarri elektronikoan gordetako corpusei eta HAPen alorrean garatutako tresnei esker, testuen azterketa linguistikoak azkarrago egiteko eta emaitza fidagarriagoak lortzeko aukerak dauzkagu. Bestetik, corpusak oinarri hartuta

¹⁸ Letra lodiz daude tesi-lan honetan aztertu eta baliatu ditugun corpus-laginak.

egindako analisiei esker, adibide errealei dagozkien emaitzak lortzen dira. Erabiltzaileen beharrak edota zailtasunak kontuan hartzen dituzten tresnak garatzeko aukera eskaintzen digute, beraz, corpusek.

Tesi-lan honen ekarpen nagusietako bat euskara-ikasleen corpusaren bilketa eta antolaketa da. Irizpide jakin batzuen arabera antolatuta eta baliabide informatikoekin aztertzeko prest dagoen ikasleen corpora bi ikerketa-alorretan egiten ditugun azterketen abiapuntua da: erroreen tratamendu automatikoaren alorrean, euskararako ortografia-, gramatika- eta estilo-zuzentzailea edota aldaki dialektalen markatzailea bezalako tresnak garatzea dugu helburu eta ordenagailuz lagundutako hizkuntzen i(r)a(k)askuntzaren alorrean, hizkuntzaren i(r)a(k)askuntza bideratzen laguntzeko tresnak diseinatzea. Alor batean zein bestean egindako lanetan, garrantzitsua da gure hizkuntza-komunitateak bizi duen egoera soziolinguistikoa kontuan hartzea; izan ere, hainbat hiztun mota bizi gara elkarrekin eta tresna informatikoak erabiltzeko beharrak askotarikoak dira guztiontzat. Horregatik, hiru hiztun motek idatzitako testuak baliatzen ditugu erroreen eta desbideratzeen azterketa eta prozesamendua egiteko: euskara-ikasleen idazlanak ezinbestekoak zaizkigu hizkuntzen i(r)a(k)askuntza prozesua ezagutzeko eta erabiltzaileen beharren arabera baliabideak prestatu ahal izateko, eta baliagarriak zaizkigu, era berean, erroreen tratamendu automatikoa bideratzeko; espezialitate-euskarako ikasleen lanak eta hiztun arruntek idatzitako testuak, berriz, erroreen tratamendu automatikoaren alorrerako bakarrik erabiltzen ditugu, ongi detektatzen diren erroreak ikusteko eta alarma faltsuak baldin badaude saihesteko.

Euskara-ikasleen corpusari dagokionez, IRAKAZI aplikazioa prestatu eta kode-sistema bat asmatu dugu testuak bildu, antolatu eta modu errazean identifikatzeko. Horrela, testuak jasotzean baliatzen ditugun erreferentzia-kodeei esker, berehala jakin dezakegu, esaterako, zein euskaltegitako testuak dauzkagun, euskaltegi eta maila bakoitzeko zenbat testu ditugun, ikasle beraren zenbat lan dauzkagun edota ariketa bera zenbat ikaslek egin duten. Ordenagailuz lagundutako hizkuntzen i(r)a(k)askuntzaren alorrean, garrantzitsua da corpora irizpide jakin batzuen arabera antolatzea, horrek baldintzatuko baitu, hein handi batean, ikerketaren arrakasta edo porrota.

Tesi-lan honetan, beraz, euskara-ikasleen corpus sendo eta adierazgarri bat osatzen jarraitzeko irizpideak definitu ditugu, corpora baita euskarazko errorearen eta desbideratzearen alorrean egiten ditugun azterketen abiapuntu nagusia. Behin aztertu beharreko testu-bilduma eskuragarri izanik, etiketatze-prozesua da hurrengo urratsa. Testuetan dauden errorearen eta desbideratzearen diagnosi linguistikoa egiteko, hurrengo kapituluan aurkezten dugun euskarazko errorearen eta desbideratzearen sailkapena hartzen dugu oinarri.

Corpusetan detektatutako adibideen diagnosi linguistikoa egiteko euskarazko errorearen eta desbideratzearen sailkapena baliatzen dugu.

V. Euskarazko errorearen eta desbideratzearen sailkapena

V.1. Sarrera

Errorearen inguruko azterketak egiteko errorearen sailkapenak edota kategoria-sistemak hartu ohi dira oinarri. Hizkuntzaren Azterketa eta Prozesamenduaren (HAP) alorrean, errorearen sailkapen baten arabera etiketatutako adibideak aztertu eta automatikoki detektatzeko teknika egokienak bilatzen dira; alegia, sailkatutako adibideetan oinarrituta bideratu ohi da errorearen detekzio eta tratamendu automatikoa. Hizkuntzen i(r)a-kuntzaren alorrean ere errore-sailkapenak baliabide garrantzitsuak dira. Izan ere, etiketatutako errore mota bakoitzaren ezaugarriak, kopuruak edota maiztasunak aztertuz hizkuntza baten ikasketa prozesua ezagutzera hel gaitezke.

Errorearen sailkapenak testuetan dauden adibide erroredunen diagnosi linguistikoaren oinarri dira. Ikertzaile bakoitzak dituen helburuen arabera finkatu ohi ditu bere sailkapena eratzekeko irizpideak (Corder, 1971; Dulay *et al.*, 1982; Santos, 1993; Fernández, 1997; Bustos, 1998; James, 1998; Yang eta Akahori, 1998; Granger, 2003). Bustos-ek (1998), esaterako, errore-taxonomiak osatzeko bost ikuspuntu nagusi bereizten ditu:

Irizpide deskribatzailean oinarritzen diren sailkapenak errorearen azaleko egiturari erreparatzen diote, errorearen forma aztertzen dute. Lau errore mota nagusi bereizten dira: elementu linguistiko baten ezabaketa, gehikuntza, ordezkapena eta ordena-aldaketa. Hitz-mailan eragiketa horiek karaktereen gainean egiten dira; sintaxi-mailan, berriz, elementu sintaktikoen gainean. Sailkapen mota hau errorearen egitura deskribatzera mugatzen da, beste alderdi batzuk (errorea eragin duten arrazoi inter edo

intralinguistikoak, esaterako) kontuan hartu gabe. Eta horixe da, hain zuzen, hizkuntzen i(ra)kaskuntzaren alorrean irizpide honi egin izan zaion kritiketako bat, lortzen den informazioa (errorearen azaleko egituraren deskripzioa) ez dela nahikoa hizkuntzaren i(ra)kaskuntzaren prozesua ezagutu ahal izateko.

Irizpide linguistikoetan oinarritutako taxonomiek hizkuntza-sisteman bereizi ohi diren moduluen arabera sailkatzen dituzte erroreak; errore fonologiko, morfologiko, sintaktiko, morfosintaktiko, lexiko edota semantikoak bereizi ohi dira irizpide honetan oinarritutako sailkapenetan. Maiz erabiltzen den eredua da hau. Hala ere, ikuspuntu honek ere badauzka hutsuneak: gramatika-osagaiak bakarrik hartzen dira kontuan, pragmatikari edota diskurtsoari dagozkion osagaiei erreparatu gabe; gramatika ez da nahikoa, ordea, hizkuntza osotasunean aztertzeko.

Irizpide pedagogikoa jarraitzen duten sailkapen-tipologiak osatu izan dira, halaber, Corder-ek proposatutako kontzeptu hauek kontuan hartuta: errore sistematikoak edo ustekabeko hanka-sartzeak, errore fosilduak edo ez-fosilduak, garapeneko erroreak edo transferentziak eragindakoak. Sailkapen mota honi, ordea, erroreen taxonomia propio bat baino ondorio batzuen zerrenda hutsa izatea kritikatu izan zaio, eta ikuspuntu pedagogiko batetik proposamen hau laburra geratzen dela.

Irizpide komunikatiboan oinarritutako taxonomiak ere eratzen dira, Corder-ek, Burt eta Kiparsky-k (1974) eta Juozulynas-ek (1994) bereizten dituzten bi errore mota oinarri hartuta: errore lokalak (komunikazioa eteten ez duten hitz edota sintagma-mailako erroreak) eta errore globalak (esaldi osoaren egiturari edota esanahiari erasan eta komunikazioa oztopatzen dutenak). Ikuspegi komunikatiboari, baina, kritikatu izan zaio errorrea lokala edo globala den esatea errorearen larritasun-mailari dagokiola, eta ez errorrea deskribatzeari. Ondorioz, erroreak sailkatu eta deskribatu ostean gehitu beharreko informazio osagarria izan behar luke honek, irizpide komunikatiboan oinarritutako erroreen sailkapena, bestela, laburregi geratuko bailitzateke.

Irizpide etiologikoa edo *esplikatzailea*, azkenik, erroreak jatorriaren arabera multzokatzear datza; helburua errorearen zergatia azaltzea da. Bi errore mota nagusi bereizten dira ikuspegi honetatik: hizkuntza arteko erroreak (beste hizkuntza baten transferentzia egon baldin bada) edo hizkuntza barneko erroreak (errorearen jatorria ikasten ari den bigarren hizkuntza horretan bertan baldin badago). Erroreen azterketa eta prozesamenduaren alorrean, baliagarria izan daiteke errorearen jatorriaren inguruko

informazioa, baina hori kontuan hartzeak konplexutasuna gehitzen dio tratamendu automatikoari (Gojenola, 2000). Irizpide etiologikoari kritikatu izan zaio, halaber, ikertzaileen interpretazio subjektiboetan oinarritzen dela.

Hainbat irizpide jarraitu ahal dira, beraz, errore-taxonomia bat eraikitzeke garaian. Guztiak izan daitezke baliagarriak, bakoitzaren helburuen arabera; baina, era berean, guztiek dituzte hutsuneak. Horregatik, hain zuzen, gai horren inguruko adituek (Dulay *et al.*, 1982; James, 1998; Meunier, 1998; Tono, 2003; Granger 2003a; Díaz-Negrillo eta Fernández-Domínguez, 2006, besteak beste) sailkapen berean irizpide edo ikuspuntu bat baino gehiago uztartzea gomendatzen dute, sailkapena mugatuegia gera ez dadin.

Kapitulu honetan euskarazko erroreak eta desbideratzeak sailkatzeko garatu dugun kategoria-sistema aurkezten dugu (Aldabe *et al.*, 2005c). Sailkapenaren oinarrian hainbat lan daude eginak lehendik IXA taldean. Hasteko, duela urte batzuk hitz-mailako taxonomia bat osatu zen XUXEN zuzentzaile ortografikoa garatzeko (Aduriz, 1994). Behin errore ortografikoen detekzioa bideratuta, errore sintaktikoen tratamendu automatikoari ekin zitzaion, esaldi mailako errore mota batzuen azterketarekin (Gojenola, 2000; Oronoz, 2009). Orduan, erroreen sailkapen zehatz, osatu eta sendo baten beharra aurreikusi zen, euskarazko erroreen eta desbideratzeen azterketa eta prozesamenduaren alorrean bide berriak ireki eta ekarpenak egiten jarraitzeko. Horrela, euskaraz idatzitako testuetan detektatzen diren erroreak sailkatzeko kapitulu honetan aurkezten dugun taxonomia eratu dugu, irizpide linguistikoan eta deskribatzailean oinarrituta. Irizpide pedagogikoa, komunikatiboa edota etiologikoa, berriz, errore-adibideen kategoria linguistikoarekin edota azaleko egituraren deskripzioarekin nahastea ez zaigu egokia iruditu; beste maila bati dagozkion zehaztapenak direla uste dugu eta memento, behintzat, ez zaizkigu baliagarriak gure helburuetarako.

Euskarazko erroreen sailkapena definitzeko, bestalde, bibliografia zabala kontsultatu dugu: euskal gramatika-liburuetan (Txillardegí, 1978; Goenaga, 1980; Euskaltzaindia, 1985 eta 1993; Zubiri eta Zubiri, 1995; Laka, 1996; Hualde eta Ortiz de Urbina, 2003) proposatu diren gramatika-kategoriak aztertu ditugu zein kategoria eta azpikategoria bereizten diren ikusteko; beste hizkuntza batzuetan zehaztu diren sailkapen-ereduak (Suri eta McCoy, 1993; Juozulynas, 1994; Fernández, 1997; James, 1998; Becker *et al.*, 1999; Granger, 2003; Díaz, 2005; Díaz-Negrillo eta Fernández-Domínguez, 2006) alderatu ditugu guretzako baliagarriak izan zitezkeen ereduen bila; euskaltegietan jarraitu ohi

diren ikas-programak eta hizkuntza-maila bakoitzean i(ra)kaskuntzen diren gramatika-edukiak ere (HABE, 1999) kontsultatu ditugu, eduki horiek ikasleek egiten dituzten errore motei lotuta egon daitezkeelakoan. Eta azterketa bibliografikoaz gain, testu errealeko errore-adibideak ere kontuan hartu ditugu euskarazko errorearen eta desbideratzearen taxonomia bat finkatzeko; izan ere, corpus errealean egon ohi diren adibideen azterketa eta prozesamendua egin nahi badugu, ezinbestekoa da adibide horiei dagozkien kategoriak eta azpikategoriak sailkapenean lekua izatea.

V.2. atalean beste hizkuntza batzuetan egin diren errore-sailkapen batzuk aurkeztuko ditugu. Ondoren, V.3. atalean, euskarazko errorearen eta desbideratzearen diagnosi linguistikoa egiteko garatu dugun sailkapenaz hitz egingo dugu: sailkapenaren egitura azalduko dugu (V.3.1.), sailkapena osatzen duten kategoria nagusiak definituko ditugu (V.3.2.) eta sailkapena egituratzeko finkatu ditugun irizpideak aipatuko ditugu (V.3.3.). Euskarazko errorearen sailkapena hizkuntzalari batzuen lankidetzarekin ebaluatu dugu eta V.4. atalean aurkeztuko ditugu ebaluaziotik ateratako ondorioak. Amaitzeko, V.5. atalean, kapitulu honen laburbilpena egingo dugu.

V.2. Beste hizkuntza batzuetan egin diren errore-sailkapenak

Hizkuntza-sistema guztiak desberdinak dira, errorearen analisiaren alorrean ikertzaileok ditugun helburuak ere askotarikoak dira eta, ondorioz, detektatzen, etiketatzen eta aztertzen diren erroreak ere mota askotakoak dira. Ez dago, beraz, hizkuntza-sistema guztietarako baliagarria den sailkapen orokor edo unibertsalik; aitzitik, ikertzaile bakoitzak, dituen helburuen arabera, mota bateko edo besteko errore-sailkapena eraikiko du, bere ikerketara hobekien egokitzen den taxonomia propioa. Guk beste hizkuntza batzuetan egin diren sailkapenak aztertu ditugu:

- Hendrickson-ek (1979), analisi sintaktiko automatikoa helburu, 4 kategoria nagusik osatutako errore-sailkapena proposatzen du, irizpide linguistikoan oinarritua. Errore lexikoak, morfologikoak, sintaktikoak eta ortografikoak bereizten ditu (A eranskina). Sailkapen horrekin alderatuta, euskarazko errorearen sailkapenean, errorearen kategoria linguistikoa zehazteaz gain, errorearen azaleko egituraren deskribzioa ere egiten da.
- Suri eta McCoy-ek (1993) diote ordenagailuz lagundutako bigarren hizkuntzen i(ra)kaskuntzaren alorrean sistema eraginkorrak garatu ahal izateko ezinbestekoa

dela ikasleek egiten dituzten erroreak aztertzea, eta horretarako erroreen sailkapena oinarrizko bitartekoa dela. Horrekin batera, hizkuntza-transferentziak (*language transfer*) duen garrantzia ere azpimarratzen dute, lehen hizkuntzak eragin zuzena izan ohi duelako bigarren hizkuntzen i(r)a(k)askuntzan. Horrexegatik, ikertzaile horien sailkapena osatzen duten 22 kategoriak eta 40 azpikategoriak hizkuntza-transferentziaren eragina oinarri hartuta sortu dira. Gramatika-kategoriez gain, diskurtso mailan egiten diren erroreak sailkatzea ere garrantzitsua iruditzen zaie. Hala, errore mota horietarako beste sailkapen zehatzago bat garatu dute, 10 kategoria eta 12 azpikategoria dituen (A eranskina). Sailkapen horrekin alderatuta, guk ez ditugu errore-kategoriak hizkuntza-transferentziaren eragina kontuan hartuz sortzen. Guretzat hizkuntza-transferentziaren eragina errorea egin izanaren zergatiari dagokion faktorea da, eta erroreaken deskripzioa (linguistikoa eta azaleko egiturarena) eta zergatia maila desberdinetan bereizten ditugu.

- Juozulynas-ek (1994) alemana ikasten ari diren ikasle amerikarren erroreak sailkatu eta horien ikas-prozesua aztertzeko helburuarekin egin du sailkapena. Lehenik, errore lokalen eta globalen arteko bereizketa egiten du. Errore lokalek errorea dagoen osagaian eragiten dute soilik (hitz bakarrean) eta sailkagarriak dira; errore globalek, aldiz, esaldi osoaren egituran dute eragina. Errore global batzuk sailkatzea zaila izan ohi da, beste batzuk sailkatzea ezinezkoa, eta beste batzuk kategoria batean baino gehiagotan sailka daitezke (Krishna, 1975; Juozulynas, 1994). Beraz, errore mota batzuen etiketatzea (zein errore mota den finkatzea) zaila izan daiteke kasu batzuetan. Errore lokalen barruan lau errore mota bereizten ditu (errore morfologikoak, sintaktikoak, puntuazio-erroreak eta errore ortografikoak) eta irizpide deskribatzailean oinarritutako hainbat azpikategoria. Errore globalen barruan, berriz, morfologia, sintaxia eta ortografiari dagozkion kategoriak sartu ditu. Errore semantikoak aparte aztertzen ditu (A eranskina). Behin errorea lokala edo globala den zehaztuta, irizpide linguistikoa eta deskribatzailea uztartzen ditu Juozulynas-ek bere taxonomian. Bi irizpide horiek baliatzea da sailkapen horrek gurearekin duen berdintasun nagusia.
- Sónsoles Fernández-ek (1997) hizkuntzen i(r)a(k)askuntzaren prozesua aztertzeko helburuarekin garatu du bere errore-sailkapena. Lau kategoria nagusi bereizten ditu (errore lexikoak, gramatikalak, diskurtsiboak eta grafikoak), eta 57 azpikategoria. Azpikategoria bakoitzak, era berean, azaleko egituraren deskripzioan oinarritzen

diren beste hainbat azpikategoria barne hartzen ditu (A eranskina). Gure sailkapenean bezala, irizpide linguistiko eta deskribatzailea uztartzen dira hemen eta errorea egin izanaren zergatiak beste maila batean kokatzen dira.

- James-ek (1998) argi eta garbi aldarrikatzen du erroreen sailkapen berean bi irizpide edo ikuspuntu uztartzeak duen garrantzia: sailkapena sendoa eta zehatza izan dadin, ezinbestekoa da errore mota (lexikoa, sintaktikoa...) eta errorearen azaleko egituren deskripzioa (hitza(k) ezabatzea, gehitzea, ordezkatzeta edota ordenaz aldatzea) kontuan hartzea. Adierazpen hori izan zen errore-sailkapenen alorrean James-ek egindako ekarpen nagusia, eta ikertzaile askok egin dugu bat bere proposamenarekin. Errore mota edo kategoria nagusien artean errore ortografikoak, lexikoak, semantikoak, gramatikalak, diskurtsiboak edota pragmatikoak bereizten ditu James-ek, eta bakoitzaren barruan hainbat azpikategoria, erroreen azaleko egituraren deskripzioan oinarritutakoak (A eranskina). Euskarazko erroreen eta desbideratzeen sailkapenean ere irizpide edota ikuspuntu honi jarraitzen diogu.
- Bustos-ek (1998) ez dio halako garrantzirik ematen errorearen egitura formalari. Aitzitik, errorea eragin ahal izan duen kausak du berarentzat lehentasuna. Bere ustez, hainbat arrazoik sorraraz dezakete errore bat eta, ondorioz, egokiagoa iruditzen zaio sailkapenaren egitura erroreen zergatien arabera izatea. Horrela, irizpide etiologikoan edo esplikatzailean oinarritutako taxonomia bat proposatzen du, eta erroreak hiru atal nagusitan sailkatzen ditu: hizkuntza artekoak (beste hizkuntza baten transferentzia egon baldin bada), hizkuntza barnekoak (errorearen jatorria ikasten ari den bigarren hizkuntza horretan bertan baldin badago) eta sailkatu ezin direnak. Atal bakoitzak, era berean, hainbat azpikategoria barne hartzen ditu (A eranskina). Guk irizpide etiologikoa beste maila bati dagokiola uste dugu eta erroreak egin izanaren arrazoiak ez ditugu adibideen kategoria linguistikoarekin edota azaleko egituraren deskripzioarekin nahasten.
- Becker *et al.*-ek (1999), erroredun corpus etiketatua eta gramatika-zuzentzaile automatikoa helburu, errore-sailkapena eraikitzeke gramatika-liburuetan zehazten diren erroreak eta, oro har, gramatikaren ikuspegitik zailtzat hartzen diren egiturak hartzen dituzte kontuan. Hainbat kategoriak eta azpikategoriak osatzen dute sailkapen hori eta hauek dira azpikategoria nagusi batzuk: errore sintaktikoak, komunztadura-erroreak, hitzen ordenari dagozkion erroreak, azpikategorizazioari

dagozkionak, errore morfologikoak, ortografikoak edota puntuazio-erroreak (A eranskina). Sailkapen horren eta gurearen arteko desberdintasun nagusia errore-adibideen jatorria da: beraiek ez dituzte ikasleen corpusak aztertzen errore mota posibleak aurreikusteko; gramatika-liburuak soilik hartzen dituzte adibide erroredunen iturritzat.

- Granger-ek, bere ikerketa-taldearen laguntzarekin, lan handia egin du ordenagailuz lagundutako hizkuntzen i(r)a kaskuntzaren alorrean (Granger, 1998b, 2002, 2003 eta 2004; Dagneaux *et al.*, 2005, besteak beste). Frantsesa ikasten ari diren ikasleek egindako erroreak detektatu, sailkatu eta aztertzeko erroreen sailkapen sendo bat garatu dute, FreeText proiektuaren¹⁹ barruan. James-ekin (1998) bat eginez, gramatika-kategorietan eta azaleko egituraren deskripzioan oinarritutako errore-taxonomien baliagarritasuna nabarmentzen dute, sailkapen berean bi irizpide uztartzeak dituen abantailak azpimarratuz. Corpusetan errore-adibideak etiketatzeko hiru maila hartzen dituzte kontuan: errore mota (*error domain*), errore-kategoria (*error category*) eta hitz erroredunaren gramatika-kategoria (*grammatical category*). Errore mota da sailkapenaren maila nagusia eta 9 errore mota bereizten dituzte: errore ortografikoak, errore morfologikoak, gramatikalak, sintaktikoak, lexikoak, erregistroari dagozkionak, estilo kontuak, puntuazio-erroreak eta errore tipografikoak. Errore mota bakoitza, era berean, hainbat errore-kategoriaz osatuta dago; 36 kategoria bereizten dituzte guztira. Gainera, aparteko sailkapen bat oinarri hartuta, hitz erroredun bakoitzari gramatika-kategoria bat ere esleitzen diote, adjektiboa, adberbioa, artikulua, juntagailua, determinatzailea, izena, preposizioa, izenordaina, aditza, puntuazio-ikurra edota sekuentzia kategoriakoa den zehaztuz. Guztira 11 gramatika-kategoria eta 54 azpikategoria bereizi dituzte bigarren sailkapen horretan (A eranskina). Gure sailkapenak badu antzekotasunik taxonomia horren egitura edota irizpideekin, guk ere errore motak (ortografikoa, lexikoa, semantikoa...), gramatika-kategoriak (deklinabidea, determinatzailea, aditza, adjektiboa...) eta

¹⁹ www.latl.unige.ch/freetext/index.html

azaleko egituraren deskripzioari dagokion informazioa zehazten baititugu (V.3.1. atala).

- Díaz-ek (2005) gaztelaniarako gramatika- eta estilo-zuzentzaile automatikoa bideratzeko garatu du errorearen sailkapena. Hizkuntza-gaitasunari lotutako errorearen –*errores de competencia*– eta jarduneko errorearen –*errores de actuación*– arteko bereizketa egiten du, eta gaitasunari lotutako erroreari erreparatzen die bereiziki. Ikertzaile honek dio gaitasunari lotutako errorearen zergatia zein den asmatzea posible dela eta errorea eragin duten arrazoiengan oinarritutako sailkapen batek jasotzen duen informazioa beharrezkoa dela gramatika- eta estilo-zuzentzaileak garatzeko, erabiltzaileei diagnostiko argigarriagoak, aberatsagoak eta baliagarriagoak eskaini ahal izateko; aldiz, jarduneko erroreak aurreikustea eta bakoitzaren diagnostika egitea ezinezkoa da, errore horiek ustekabekoak direlako. Irizpide etiologikoa kontuan hartzen du, beraz, bereizketa nagusi hori egiteko. Horrekin batera, irizpide linguistikoa eta deskribatzailea uztartzen ditu errorearen sailkapenean. 5 kategoria nagusik (puntuazio-ikurrei dagozkien erroreak, errore lexikoak, sintaktikoak, komunztadura-erroreak eta errore semantikoak) eta 28 azpikategoriak osatzen duten errorearen sailkapena garatu du (A eranskina). Errore pragmatikoak ez ditu sailkatzen, maila horretako analisi linguistiko-konputazionalerako baliabideak urriak direlako oraindik. Sailkapen horren eta euskarazko erroreak sailkatzeko guk sortu dugun taxonomiaren arteko desberdintasun nagusia da erroreak sailkatzeko guk ez dugula irizpide etiologikoa kontuan hartzen.
- Euskararen kasuan, errorearen azterketaren inguruan lan gutxi egin da orain artean. IXA taldean egin diren lanez gain, Maia *et al.*-ek (2005) euskara-ikasleek egindako hainbat errore-adibide bildu eta multzokatu dituzte. Zortzi atal bereizi dituzte, hain zuzen, beraien errore-sailkapenean: deklinabide-arazoak, aditz-arazoak, perpauseko elementuen ordena, perpaus bakunaren barneko beste puntu batzuk, lokailuak, perpaus elkartuaren sintaxia, arlo lexiko-semantikoa eta ondo dauden arren beharbada zuzendu egingo liratekeen esaldiak tradizioa ezkutuan gelditzen delako. Atal nagusi horien barruan, azpialal gehiago daude: deklinabide-kasuak oker, adizki jokatu gabeak eta jokatuak, esapide kalkatu batzuk, zehar-galderak, nominalizazioak, erdarazko *que* menperatzaile batzuen eragina, etab. (A eranskina). Badirudi sailkapena testuetako errore-adibideetan oinarritzen diren errore-motek

osatzen dutela. Guri, ordea, garrantzitsua iruditzen zaigu irizpide jakin batzuetan oinarritutako taxonomiak eratzea. Hala ere, interesgarria izan da guretzat euskarazko erroreak sailkatzeko egin den beste kategoria-sistema bat kontsultatzeko aukera izatea; izan ere, erroreen analisia bezain konplexua den alorrean eta euskara bezalako hizkuntza txiki batean egiten diren ekarpenak alderatzea aberasgarria dela uste dugu.

Aztertutako hamar errore-taxonomien artean, ez dugu aurkitu irizpide bakar batean oinarritzen den sailkapenik, hau da, irizpide linguistikoan, deskribatzailean, etiologikoan, komunikatiboan edota pedagogikoan soilik oinarritutakorik. Aitzitik, sailkapen guztietan uztartzen dira ikuspegi bat baino gehiago. Sailkapen gehienetan, joera nagusia da lehenik eta behin errore-adibide bakoitza errore mota edota gramatika-kategoria baten barruan kokatzea eta ondoren bere ezaugarrien arabera deskribatzea. Errore ortografikoak, gramatikalak eta lexikoak sailkapen guztietan agertzen dira; fonetika, pragmatika edota diskurtso mailako erroreak, berriz, orain artean gutxi aztertu diren kategoriak dira.

Erroreak aztertzeko finkatu ohi diren irizpideak, oro har, ikerketa bakoitzaren helburuen arabera dira. Horregatik, hain zuzen, taxonomien egokitasuna edota baliagarritasuna ere ikerketa bakoitzaren helburuen arabera ebaluatu beharrekoa da (Tono, 2003). Helburuen arabera sailkapenak egin ohi direnez, asko dira erroreen analisiaren alorrean eratu diren sailkapenak, nahiz eta tesi-lan honetan hamar baino ez ditugun aipatu.

V.3. Euskarazko erroreen eta desbideratzeen sailkapena

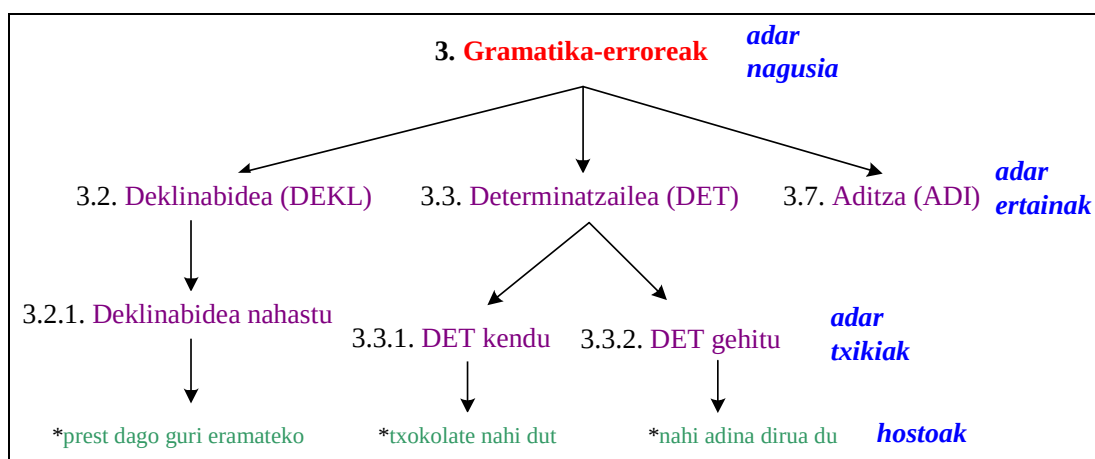
V.3.1. Sailkapenaren egitura

Euskarazko erroreak eta desbideratzeak sailkatzeko taxonomia hierarkiko eta dinamiko bat eraiki dugu. Sei kategoria nagusi bereizi ditugu:

1. Errore ortografikoak
2. Lexiko erroreduna
3. Gramatika-erroreak
4. Puntuazio-ikurrei dagozkien erroreak
5. Estilo kontu orokorrak
6. Errore semantiko-pragmatikoak

Kategoria nagusi horiek, aldi berean, gramatika-kategorietan edota azaleko egituraren deskripzioetan oinarritutako hainbat azpikategoriaz osaturik daude, sailkapenaren egitura definitzeko bi irizpide uztartu baititugu: linguistikoa eta deskribatzailea.

Euskarazko errorearen eta desbideratzearen sailkapenaren egitura hierarkikoa du: hainbat kategoria nagusik eta azpikategoria zehatzagoek osatzen dute, edo metaforikoki esanda, zuhaitz moduko hierarkia bat sortzen da: kategoria nagusiak eta horien barruko azpikategoriak zuhaitz baten maila desberdinetako adarrak lirateke (adar nagusiak, ertainak eta txikiak); eta zuhaitzeko hostoak, berriz, bildutako adibideak. 16. irudiak erakusten digu sailkapenaren egitura hierarkikoa, gramatika-erroreen kategoria eta azpikategoria batzuk adibide gisa hartuta:



16. irudia. Erroreen eta desbideratzearen sailkapenaren egitura hierarkikoa.

Sailkapena dinamikoa ere bada. Horri esker, corpusetan detektatzen ditugun adibideak oinarri hartuta, sailkapena osatzen eta zehazten joango gara; hau da, azpikategorien kopurua aldatuz joango da, adibide berriak detektatu ahala, edo metaforikoki esanda, hizkuntzalari konputazionalak historik gabeko adarrak kimatu ahal izango ditu, edota berriak txertatu, adar beretik zintzilik adar berri batean multzoka

daitezkeen hosto gehiegi baldin badaude. Dinamikotasun horri baliagarritasun handia ikusten diogu; izan ere, corpusen gainean lan egiten dugunez, edozein unetan ager daitezke sailkapenean gehitu, ezabatu edota aldatu beharreko azpikategoriak. Azken finean, sailkapenaren zehaztasun- eta sakontasun-maila (azpikategoria kopurua) corpusetan etiketatutako adibideek baldintzatzen dute: zenbat eta errore mota gehiago etiketatu, orduan eta zehatzagoa, sakonagoa eta osatuagoa izango da sailkapena.

Azpikategoria zehatzenei (edo zuhaitzeko hostoei) erroreak edota desbideratzeak etiketatzeko baliatzen ditugun kodeak dagozkie. Sailkapenean, beraz, adibide bakoitzari dagokion errore-kategoria, gramatika-kategoria edota azaleko egituraren deskripzioa zehazten ditugu, azpikategoria bakoitzari dagokion kodearekin eta adibide batekin, gutxienez. Tesi-lan honetan aztergai ditugun determinatzaile-erroreak, esaterako, gramatika-erroreen kategoriakoak dira, determinatzailea da dagokien azpikategoria eta azpikategoria horren barruan erroreen azaleko egituraren deskripzioaren arabera bereizten ditugu determinatzailea kentzeagatik, gehitzeagatik, errepikatzeagatik... egindako erroreak²⁰ (17. irudia):

Kategoria nagusia:
Gramatika-erroreak
Gramatika-kategoria:
Determinatzaile-erroreak
Errorearen azaleko egituraren deskripzioa:
DE Terminatzailea K entzea (DETK): * <i>txokolate nahi dut</i>
DE Terminatzaile atzizkia G ehitzea (DETG): * <i>nahi adina dirua dauka</i>
DE Terminatzailea E rrepikatzea (DETE): * <i>gizona bat etorri da</i>
...

17. irudia. Determinatzaile-erroreei dagokien egitura hierarkikoa.

Euskarazko erroreen eta desbideratzeen sailkapenaosatzen duten kategoria eta azpikategoria guztiak B eranskinean daude ikusgai.

²⁰ Determinatzaile-erroreen kategoriari dagozkion azpikategoria guztiak hurrengo kapituluaz azalduko ditugu xehetasunez, *Determinatzaile-erroreen sailkapena* atalean (VI.4. atala).

V.3.2. Sailkapeneko kategoria nagusien definizioa

Erroreen sailkapen bat osatzen duten kategoriak eta azpikategoriak definitzea eta antolatzea ez da askotan erraza izaten, ez baitago, oraindik, sailkapenak egiteko eredu estandar edo unibertsalik. Beraz, ikertzaile bakoitzak, dituen helburuen arabera, definituko ditu sailkapenaren egitura eta sailkatuko dituen errore motak: *“corpus researches have yet to agree on a general scheme of error annotation (...) shared tendencies may be observed but, in general, research groups often appear to design their own error tagging systems and explore different tagging models and error typologies”* (Díaz-Negrillo eta Fernández-Domínguez, 2006).

Guk ere gure helburuei hobekien egokitzen zaien sailkapena eraiki dugu, aztergai ditugun bi ikerketa-alarretan erroreen eta desbideratzeen diagnosi linguistikoa egiteko baliagarri zaiguna. Hainbat irizpide finkatu eta erabaki batzuk hartuta, sei errore mota nagusik osatzen duten sailkapena eratu dugu. Kategoria nagusi bakoitzaren barruan, halaber, gramatika-kategoriaren edota erroreen azaleko egituraren deskripzioaren arabera zehaztutako hainbat azpikategoria daude.

Bestalde, erroreak eta desbideratzeak bereizten ditugu guk (II.3.4. atala): ‘erroreak’ euskal gramatiketan zehaztutako edota Euskaltzaindiak ezarritako arauak betetzen ez dituzten egiturak dira, edozein kasutan okerrak izango direnak; ‘desbideratze’, berriz, testuinguru jakin batean egokiak ez diren egiturei esaten diegu. Erroreen eta desbideratzeen ezaugarriak desberdinak diren arren, kapitulu honetan aurkezten dugun sailkapenak biak barne hartzen ditu; izan ere, erroreen eta desbideratzeen tratamendu automatikoa bideratzea ekarpen garrantzitsua da gure hizkuntza-komunitaterako tresna eraginkorrak garatu nahi badira. Bada, erroreen detekzioa eta diagnosi linguistikoa ortografia- eta gramatika-zuzentzaileari dagokie; desbideratzeena, berriz, estilo-zuzentzaileari eta aldaki dialektalen markatzaileari.

Atal honetan azalduko dugu euskarazko erroreen eta desbideratzeen sailkapena osatzen duten kategoria eta azpikategoria nagusiak nola definitu ditugun, bakoitzari dagozkion kodeak eta adibideak zehaztuz²¹:

Errore ortografikoak arau ortografikoak ez errespetatzeagatik sortzen dira eta unitate gisa beti gaizki dauden lemak izango dira: “*ortografia bat eta bera da Baturako zein euskalkietarako erabaki dena; hemen ez da aukeretarako lekuri*” (Zuazo, 2000). Kategoria honetako azpikategoriak irizpide deskribatzailean oinarritzen dira, erroreen azaleko egitura deskribatuz: errore ortografikoak letra(k) kendu, gehitu, ordezkatu edota ordenaz aldatzeagatik sortzen diren erroreak dira:

- 1.1. **LEtra(k) Kentzea (ORT_LEK)**: *euki, *muitu, *biotza, *hoiek
 - 1.1.1. **LEtra Kentzea Hasieran, H (ORT_LEKHH)**: *erri, *emen, *artu
 - 1.1.2. **LEtra Kentzea Hasieran, E (ORT_LEKHE)**: *ta
 - 1.1.7. **LEtra Kentzea Bukaeran, R (ORT_LEKBR)**: *ze
- 1.2. **LEtra(k) Gehitzea (ORT_LEG)**: *haundi, *bainan, *zehatzki
 - 1.2.1. **LEtra Gehitzea Hasieran, H (ORT_LEGHH)**: *harrisku, *heman
 - 1.2.2. **LEtra Gehitzea Tartean, I (ORT_LEGHI)**: *laister
 - 1.2.3. **LEtra Gehitzea Tartean, U (ORT_LEGHU)**: *haundi
- 1.3. **LEtra(k) Ordezkatzeta (ORT_LEO)**: *eritzi, *inbiria, *ospatsuak, *araso
 - 1.3.2. **LEtra Ordezkatzeta Edozein tokitan, A beharrean E (ORT_LEOEAE)**: *erreztasun
 - 1.3.6. **LEtra Ordezkatzeta Edozein tokitan, H beharrean J (ORT_LEOEHI)**: *zijoala
 - 1.3.9. **LEtra Ordezkatzeta Edozein tokitan, X beharrean J (ORT_LEOEXJ)**: *keja
- 1.4. **LEtra(k) Ordenaz Aldatzea (ORT_LEOA)**: *lehioa, *leihaketa
 - 1.4.1. **LEtra Ordenaz Aldatu, H (ORT_LEOAH)**: *lehioa, *ohiana, *leihaketa

Errore ortografikoen artean sailkatzen ditugu, halaber, -A organikoa kentzeagatik egiten diren erroreak eta hitzak elkartzeagatik edota banatzeagatik egiten direnak:

- 1.5. **-A ORGANikoa kentzea (ORT_AORG)**: *Donosti, *panpin, *bidai
- 1.6. **Hitzak ELKartzea edo BANatzea**:
 - 1.6.1. **Hitzak ELKartzea (ORT_ELK)**: *hitzegin
 - 1.6.2. **Hitzak BANatzea (ORT_BAN)**: *bait du

Eta errore ortotipografikoak (hitz elkartuetan marratxoa gehitzea edo kentzea edota letra larrien edo xeheen erabilera okerra) ere kategoria honen barruan sailkatu ditugu:

²¹ Hemen sailkapena osatzen duten kategorien eta azpikategorien adibide batzuk emango ditugu; sailkapen osoa B eranskinean ikus daiteke.

1.7. Errore orto(tipo)grafikoak (ORTOTIPO):

1.7.1. Marratxoaren erabilera okerra, hitz elkartuetan:

1.7.1.1. Hitz elkartuetan MARRatxoa Kentzea (ORTOTIPO_MARK):

**jaun andre*

1.7.1.2. Hitz elkartuetan MARRatxoa Gehitzea (ORTOTIPO_MARG):

**giza-jarduer*

1.7.2. MAIuskula eta minuskulen erabilera okerra (ORTOTIPO_MAI-min):

**joxe garin, *. etxera*

Errore horiek ortografia-zuzentzaileari dagozkio. Errore ortografiko askoren detekzio automatikoa bideratu da jada IXA taldean, EDBL datu-base lexikala (Aldezabal *et al.*, 2001) eta bi mailatako morfologia (Koskeniemi, 1983) baliatuz. Ordenagailuz lagundutako hizkuntzen i(ra)kaskuntzaren alorrerako ikasleen testuetan detektatutako errore ortografikoen bilduma bat eskura izatea, berriz, baliagarria da euskara ikasten ari diren ikasleek dituzten arazo ortografikoak ezagutzeko eta horren arabera ortografia lantzeko aplikazioak garatzeko. Errore ortografikoen bilketa baliagarria izan liteke, era berean, errore horien tratamendu automatikoa hobetzen joateko eta errore ortografikoak detektatzeaz gain diagnosi linguistikoa ere egin ahal izateko.

Lexiko erroreduna kategoriak erroreak eta desbideratzeak barne hartzen ditu. Erroreek asmakuntzarekin edota sorkuntzarekin dute zerikusia, hau da, asmatutako edota gaizki sortutako hitzekin. Errore hauek ez dira ortografikoki zuzenak izango, baina ez hizki bat kendu, gehitu, ordeztu edo nahastu egin delako (errore ortografikoen kasuan gertatzen den moduan), hitza, unitatea, gaizki osatu delako baizik (2.1. azpikategoria). Lexiko erroredunari dagozkio, halaber, gaur egun jada erabiltzen ez diren hitz-formak (2.2. azpikategoria) eta maileguen egokitzapenean sortzen diren hitz erroredunak (2.3. azpikategoria). Errore horien tratamendu automatikoa bideratu ahal izateko, kategoria honi dagozkion adibideen bilduma bat egin beharko da:

2.1. Errore LEXikoa (LEX): **afaltzaile, *haurtoki*

2.2. Lexiko EZ-ERABilia (LEX_EZ-ERAB): **urrutizkina, *orlegi, *astegaitz*

2.3. Maileguen egokitzapen okerrak (LEX_MAIL): **afamatu, *heredentzia*

Kategoria honetan sartzen ditugu, halaber, desbideratze mota batzuk: aldaki dialektalak ez dagokien testuinguruan erabiltzea (2.4. azpikategoria):

2.4. Aldaki dialektalak (LEX_DIAL): **nere*, **haundi*, **barri*, **ipui*

Desbideratzeak aldaki dialektalen markatzaileari dagozkio. Forma estandarrek eta ez-estandarrek bereizi eta forma ez-estandarrek aldaki dialektal gisa sailkatzea eta tratatzea garrantzitsua da HAPen alorrean garatzen ditugun bitartekoak gure hizkuntza-komunitaterako baliagarriak izatea nahi baldin badugu. Horretarako, tresna informatikoei gai izan behar dute hizkuntza estandarrekin batera forma ez-estandarrek edota dialektalak ere ezagutzeko: esaterako, *gera* aditz-laguntzailea EZEZAG bezala analizatu beharrean, aldaera dialektal bati dagokion ADL (aditz-laguntzailea) gisa analizatu beharko litzateke, *gara* baliokide estandarrekin. Izan ere, modu horretara erabiltzaileari adierazi ahalko zaio forma dialektal hori euskara estandarra erabili behar den komunikazio-egoeretan egokia ez den arren, euskalki jakin bati dagokion aldakia dela, beste komunikazio-egoera batzuetan zuzena eta erabilgarria dena (II.3.4. atala).

Gramatika-erroreak gramatika-arauren bat hausteagatik sortzen diren erroreak dira, testuinguru morfosintaktikoan gaizki dauden gramatika-egitura okerrak. Kategoria nagusi honetan hainbat azpikategoria daude, irizpide linguistikoan eta deskribatzailean oinarritzen direnak:

3.1. Errore morfologikoak

3.1.1. Izen oinarriari dagozkionak: **erlien*, **etxian*

3.1.2. Aditzari dagozkionak: **leike*, **gordetu*, **jasatzen*, **dabiltzate*

3.2. Deklinabide-erroreak

3.2.1. DEKLabinabide-atzizkia NAHastea (DEKL_NAH): **autoarekin etorri gara*

3.2.2. DATA-egitura erroredunak (DEKL_DATA): **Donostia, 1995eko urtarilak 15ean*

3.3. Determinatzaile-erroreak

3.3.1. DETErminatzailea Kentzea (DETK): **txokolate nahi dut*

3.3.2. DETErminatzaile atzizkia Gehitzea (DETG): **nahi adina dirua dauka*

3.3.3. DETErminatzailea Errepikatzea (DETE): **mendia bat ikusten dut*

3.3.4. DSko osagaien Ordena okerra (DETO): **gutxi pertsona ginen*

3.3.5. MUGatasun kontuak (DETMUG): **edozein aldizkarietan irtetzen da*

3.3.6. DETErminatzailea/Zenbatzailea Nahastea (DETN): **asko beroa egin zuen*

3.3.7. DETG eta DETK sintagma berean (DETGK): **mozkorra handi hartu genuen*

3.3.8. Anbiguoak diren kasuak (DET?): **hamar garagardoak erosi nituen*

3.3.9. Zalantzazko kasuak (DET)

3.4. Izenordainen inguruan egindako erroreak

3.5. Adjektibo eta Adberbioen inguruko erroreak

- 3.6. Aposizio-erroreak
- 3.7. Postposizio-erroreak
- 3.8. Komunztadura-erroreak
 - 3.8.3. KOMunztadura eza Perpausean
 - 3.8.3.1. Aditza - Subjektua:
 - 3.8.3.1.1. NUMeroari dagokionean: **gizonek egin du*
 - 3.8.3.1.2. KASuari dagokionean: **zuk etorri zara*
 - 3.8.3.4. Aditza - Predikatua: **gure erleak oso soziablea dira*
 - 3.8.4. KOMunztadura eza Mendekoetan: **goxoki asko jaten duten umeak kariesa du*
- 3.9. Aditzaren inguruan egindako erroreak
 - 3.9.1. Aditzaren Denbora, Aspektua edota Modua nahastea
 - 3.9.2. Aditz-paradigmen nahasketa
 - 3.9.2.1. Nor eta Nor-NorK aditz-paradigmak nahastea: **ez da funtzionatzen*
 - 3.9.2.4. Nor-NorK eta Nor-NorI-NorK aditz-paradigmak nahastea: **Joni ikusi diot*
 - 3.9.2.5. Nor-NorI eta Nor-NorI-NorK aditz-paradigmak nahastea: **gustatzen dit*
- 3.10. Mendeko perpaus erroredunak
- 3.11. Juntagailuen eta lokailuen inguruan egindako erroreak
- 3.12. Perpausen egitura okerra (sintaxiari dagokionean)

Errore horien analisiak gramatika-zuzentzailea du helburu. Gramatika-errore batzuk tratatzeko hainbat azterketa egin dira jada IXA taldean (Gojenola, 2000; Oronoz, 2009; eta tesi-lan hau). Lan horien bidez, komunztadura-, postposizio- eta determinatzaile-errore batzuen tratamendu automatikoa bideratu da, baita data-egituretan egin ohi diren errore batzuen ere. Errore mota horien detekzio automatikoa urrats garrantzitsua da gramatika-zuzentzaileari begira.

Puntuazio-ikurren erabilerari dagozkion erroreak: puntuazio-arauen hausteari dagozkion adibideak sailkatzea eta aztertzea beharrezkoa zaigu puntuazio-ikurren inguruan egiten diren erroreak detektatu eta tratatu ahal izateko. Kategoria honek hartzen ditu, beraz, puntuazio-ikurra behar ez denean erabiltzeagatik (4.1. azpikategoria), behar denean ez erabiltzeagatik (4.2. azpikategoria) edota ikur okerrak erabiltzeagatik (4.3. azpikategoria) egiten diren erroreak:

- 4.1. **Puntuazio-ikurra** behar ez denean **Erabiltzea** (**PI_E**)
- 4.2. **Puntuazio-ikurra** behar denean **Ez Erabiltzea** (**PI_EE**)
- 4.3. **Puntuazio-ikurrak** **Ordezkatzea** edo **nahastea** (**PI_O**)

Tesi-lan bat bidean da, hain zuzen, puntuazio-ikurren erabilera okerraren inguruan: komaren erabilera desegokiak detektatu nahi dira ikasketa automatikoko teknikak erabiliz²² (Alegria *et al.*, 2006), euskararako gramatika-zuzentzailean integratzeko.

Estilo kontu orokorrak kategoriari dagozkio sintaxi mailan zuzenak izan arren, idazketaren estiloari dagokionean hobetu litezkeen egiturak. Kategoria honetan sartu ditugu esaldi luzeegiak erabiltzea (5.1. azpikategoria), hitzen ordena desegokia (estiloari dagokionean) (5.2. azpikategoria), hitz edota egitura berdinak maizegi errepikatzea (5.3. azpikategoria) edota egiturak ekiditea (5.4. azpikategoria) bezalako azpikategoriak:

- 5.1. **ESALdi LUZeegiak** (**ESALUZ**)
- 5.2. **HITZen ORDena, Estiloari dagokionean** (**HITZOR-E**)
- 5.3. **MAIZegi erabiltzen/errepikatzen diren hitzak/egiturak** (**MAIZ**)
- 5.4. **EKIDIIn diren egiturak** (**EKIDI**)

Estilo kontu orokorrak ezin dira erroretzat hartu, desbideratzetzat baizik. Eta interesgarria iruditu zaigu euskarazko erroreen eta desbideratzeen sailkapenean kategoria hau gehitzea, horrelako adibideak sailkatuz estilo-zuzentzailearen garapenerako informazioa eskura izango baitugu.

Azkenik, **errore semantiko-pragmatiko** esaten diegu esanahiari dagokionean testuinguru jakin batean zentzurik ez duten egiturei. Kategoria honetan daude hitz bat beste baten ordeztu erabiltzea (antzekotasunagatik, gehienetan) (6.1. azpikategoria), sasi-adiskideak (6.2. azpikategoria), pare dikotomikoak (6.3. azpikategoria), kalkoak (6.4. azpikategoria), gaizki eratutako esamoldeak (6.5. azpikategoria) eta errore pragmatikoak (6.6. azpikategoria):

²² Ikasketa automatikoko tekniken bidez, corpusak aztertzen dira ahalik eta informazio linguistiko gehien erauzteko, eta informazio horretan oinarrituta makinak komak non jarri behar diren ikasten du. Puntuazio-ikurren erabilera okerrari dagozkion erroreen detekzioa ere gramatika-zuzentzaileari dagokio.

- 6.1. Hitz bat beste batekin nahastea: *froga - proba, *haziera - hazkuntza
- 6.2. Sasi-adiskideak: *xelebre (célebre), *azienda (hacienda)
- 6.3. Pare dikotomikoak: *hura - ura; *ari - hari; *hasi - hazi, *atso - atzo
- 6.4. Kalkoak: *siesta bota, *ilea hartu
- 6.5. Esamolde okerrak: *lur eta zur
- 6.6. Errore pragmatikoak: jaten ari den bati *bejondeizula; gauean *egun on

Kategoria honetako erroreak dira automatikoki tratatzeko konplexuenak; izan ere, testuingurua edota informazio diskurtsiboa ezinbestekoak dira errore horien detekziorako, eta testu-mailako hainbat azterketa egitea ez zaigu posible oraindik. Hala ere, errore horiek sailkatzeko kategoriak eta azpikategoriak gehitu ditugu sailkapenean, adibide horien bilduma baliagarria izango baita, adibidez, hizkuntzaren i(ra)kaskuntzaren alorrean ikasleek izan ditzaketen arazo lexiko, semantiko edota pragmatikoak lantzen hasteko. Adibide horiek baliatu ahalko ditugu, halaber, estilo-zuzentzailea garatzen hasten garenean.

Oro har, ez dago errore-sailkapenak osatzeko jarrai daitekeen eredu estandarrik. Ondorioz, ikertzaile bakoitzaren helburuen arabera sailkapenak egin ohi dira. Horrelako taxonomiak eraikitzea ez da ataza erraza, ordea, zalantza ugari sortu ohi baitira. Gu saiatu gara euskarazko erroreen eta desbideratzeen sailkapena eratze alarik eta irizpide objektiboenak finkatzen, sailkapen erabilerraz bat definitzen. Izan ere, sailkapena osatzen duten kategoriak eta azpikategoriak dira testuetan etiketatzen diren adibideen diagnosi linguistikoaren oinarria eta hainbat tresna informatikoren abiapuntua. 6. taulak erakusten digu, hain zuzen, sailkapeneko kategoria nagusi bakoitzari dagokion tresna informatikoa zein den:

Erroreen eta desbideratzeen sailkapena	Tresna informatikoa
1. Errore ortografikoak	Ortografia-zuzentzailea
2. Lexiko erroreduna: Lexiko mailako erroreak Forma ez-estandarrak (desbideratzeak)	Gramatika-zuzentzailea Aldaki dialektalen markatzailea
3. Gramatika-erroreak	Gramatika-zuzentzailea
4. Puntuazio-ikurrei dagozkien erroreak	Gramatika-zuzentzailea
5. Estilo kontu orokorrak	Estilo-zuzentzailea
6. Errore semantiko-pragmatikoak	Estilo-zuzentzailea

6. taula. Erroreen eta desbideratzeen sailkapeneko kategoriak eta horien tratamendurako tresnak.

Ortografia-zuzentzailea garatuta, gramatika-zuzentzailearen garapena dugu orain esku artean. XUXENg-n integratuko diren hainbat errore motaren detekzio automatikoa bideratu da jada: komunztadura-erroreak, data-egituretan egindako okerrak, postposizio-lokuzioetan egindakoak (Ornoz, 2009) eta determinatzaile-erroreak (Uria *et al.*, 2009), besteak beste. Lan honetan landu dugu, hain zuzen, determinatzaile-erroreen detekzio automatikoa. Euskararako gramatika-zuzentzailea osatzeko ekarpena da, beraz, tesi hau.

V.3.3. Sailkapena egituratzeko irizpideak

Euskarazko erroreen eta desbideratzeen azterketa berbera bi ikerketa-alorretan ekarpenak egiteko baliatzen dugu guk: erroreen tratamendu automatikoan eta ordenagailuz lagundutako hizkuntzen i(r)a kaskuntzan. Alor batean zein bestean, erroreak eta desbideratzeak ditugu aztergai, corpus errealean detektatutako adibideak. Adibide horiek etiketatzeko garatu dugu, hain zuzen, euskarazko erroreen eta desbideratzeen sailkapena, irizpide batzuetan oinarrituta.

Errore-sailkapenak eraikitzeke, oro har, bi bide nagusi daude: erroreak aurrez finkatutako kategorietan sailkatzea, bata; kategoriak errore-adibideen arabera sortzea, bestea (Norrish, 1981). Guri aukera biak konbinatzea iruditu zaigu egokiena: sailkapeneko kategoria edota azpikategoria batzuk definitzeko gramatika-liburuetan zehaztutako kategoriak hartu ditugu abiapuntutzat; eta hortik aurrera, corpusetan detektatzen ditugun adibideen arabera goaz euskarazko erroreen eta desbideratzeen sailkapena osatzen eta zehazten.

Sailkapenean, bestalde, irizpide linguistikoa eta deskribatzailea uztartu ditugu, gramatika-kategoriak eta azaleko egituraren deskripzioak zehaztuz. Irizpide etiologikoak (errorea zerk eragin ote duen edo errorea egin izanaren arrazoiak), berriz, ez du lekurik hemen; izan ere, ezaugarri horiek maila sakonago bati dagozkiola iruditzen zaigu, diagnosi psikolinguistikoari, hain zuzen ere. Tesi-lan honetan ez gara alderdi psikolinguistikoaren analisisian sartuko, baina ordenagailuz lagundutako hizkuntzen i(r)a kaskuntzaren alorrean ikertzen jarraitzeko garatu ditugun tresna batzuetan azterketa psikolinguistikoak egiteko aukerak prest daude dagoeneko. Irizpide pedagogikoa eta komunikatiboa ere ez ditugu kontuan hartzen diagnosi linguistikoak egiteko.

Errore-sailkapenen zehaztasun mailari dagokionez, askotariko iritziak daude: batzuen ustez, taxonomia oso zehatzek hainbat errore mota gainjartzeko arriskua erakusten dute, eta adibide bera kategoria edota azpikategoria batean baino gehiagotan sailkatzeko aukera izateak zaildu egiten du errore-adibideak sailkatzea; beste batzuen iritziz, berriz, azpikategoriarik sortu ezean erabilgarriak ez diren sailkapen orokorregiak garatzeko arriskua sortzen da. Guk sailkapen orokor bat baino sailkapen zehatz bat egin nahi izan dugu; izan ere, testu errealetan detektatzen diren erroreen eta desbideratzeen analisia dugu helburu, eta corpusetan mota guztietako adibideak aurki daitezke. Bestalde, erabiltzaile mota desberdinentzat baliagarria den sailkapen bat eratu nahi badugu, sailkapenak erabilerraza izan behar du, irizpide objektiboetan oinarritutakoa; eta etiketatzaileei erroreen aurrean nola jokatu jakinarazi behar zaie, azpikategoria bakoitzari zein errore mota dagokion argi eta garbi azalduz.

Euskarazko erroreen eta desbideratzeen sailkapenari dagozkion irizpideak finkatzeko ezinbestekoa izan da hizkuntzalarien eta informatikarien arteko elkarlana. Izan ere, bi ikuspuntuak (linguistikoa eta konputazionala) uztartzea, alderatzea eta kontuan hartzea garrantzitsua da bi ikerketa-alorretan baliatuko dugun sailkapena garatu ahal izateko. Sailkapen hau, baina, ez da behin betikoa; aitzitik, etengabeko eraikitze-prozesuan jarraituko du: kategoriak eta azpikategoriak gehituz, kenduz, aldatuz edota zehaztuz joango gara, batik bat testuetan detektatzen ditugun adibideak oinarri hartuta.

V.4. Sailkapenaren ebaluazioa

Euskarazko erroreak eta desbideratzeak sailkatzeko garatu dugun sailkapena ebaluatu egin dugu, bi helburu nagusirekin: ebaluatzaileen ustetan sailkapenaren egitura egokia ote den jakitea eta ebaluatzaileek erroreen aurrean nola jokatzen duten ikustea. Ebaluazioa egiteko, hamaika hizkuntzalariri hogeita zortzi esaldi eman zaizkie, gramatika-errore bat edo gehiago zituztenak, eta adibide horiek sailkapeneko kategorien edota azpikategorien arabera etiketatzeko eskatu zaie. Guk ere, gure aldetik, esaldi horietako adibideak etiketatu ditugu, hizkuntzalariek guk aurreikusitako errore kopuruarekin eta errore bakoitzari esleitutako kategoriekin bat egin ote duten ikusteko (Aldabe *et al.*, 2005d).

Ebaluazioan zenbaki edota kopuru txikiak erabili ditugu: hogeita zortzi esaldi erroredun eta hamaika hizkuntzalariren parte-hartzea. Ebaluatzaile gehienek, gainera, ez dute esperientzia handirik errore-analisiaren alorrean. Kontuan hartu behar dugu, era

berean, sailkapena osatzen duten azpikategoria guztiak ez daudela sakon landuta, hainbat kategoria oraindik aztertzeke daudela, eskura dauzkagun bitartekoekin errore mota guztien azterketa aldi berean bideratzea ezinezkoa izan baita. Erroreak sailkatzean, bestalde, etiketatzailer bakoitzak bide bat edo beste jarrai dezake; alegia, etiketatze-lana subjektiboa izan daiteke. Eta horrek guztiak, noski, eragina izan lezake eskuratutako emaitzetan. Hala ere, sailkapenaren ebaluazio orokor bat egitea eta ebaluatzaileen iritziak ezagutzea garrantzitsua iruditu zaigu.

Sailkapenaren egiturari dagokionez, hizkuntzalariei galdetu diegu ea soberan dagoen edo faltan bota duten azpikategoriarik; sailkapena intuitiboa, erabilerraza, zehatzegia edota orokorregia iruditu zaien; aldaketak proposatuko ote lituzketen, etab. Horrez gain, testu errealeko adibideak sailkatzeko edota zuzentzeko garaian askotariko jarrerak egon ohi direnez, ebaluazio honen bidez aztertu nahi izan dugu etiketatzailerak errore guztiak sailkatzeko joera duten ala larrienei bakarrik erreparatzen dieten; errore bakoitza kategoria edota azpikategoria posible guztietan sailkatzen duten ala kategoria edo azpikategoria bakar batean sartzearekin konformatzen diren; erroreak zuzendu ondoren edota zuzendu aurretik sailkatzen dituzten; errorea bakarrik zuzentzen duten ala esaldi osoa berridazten duten...

Ebaluaziorako eskatutako ariketan, etiketatzailer gehienek ahalik eta azpikategoria zehatzenetara jo dute erroreak sailkatzean, goiko kategorietan geratu beharrean. Oro har, ebaluatzaileei erraza egin zaie errore mota batzuk sailkatzea, baina beste batzuen aurrean zalantzak izan dituzte. Komunztadura-erroreak eta errore ortografikoak etiketatzeko, esaterako, ez dute arazorik izan. Errore semantikoak detektatu eta sailkatzea, berriz, zailagoa dela ikusi dugu: oso kasu gutxitan egin dute bat ebaluatzaile guztiek. Espero genuen moduan, errore asko dituzten esaldietan (hiru esalditan zeuden hiru erroretik gora) desadostasunak handiak izan dira hizkuntzalarien artean; izan ere, esaldiak luzeak, aldrebesak eta ulergaitzak diren kasuetan oso zaila gertatzen da errore bakoitzari kategoria zehatz bat esleitzea, eta horrelakoetan hoberena esaldia berridaztea izan ohi da.

Bestalde, hamaika adibidetan detektatu ditugu guk kategoria batean baino gehiagotan sailka daitezkeen erroreak. Ebaluatzaileei errore bera kategoria batean baino gehiagotan sartzeko aukera hobetsita dugula jakinarazi arren, etiketatzailerak artean adibide erroredun bakoitzari kode bakar bat esleitzea izan da joera orokorra. Esaldi bakar batean egin dute bat sei hizkuntzalarik errorea bi kategoriatan sailkatzen, eta pare bat

esalditan bakarrik sailkatu dituzte hiru hizkuntzalarik errore-adibide berberak kategoria batean baino gehiagotan. Hala ere, uste dugu sailkapena hobeto ezagututa eta erroreak sailkatzeko ohitura izanda, errazagoa dela errore mota bakoitzari kategoria bat edo gehiago dagokion aurreikustea.

Erroreak etiketatzeko eta zuzentzeko prozesuari dagokionez, berriz, bi aukera daude: erroreak sailkatu eta ondoren zuzentzea, edo errorea zuzendu eta ondoren kategoria-sistemaren arabera sailkatzea. Bost hizkuntzalarik adierazi digute errazagoa egin zaiela erroreak sailkatu aurretik zuzendu eta ondoren kategoria-sistemaren arabera mailakatzea. Sailkapena pixka bat ezagutzen zuten hiru hizkuntzalariek, aldiz, lehenik eta behin erroreak sailkatu dituzte eta ondoren zuzendu. Gainerako hiru etiketatzailleek kasu batzuetan sailkatu ondoren zuzendu dituzte erroreak eta, beste batzuetan, zuzendu ondoren sailkatu, inongo ordena finkorik jarraitu gabe.

Ebaluazioan parte hartu duten hamaika hizkuntzalariek sailkapenaren edo kategoria-sistemaren inguruan iritzi baikorra azaldu dute. Sailkapena ezagutzen ez zuten zazpi hizkuntzalarik batzuetan zehatzegia egin zaiela adierazi digute; eta zehaztasun-maila horrek agian etiketatzeko-prozesua zail dezakeela. Sailkapenaren eraikitze-prozesuan parte hartu duten beste lau hizkuntzalariek, ordea, sailkapen ahalik eta zehatzenaren aldeko iritzia eman dute, hain zuzen ere, corpus errealetan aurkitutako erroreak etiketatzeko aukera zabalagoa eskainiko digulako. Ikuspegi konputazionaletik ere uste dugu sailkapen zehatzek sendotasun handiagoa ziurtatzen dutela eta informazio-iturri aberatsagoa osatzen dutela. Eta hori da, oro har, errore-sailkapenekin lanean dihardutenen iritzia ere (Dulay *et al.*, 1982; Dagneaux *et al.*, 1998; James, 1998; Díaz-Negrillo eta Fernández-Domínguez, 2006). Azkenik, ezinbestekoa da sailkapenaren egituraren eta erabileraren inguruan irizpide batzuk finkatzea eta erabaki horiek erabiltzaileei jakinaraztea sailkapenaren baliagarritasuna bermatu nahi baldin bada.

V.5. Laburbilpena

Kapitulu honetan aurkeztu dugun euskarazko erroreen eta desbideratzeen sailkapena tesi-lan honen ekarpen nagusietako bat da. Izan ere, mota honetako sailkapenak ezinbestekoak dira corpusetan detektatutako erroreen edota desbideratzeen analisiak egiteko, sailkapeneko kategoriak eta azpikategoriak oinarri hartuta egiten baita erroreen eta desbideratzeen diagnosi linguistikoa.

Erroreen azterketa eta prozesamenduaren alorrean hain baliagarriak diren taxonomiak osatzeko guztiontzako baliagarria den eredurik ez dagoenez, ikertzaile bakoitzak bere helburuen arabeko sailkapenak definitu ohi ditu. Ondorioz, errore-taxonomia ugari eratu dira ikerketa-alor honetan.

Euskarazko erroreen eta desbideratzeen analisia bi xede nagusirekin egiten dugu: euskararako ortografia-, gramatika- eta estilo-zuzentzailea edota aldaki dialektalen markatzailea bezalako tresnak garatzeko eta euskararen i(ra)kaskuntza prozesua bideratzen laguntzeko tresnak prestatzeko. Hain zuzen, erroreak ortografia- eta gramatika-zuzentzaileari dagozkio eta desbideratzeak, berriz, estilo-zuzentzailea eta aldaki dialektalen markatzaileari. Helburuak desberdinak izan arren, sailkapen berbera baliatzen dugu ikerketa-alor bietan.

Kapitulu honetan aurkeztu dugun sailkapenaren abiapuntua IXA taldean kokatzen da. XUXEN zuzentzaile ortografikoa (Agirre *et al.*, 1992) garatu ostean, errore sintaktikoak lantzen hasteko sailkapen baten lehen hurbilpenak egin ziren (Gojenola, 2000). Lan horiek abiapuntutzat hartutaosatzen joan gara erroreen eta desbideratzeen sailkapen hau.

Sailkapena osatzen duten kategoriak eta azpikategoriak definitzean, kontuan hartu ditugu euskal gramatika-liburu batzuetan zehaztutako gramatika-kategoriak, beste hizkuntza batzuetan egin diren errore-taxonomiak, euskaltegietan irakasten diren gramatika-edukiak eta, batez ere, corpus errealetan etiketatutako adibideak. Horrela, sailkapen hierarkiko eta dinamiko bat eraiki dugu, irizpide linguistiko eta deskribatzailean oinarrituta, testuetan dauden adibideen diagnosi linguistikoak egiteko.

Sei kategoria nagusi bereizi ditugu sailkapenean: errore ortografikoak, lexiko erroreduna, gramatika-erroreak, puntuazio-ikurrei dagozkien erroreak, estilo kontu orokorrak eta errore semantiko-pragmatikoak. Kategoria bakoitzaren barruan, gramatika-kategoriaren arabera edota azaleko egituraren deskripzioaren arabera sailkatzen ditugu adibideak, azpikategorietan. Sailkapena dinamikoa da, irekia. Horri esker, kategoriak edo azpikategoriak kentzeko, gehitzeko edota aldatzeko aukera izango dugu, hein handi batean, testuetan etiketatutako adibideen arabera.

Bestalde, corpuseko adibide guztiak aldi berean detektatzea, etiketatzea eta aztertzea litzateke egokiena. Eskura ditugun bitartekoekin, ordea, hori ezinezkoa zaigu, eta ondorioz, errore errorez goaz euskarazko erroreen eta desbideratzeen sailkapena

zehazten. Tesi honetan determinatzaile-erroreari dagokien azpikategoria landu dugu (VI. kapitulua). Beste errore batzuen analisia ere egin da dagoeneko: komunztadura-erroreak, postposizio-erroreak, data egituretan egindako erroreak... (Oronoz, 2009). Eta etorkizuneko lanetan bideratu beharko da lantzeke dauden errorearen eta desbideratzearen azterketa eta prozesamendua.

Orain artean definitutako sailkapenaren ebaluazio bat egin dugu, kategoria-sistemaren egitura orokorraren inguruan ebaluatzaileek duten iritzia eta erroreak etiketatzean izan dituzten zailtasunak ezagutzeko. Kontuan hartu behar dugu kopuru txikiak erabili ditugula, sailkapena osatzen duten azpikategoria guztiak ez daudela sakon landuta eta ebaluatzaileek alor honetan duten esperientzia ere txikia dela kasu gehienetan. Baina ebaluazioa beharrezkoa iruditu zaigu sailkapenaren egitura nagusia ongi definituta dagoen ikusteko. Oro har, balorazio positiboa jaso du sailkapenak ebaluatzaileengandik.

Gure helburuetako bat euskarazko erroreak eta desbideratzeak sailkatzeko taxonomia zehatz eta erabilerraz bat eraikitzea izan da, sailkapena osatzen duten kategoriak edota azpikategoriak hartzen baititugu oinarri corpusetan detektatutako adibideen diagnosi linguistikoa egiteko; alegia, testuak etiketzeko. Horrela, euskararen azterketa eta prozesamenduaren alorrean ekarpenak egin ahal izateko informazio-iturri aberatsa izango den errorearen eta desbideratzearen bildumaosatuz goaz.

Erroreen edota desbideratzeen eskuzko etiketatzea eta azterketa linguistikoa funtsezko urratsak dira ondoren erroreen tratamendu automatikoa bideratu ahal izateko. Tesi-lan honetan determinatzaile-erroreak ditugu aztergai.

VI. Determinatzaile-erroreen eskuzko etiketatzea eta azterketa linguistikoa

VI.1. Sarrera

Erroreen eta desbideratzeen eskuzko etiketatzea eta azterketa linguistikoa izan ohi dira tratamendu automatikoaren abiapuntu. Erroreen eta desbideratzeen analisiaren alorrean, *detekzioa*, *diagnosi linguistikoa*, *diagnosi psikolinguistikoa*, *etiketatzea* eta *zuzenketa* bereizten ditugu. *Detekzioa* hitz, esaldi edota egitura jakin batean errorea edo desbideratzea aurkitzea da. Ondoren, erroreak edota desbideratzeak *etiketatu* egiten dira, hau da, informazio linguistikoa gehitzen zaie, informazio linguistikoaz hornitutako corpusak baliabide ahaltsuak baitira askotariko azterketak bideratu ahal izateko. Erroreen eta desbideratzeen sailkapena oinarri hartuta zehazten da detektatutako adibide bakoitza zein motatakoa den. Horri *diagnosi linguistikoa* egitea esaten diogu, eta baliagarria da bai erroreen tratamendu automatikoaren alorrean eta bai hizkuntzen i(r)a kaskuntzarenean ere. Erroreen eta desbideratzeen *diagnosi linguistikoaz* gain, hizkuntzen i(r)a kaskuntzaren alorrean *diagnosi psikolinguistikoa* ere egiten da: hizkuntza-ikasleari dagokion informazioa (ama-hizkuntza, xede-hizkuntzaren maila, ikasleak hizkuntza ikasten daraman denbora, etab.) kontuan hartuz aztertzen dira erroreak eta desbideratzeak eta jabekuntza edo ikasketa prozesuarekin lotzen dira. Analisiaren azken urratsa *zuzenketa* da, erroreaki dago(z)kion baliokide zuzena(k) ematea.

III. kapituluan aurkeztu dugun euskarazko erroreen eta desbideratzeen analisirako lan-ingurunea edozein errore motaren detekzioa, etiketatzea, diagnosia (linguistikoa edota psikolinguistikoa) eta zuzenketa bideratzeko prestatuta dago. Bertan, errore eta desbideratze mota guztien azterketa eta prozesamendua aldi berean bideratzea litzateke egokiena. Hori, ordea, ezinezkoa izan zaigu eskura ditugun baliabideekin: errore-etiketatzailerik bakar baten laguntza izan dugu corpusetan detektatutako adibideak etiketatu eta aztertzeko. Ondorioz, erroreen eta desbideratzeen analisia urratsez urrats egin beharra daukagu, errore-kategoriak edo motak: errore edota desbideratze mota bati dagozkion adibideak etiketatu, etiketatutako adibideak aztertu, detekzio automatikoa bideratu eta beste errore edo desbideratze baten analisiari ekiten diogu²³. Gure helburua euskarazko erroreen eta desbideratzeen azterketa oso bat egitea da. Errore ortografikoen detekzio automatikoa egin zen lehenik (Agirre *et al.*, 1992). Ondoren, hainbat gramatika-erroren azterketa eta prozesamendua bideratu da: postposizio-erroreak, data-egitura okerrak eta komunztadura-erroreak, besteak beste (Gojenola, 2000; Oronoz, 2009). Tesi-lan honetan, berriz, determinatzaile-erroreen inguruan egin dugun azterketa linguistikoa eta detekzio automatikoa aurkezten ditugu.

Determinatzaile-erroreen analisia egiteko, kontuan izan ditugu euskarazko gramatika-erroren tratamendu automatikoa bideratzeko IXA taldean egin diren gogoetak, hartu diren erabakiak eta finkatu diren irizpideak. Gramatika-liburuetako oharra eta adibideak ere kontsultatu ditugu. Baina, batez ere, euskara-ikasleen corpora hartu dugu errore-analisiaren abiapuntutzat; izan ere, garatzen ditugun tresnak baliagarriak izango badira, funtsezkoa da erabiltzaileek egiten dituzten erroreak aintzat hartzea.

²³ Beste aukera bat izan zitekeen corpusaren lagin txikietan detektatzen diren errore edota desbideratze mota guztiak etiketatu eta aztertea. Aukera horrek badu, baina, eragozpen bat: errore eta desbideratze mota bakoitzaren agerpen kopurua edo maiztasuna ez da nahikoa ondorioak atera, oinarriak finkatu edota tratamendu automatikoa bideratu ahal izateko.

Determinatzaile-erroreak sintagma-mailako erroreak dira, errore sintaktiko lokalen multzokoak, eta ez esaldi- edota testu-mailako errore globalak. Elkarren ondoan dauden bi edo hiru hitzek osatu ohi dute, beraz, errore-eremua. IXA taldean Hizkuntzaren Azterketa eta Prozesamenduaren (HAP) alorrean garatu diren tresnekin errore sintaktiko lokalen detekzio automatikoa bideragarria izan daitekeela aurreikusita ekin diogu errore mota honen azterketa eta tratamendu automatikoari: Murritzapen Gramatika (MG) formalismoan (Karlsson *et al.*, 1995) oinarritutako erregelak idatzi ditugu determinatzaile-errore batzuen detekzio automatikoa bideratu ahal izateko (VII. kapitulua).

Erroreen eta desbideratzeen detekzio automatikoaren abiapuntua eskuzko etiketatzea eta azterketa linguistikoa izan ohi dira, eta horretaz hitz egingo dugu, hain zuzen ere, kapitulu honetan. VI.2. atalean, euskal determinatzaileen deskripzio linguistiko orokorra egingo dugu. Atal horretan hitz egingo dugu euskal determinatzaileen inguruan egin diren sailkapenez (VI.2.1.) eta euskal determinatzaileen eta zenbatzaileen jokaera sintaktikoaz (VI.2.2.). Ondoren, VI.3. atalean, determinatzaile-erroreak etiketatzeko finkatu ditugun irizpideak azalduko ditugu, eta determinatzaile-erroreen diagnosi linguistikorako oinarri den sailkapena aurkeztuko dugu VI.4. atalean. Sailkapen hori euskarazko erroreen eta desbideratzeen sailkapen orokorrari dagokion azpikategorietako bat da. VI.5. atalean, determinatzaile-erroreen eskuzko etiketazetik eskuratutako datuak aurkeztuko ditugu. Kapitulua laburbilpen batekin eta ondorio batzuekin itxiko dugu (VI.6. atala).

VI.2. Euskal determinatzaileen deskripzio linguistikoa

VI.2.1. Euskal determinatzaileen sailkapena

Hizkuntzalari tradizionalak (Azkue, 1925; Lafitte, 1944; Villasante, 1972 eta 1983; Txillardegi, 1978) *artículos* / *articles* / *artikuluak* eta *demonstrativos* / *demonstratifs* / *erakusleak* deritzen hizkuntza-elementuei erreferentzia egiteko *determinante* terminoa erabiltzen duen lehen euskal hizkuntzalari Goenaga (1980) da. Determinante kategoriaren barruan mugatzaileak eta zenbatzaileak bereizten ditu. Mugatzaileen artean sartzen ditu artikulua eta erakusleak; zenbatzaileen artean, hiru mota bereizten ditu: zehaztuak (kopuru jakin bat adierazten dutenak), zehaztugabeak

(zehaztu gabeko kopuru bat adierazten dutenak) eta orokorrak (7. taula). Horiez gain, *beste* ‘nolabaiteko’ determinantetzat hartzen du Goenagak eta izenaren ezkerrean kokatzen diren elementuei (zenbatzaileei eta baita *edozein*, *zenbat* eta antzeko elementuei) predetermine esaten die. Euskal ISaren egitura honela azaltzen du:

IS → (Predeterminantea) + I + (mug)

DETERMINANTEAK		
MUGATZAILEAK	artikuluak	-a, -ak, -ok
	erakusleak	<i>hau, hori, hura, hauek, horiek, haiek...</i>
ZENBATZAILEAK	zehaztuak	numeralak: <i>bat, bi, hiru, lau, bost...</i> ordinalak: <i>bigarren, laugarren...</i> distributiboak: <i>hiruna, seina...</i>
	zehaztugabeak	<i>zenbait, asko, gutxi, anitz, honenbeste, horrenbeste, hainbeste...</i>
	orokorrak	<i>guzti, dena, oro</i>
BESTE DETERMINANTE BATZUK: <i>beste</i>		

7. taula. Determinanteen sailkapena Goenagaren (1980) arabera.

Kategoria eta azpikategoria bakoitzean sailkatu behar diren elementuei dagokienez, Goenagak (1980) arazotzat hartzen du *bat* elementuaren sailkapena: zenbatzailea ala mugatzailea da? Bi arrazoi ikusten ditu *bat* mugatzailetzat hartzeko: batetik, beste mugatzaileen lekuan ager daiteke (*{gizona / gizon **hau** / gizon **bat**} etorri da*); bestetik, bere kokagunea beste zenbatzaile zehaztuena ez bezalakoa da (*gizon **bat** / **bi** gizon edo gizon **bi** / **hiru** gizon / **lau** gizon...*). Jokaera horiek kontuan hartuz gero, *bat* zenbatzailea eta *bat* artikulua zehaztugabea bereizi beharko liratekeela dio Goenagak. Baina planteamendu horrek hiztegian bi *bat* desberdin egongo liratekeela onartzera eramango gintuzke eta, beraz, ekonomikoagoa deritzo *bat* zenbatzailetzat hartzeari.

Euskaltzaindiak (1985) Goenagaren sailkapenari jarraitzen dio, baina *determinatzaile* izendapena erabiltzen du, eta izendapen hori erabiltzen dute hortik aurrera euskaraz idatzitako gramatika-lanek, oro har. Euskaltzaindiak (1993) hirugarren talde bat gaineratzen die Goenagaren lanean bereizitako biei (mugatzaileei eta zenbatzaileei). Hirugarren talde horri *determinatzaile zehaztugabeak* esaten dio eta ondoko elementuak sartzen ditu azken talde horretan: *zein, edozein, zeinahi, zeingura, zer, edozer, zernahi, zergura*. Bestalde, Euskaltzaindiak (1993) ere *bat* eta *batzuk* elementuak zenbatzailetzat hartzen ditu, ez artikulutzat (8. taula).

DETERMINATZAILEAK		
MUGATZAILEAK	artikuluak	-a, -ak, -ok
	erakusleak	<i>hau, hori, hura, hauek, horiek, haiek...</i>
ZENBATZAILEAK	zehaztuak	numeralak: <i>bat, bi, hiru, lau, bost...</i> ordinalak: <i>bigarren, laugarren...</i> distributiboak: <i>hiruna, seina...</i>
	zehaztugabeak	<i>batzuk, zenbait, asko, gutxi, anitz, honenbeste...</i>
	orokorrak	<i>guztia, dena, oro</i>
DETERMINATZAILE ZEHAZTUGABEAK	zein, edozein, zeinahi, zeingura, zer, edozer, zernahi, zergura beste	

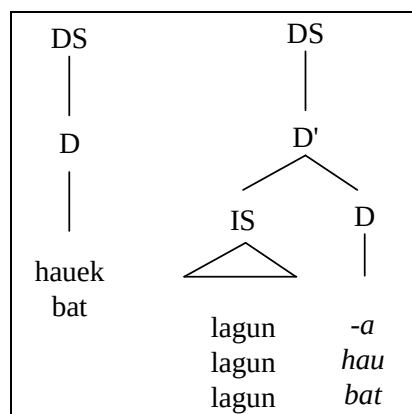
8. taula. Determinatzaileen sailkapena Euskaltzaindiaren (1993) arabera.

Zubiri eta Zubiri (1995) gramatika didaktikoak Euskaltzaindiaren (1993) sailkapenari jarraitzen dio, baina *bat* eta *batzuk* elementuak artikulu zehaztugabetzat hartzen ditu; gainera, zenbatzaile zehaztuen artean ere sartzen du *bat* (9. taula).

DETERMINATZAILEAK		
MUGATZAILEAK	artikulu zehaztuak	-a, -ak, -ok
	artikulu zehaztugabeak	<i>bat, batzuk</i>
	erakusle arruntak	<i>hau, hauek, hori, horiek, hura, haiek</i>
	erakusle indartuak	<i>hauxe, hauxek, horixe, horiexek, huraxe, haixek, berau, berauek, berori, beroiek, bera, beraiek</i>
ZENBATZAILEAK	zehaztuak	kardinalak: <i>huts, zero, bat, bi, hiru, hamar, hoge...</i> ordinalak: <i>lehen, bigarren, hirugarren...</i> zatikiak: <i>erdia, herena, laurdena...</i> ehunekoak: <i>ehuneko bost...</i> banatzaileak: <i>bana, bosna, bakoitza/bakoitzeko</i> zenbait esapide: <i>bizpahiru, hoge bat, hogeiren bat, baten bat, ehun inguru, milatik gora, hamarretik, hogeira bitartean, milaka...</i>
	zehaztugabeak	<i>zenbat, batzuk, zenbait, asko, gutxi, anitz, franko, hagitx, hainbeste, hainbat, ugari, dezente, nahiko(a), aski, pilo bat, makina bat, apur bat, pixka bat...</i>
	orokorrak	<i>dena, guztia, oro, oso, bete</i>
DETERMINATZAILE ZEHAZTUGABEAK	edozein, zein, zeinahi, zer, zernahi, edozer	

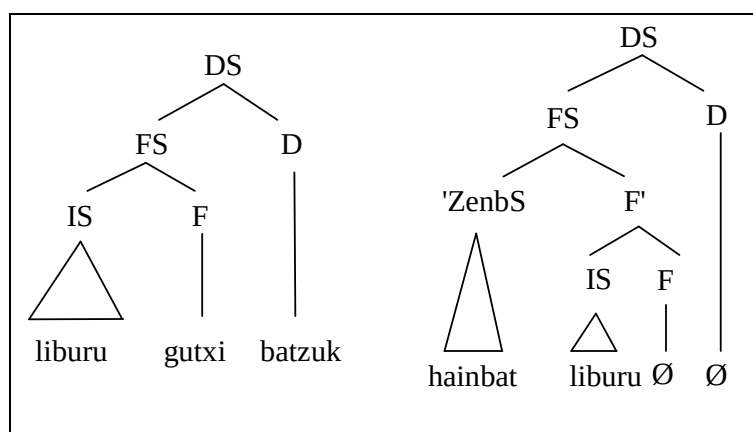
9. taula. Euskarazko determinatzaileen sailkapena (Zubiri eta Zubiri, 1995).

Abney-ren (1987) Determinatzaile Sintagmaren Hipotesiaren (DSH) arabera euskal determinatzaile-sintagmaren egituraren inguruan egindako azterketek (Artiagoitia, 1998 eta 2004, besteak beste) agerian uzten dute determinatzaileak eta zenbatzaileak bereizi beharreko kategoriak direla, gunen desberdinak betetzen baitituzte determinatzaile-sintagmaren barruan. Determinatzaile-sintagmaren burua betetzen dute *-a, hau* eta *bat* bezalako elementuek (18. irudia):



18. irudia. Determinatzaile-sintagmaren egitura, (Artiagoitia, 1998) lanetik hartua eta moldatua.

Zenbatzaileek aldiz, F sintagmaren burua edo espezifikatzailea beteko lukete²⁴ (Artiagoitia, 2004). Artiagoitiaren analisia abiapuntu ona da azaltzeko zergatik *-a*, *hau* eta *bat* bezalakoek beti agertzen diren ISaren eskuinaldean, eta zergatik ezin ditugun horrelako bi determinatzaile batera aurkitu (**liburua hau*) eta bai, ordea, determinatzaile horietakoren bat zenbait zenbatzailerekin (*hiru trenak*, *hiru lagun horiek*, *liburu gutxi batzuk*, *zuk duzun diru gutxia*); izan ere, zenbait zenbatzailek determinatzailearen buruak Ø balioa izatea eskatzen dute (*hainbat liburu* / **hainbat liburuak*) (19. irudia).



19. irudia. Determinatzaileen eta zenbatzaileen kokagune sintaktikoa, Artiagoitia (2004) lanetik hartua eta moldatua.

²⁴ F burua numeroarekin lotutako buru funtzionala da.

Artiagoitiak (1998), beraz, determinatzaileen kategorian sartzen ditu mugatzaileak (artikuluak eta erakusleak), partitiboa, *batzuk* eta zenbait determinatzaile zehaztugabe (*zein*, *ze(r)* eta *edozein*). Azken elementu horiek determinatzailetzat hartzeko arrazoitzen hartzen du zenbaitetan zenbatzaileekin batera agertu ahal izatea (***zein hiru lagun ikusi dituzu mendian?***). Baina zalantza batzuk ere azaltzen ditu, izenaren ezker aldean kokatzen direlako (***edozein liburu***) eta zenbaitetan artikulua zehaztuarekin batera ager daitezkeelako (***zein idazle ona den!***)²⁵.

Bestalde, determinatzaile-sintagmaren kategoriatzat hartzen dira izen propioak eta izenordainak (Abney, 1987). Izen propioek Ø balioa eskatzen diote D buruari (***Mikel etorri da***), nahiz eta hizkuntza erromantze batzuetan artikulua ageria azal dezaketen (*la Ana, el Pedro...*). Izenordainek, determinatzaileek bezala, singularra eta plurala bereizten dute, euskaraz determinatzailearen ezaugarria dena²⁶. Zenbatzaileek aldiz, ez dute numero-bereizketarik (Artiagoitia, 1998): *lagun asko etorri {da / dira} jaialdira*.

Lakak (1996) ere *determiners* eta *quantifiers* bereizten ditu. Ondoko eskeman ikus daitezke proposatzen dituen kategoriak (10. taula):

DETERMINERS	Determiners: -a / -ak	
	The proximity determiner: -ok	
	Demonstratives: <i>hau, hori, hura, hauxe, horixe...</i>	
QUANTIFIERS	Numerals	Numerals: <i>bat, bi, hiru, lau...</i>
		Ordinals: <i>lehen, bigarren, hirugarren...</i>
		Distributives: <i>bana, bina, hiruna...</i>
	Quantifiers that require determiners: <i>guzti, bakoitz, gehien</i>	
	Determinerless quantifiers: <i>zenbait, asko, hainbat, gutxi, edozein, oro, zenbat, zein...</i>	

10. taula. Euskal determinatzaileak eta zenbatzaileak Lakaren (1996) arabera.

Trask-ek (2003) Euskaltzaindiaren (1993) sailkapenaren antzekoa erabiltzen du, baina elementu gehiago sartzen ditu *determiners* deritzon kategorian, eta artikulua

²⁵ Nolanahi ere, harridurazko perpausen kasua berezia dela dirudi. Izan ere, *zein* adjektiboaren graduatzaile modura interpretatzen dugu determinatzaile bezala baino areago.

²⁶ Euskarazko izenek ez dute numerorik. Hori dela eta, determinatzaile baten beharra izaten dute numeroa adierazteko (Artiagoitia, 2004).

zehaztugabe eta zenbatzaile gisa sailkatzen du *bat*. Hona hemen Trask-en sailkapenaren eskema bat (11. taula):

DETERMINERS	
ARTICLES	The definite articles: <i>-a</i> , <i>-ak</i>
	The proximate plural article: <i>-ok</i>
	The indefinite article: <i>bat</i>
DEMONSTRATIVES	Ordinary demonstratives: <i>hau</i> , <i>hori</i> , <i>hura</i> ...
	Intensive demonstratives: <i>hauxe</i> , <i>horixe</i> , <i>huraxe</i>
THE PARTITIVE	<i>-(r)ik</i>
NUMERALS	Cardinal numerals: <i>huts</i> , <i>bat</i> , <i>bi</i> , <i>hiru</i> ...
	Distributive numerals: <i>bana</i> , <i>bina</i> , <i>hiruna</i> ...
	Ordinal numerals: <i>lehen</i> , <i>bigarren</i> , <i>hirugarren</i> ...
	Fractions and percentages: <i>erdi</i> , <i>heren</i> , <i>laurden</i> ...
	Measure noun phrases: <i>lau libera</i> , <i>bi kilo</i> , <i>botila bat</i> ...
QUANTIFIERS	<i>zenbat</i> , <i>hainbat</i> , <i>hainbeste</i> , <i>batzuk</i> , <i>gutxi</i> , <i>gehiago</i> , <i>anitz</i> , <i>zenbait</i> , <i>asko</i> , <i>oro</i> , <i>franko</i> ...
INDEFINITE AND INTERROGATIVE DETERMINERS	<i>zein</i> , <i>zer</i> , <i>zenbat</i> , <i>edozein</i> ...

11. taula. Euskal determinatzaileak eta zenbatzaileak Trask-en (2003) arabera.

Trask-ek *-(r)ik* partitibo atzizkia ere determinatzailetzat hartzen du, eta polaritate-elementua (*polarity item*) dela zehazten du, alegia, polaritate-testuinguruetan (ezeztapena, baieztapena, baldintza...) bakarrik ager daitekeen determinatzailea dela. Lakak (1993) ere *polar determiner* denominazioa erabiltzen du partitibo atzizkiari erreferentzia egiteko. Nolanahi ere, determinatzaileen jokaera ikuspegi teorikotik aztertu duten hizkuntzalariek ez ezik (Laka, 1993; Artiagoitia, 1998; Trask, 2003) Lafitte-k (1944) berak ere *l'article partitif* esaten dio *-(r)ik* atzizkiari bere gramatika ospetsuan.

Tesi-lan honetan, azpikategorien izendapenetarako Euskaltzaindiaren (1993) sailkapenari eutsiko diogu neurri handian, baina determinatzaileak eta zenbatzaileak bereiziko ditugu, bereizketa horrek erroreen diagnosi linguistikoan lagunduko baitigu. Hala, determinatzaileen barruan sartzen ditugu artikulua zehaztuak eta zehaztugabeak, erakusleak, partitiboa, determinatzaile zehaztugabeak eta galderazko determinatzaileak; zenbatzaileen artean, berriz, zenbatzaile zehaztuak, zehaztugabeak eta orokorrak bereizten ditugu. Bestalde, *beste* elementurako predeterminatzaile izendapena erabiliko dugu, zein kategoria egotzi behar zaion eztabaidatu gabe; izan ere, *beste* determinatzaile eta zenbatzaile guztien ezkerretara kokatzen da eta

elementu horiekin guztiekin batera ager daiteke: *beste liburua*, *beste liburu bat*, *beste edozein liburu*, *beste hiru liburu*, *beste liburu asko*...

Bestalde, eta aurrerago ikusiko dugun bezala, determinatzaile-erroreetan oso maiz inplikatur dago ‘artikulu zehaztu’ esaten zaion *-a* atzizkia. Nolanahi ere, euskaraz, hizkuntza erromantzeetan ez bezala, ‘artikulu zehaztu’ esaten zaion kategoriak ez dakar beti irakurketa zehaztua edo espezifikoa (Laka, 1993). Atzizki horrekin lotutako erroreak gehienetan ez daude zehaztasun edo mugatasun ezaugarriekin lotuta; aitzitik, euskararen morfologia flexiboaren arauekin zerikusia duten erroreak izan ohi dira. Hori dela eta, determinatzaile-erroreez hitz egingo dugunean, *determinatzaile atzizki* izendapena erabiliko dugu.

Hona hemen, beraz, gure sailkapena (12. taula):

DETERMINATZAILEAK	artikulu zehaztuak	<i>-a, -ak, -ok</i> (determinatzaile atzizkiak)
	artikulu zehaztugabeak	<i>bat, batzuk</i>
	erakusleak	<i>hau, hori, hura, hauek, horiek, haiek...</i>
	artikulu partitiboa	<i>-(r)ik</i>
	zehaztugabeak	<i>edozein, edozer, zeinahi, zernahi</i>
	galderazkoak	<i>zer, zein, zertzuk, zeintzuk</i>
ZENBATZAILEAK	zehaztuak	kardinalak: <i>huts, zero, bat, bi, hoge</i> ... ordinalak: <i>lehen, bigarren, hirugarren...</i> zatikiak: <i>erdia, herena, laurdena...</i> ehunekoak: <i>ehuneko bat, ehuneko bi...</i> banatzaileak: <i>bana, bina, launa...</i> neurri-sintagmak: <i>kilo bat, bi litro...</i> bestelako esapideak: <i>bizpahiru, hoge bat, hogeiren bat, baten bat...</i>
	zehaztugabeak	<i>zenbat, zenbait, hainbat, asko, ugari, gutxi, hainbeste, dezente, nahiko(a), aski, pilo bat, makina bat, apur bat, pixka bat, ...</i>
	orokorrak	<i>den, guzti, oro, oso, bete</i>
PREDETERMINATZAILEA	<i>beste</i>	

12. taula. Oinarritzat hartu dugun determinatzaileen sailkapena.

VI.2.2. Euskal determinatzaileen eta zenbatzaileen jokaera sintaktikoa

Atal honetan bilduko ditugu euskal determinatzaileen eta zenbatzaileen jokaera sintaktikoaz aritu diren lanetan (Goenaga, 1980; Euskaltzaindia 1985 eta 1993; Zubiri eta Zubiri, 1995; Laka, 1996; Artiagoitia, 1998 eta 2004; Trask, 2003, besteak beste) deskribatu diren sintaxi-arauak. Izan ere, determinatzaile errore bat detektatu

eta diagnosi linguistikoa egiten dugun bakoitzean, arau horietakoren bat hautsi dela finkatzen ari gara:

- a) Artikuluak (zehaztuak, zehaztugabeak eta partitiboa) eta erakusleak beti azaltzen dira determinatzaile-sintagmaren eskuinaldeko muturrean:

*liburua, liburu **bat**, liburur**ik**, liburu **hori**,
liburu zuria, liburu zuri **bat**, liburu zuri lodir**ik**,
liburu zuri askor**ik**, liburu zuri gutxi batz**uk***

- b) Determinatzaile zehaztugabeak eta galderazko determinatzaileak beti azaltzen dira determinatzaile-sintagmaren ezker aldeko muturrean:

***edozein** liburu, **edozein** liburu zuri, **edozein** bi liburu zuri,
zein liburu?, **zein** liburu zuri lodi?*

- c) Zenbatzaile batzuk izenaren eskuinaldean kokatzen dira, beste batzuk izenaren ezker aldean eta beste batzuek eskuinaldeko zein ezker aldeko kokagunea onartzen dute:

*Liburu **bat** / ***Bat** liburu
Liburu **bi** / **Bi** liburu
Hiru liburu / ***Liburu hiru**
Liburu **gutxi** / ***Gutxi** liburu
Zenbait liburu / Liburu **zenbait***

- d) Determinatzaile-sintagma batean determinatzaile bakarra azal daiteke:

****Liburua** zuria
Liburua** zuri batz**uk

- e) Argumentu-guneak (subjektua, osagarri zuzena eta zeharkako osagarria) betetzen dituzten DSei beti azaltzen dute determinatzaile ageri bat. Euskal gramatiketan 'artikulu zehaztu' (-a/-ak) esan ohi zaienek ez dute beti ematen irakurketa zehaztu edo espezifikorik (Laka, 1993). Artikulu partitiboak, determinatzaile zehaztugabeak eta galderazko determinatzaileak, aldiz, beti daramate irakurketa inespezifikoa:

Liburua ekarri dut [+ espezifikoa]
Garagardoa edango dugu [+/-espezifikoa]
**Garardo* edango dugu
Ez dago garagardorik [-espezifikoa]
**Ez dago garardo*
Edozein garardo edango dugu [-espezifikoa]
Zer garardo nahi duzu? [-espezifikoa]

- f) Zenbatzaile zehaztugabeek eta zenbait zenbatzaile zehaztuk (kardinalak, banatzaileek, ehunekoek, neurri-sintagmek eta zenbait esapidek) Ø determinatzailearen agerpena zilegiztatzen dute:

Hainbat liburu irakurri ditut
Garardo asko edan dugu
Bi garardo edan ditugu
Hiruna liburu oparitu dizkie
Bi botila ardo edan dute
Bizpahiru liburu erosiko ditut

- g) Zenbatzaile zehaztu kardinalak determinatzaile ageriak ere onartzen dituzte, irakurketa espezifikoa duten determinatzaile-sintagmetan:

Zuk oparitutako lau liburuak irakurri ditut
Mahaian dauden bi garardo horiek edango ditut

- h) Zenbatzaile ordinalak, zatikiek eta zenbait zenbatzaile orokorrek determinatzaile ageri bat eskatzen dute:

Lehen etxea da nirea / **Lehen etxe* da nirea
Tarta laurden bat jan du / **Tarta laurden* jan dut
Tarta osoa jan du / **Tarta oso* jan dut
Zikinkeria guztia utzi duzu / **Zikinkeria guzti* utzi duzu

- i) *Beste* predeterminatzaileak beti eskatzen du determinatzaile ageri bat, Ø determinatzailea zilegiztatzen duen zenbatzaile batekin agertzen ez bada:

*Ekarri **beste** liburu
 Ekarri **beste** liburua
 Ekarri **beste** liburu **hori**
 Ez dago **beste** libururik
 Ekarri **beste edozein** liburu
 Ekarri **beste zenbait** liburu

- j) Euskara hizkuntza aglutinatzailea delarik, determinatzailearen eskuinaldean kasu-atzizkiak (ergatiboa edo datiboa) edota postposizioak lotu ohi dira. Lotura horietan aldaketa morfofonologikoak gertatzen dira eta hiztun arruntentzat zaila izan ohi da ‘deklinabide-atzizki’ esan ohi zaienak biltzen dituzten morfemak bereiztea. Artikulu zehaztu singularrarekin (*liburua*) eta pluralarekin (*liburuak*) deklinabide-atzizki mugatu singularrak (*liburuan*, *liburutik*, *liburuko*...) eta pluralak (*liburuetan*, *liburuetatik*, *liburuetako*...) esaten zaienak eratzen dira. Ø determinatzailearekin aldiz (*hainbat liburu*), deklinabide-atzizki mugagabeak esaten zaienak eratzen dira (*hainbat liburutan*, *hainbat liburutatik*, *hainbat liburutako*...).
- k) Badira determinatzailearik eta zenbatzailearik azaltzen ez duten aditz-esapideak: *lan egin*, *lo egin*, *min eman*... Horrelako esapide lexikalizatuetan azaltzen diren elementuak ez dira DS kategoriagoak, IS kategoriakoak baizik (Laka, 1993). Beste aditz-esapide batzuek, aldiz, determinatzailea azaltzen dute, baina lexikalizatutzat daudenez gero, determinatzaile edota numero jakin bat baino ezin dute azaldu: *larrua jo* / **larru jo*; *harrikoa egin* / **harriko hori egin* / **harrikoak egin* (Zabala, 2004). Horrelako esapideak HAUL modura aztertu eta hiztegi-mailan tratatu behar dira.
- l) Lexikalizatutako esaeretan ere azaldu ohi dira determinatzailearik gabeko ISak: *Alargun*, *begi ilun*. *Atzerri otserri*.
- m) Azkenik, izenki-predikatuek zenbaitetan *-a/-ak* edota *-rik* atzizkiak azaldu ohi dituzte. Artiagoitiak (1997) horrelakoetan azaltzen diren atzizkiak determinatzailearen kategoriakoak direla defendatu du, baina izenki-predikatuetan agertzen diren elementu horien kategoria oraindik

ere eztabaidagai dela esan dezakegu. Adibidez, Zabalak (1993) defendatu du izenki-predikatuetan agertzen den atzizkia numero-atzizkia dela. Eguren (2006), aldiz, kopula pronominala dela aldarrikatu du. Bestalde, izenki-predikatuetan *-a/-ak* atzizkiak agertzea gobernatzen duten arauak argumentuak diren sintagmenak ez bezalakoak dira. Hona, hemen, besteak beste Zabalaren (1993, 2004) lanetan deskribatzen diren zenbait arau:

- Bigarren mailako predikatuek edo predikatu adjunktuek²⁷ ez dute inoiz hartzen atzizkirik: *Artzain joan da Ameriketara*.
- Izenki-predikatuak osagarri modura hartzen dituzten lotura-aditzek (kopulek) erabakitzen dute izenki-predikatuek zein atzizki har ditzaketen²⁸:

*Mikel azkarra da / *Mikel azkar da*
*Ura ardo bihurtu zuen / *Ura ardo**a** bihurtu zuen*
*Zurbil jarri da / ?Zurbil**a** jarri da*
*Biluz**ik** gelditu da*
Lodi dago
*Poz**ik** dago*
*Interesgarrit**zat** hartu dute*

- Izenki-predikatuaren zehaztapen lexikoek erabakitzen dute zer nolako atzizkiak onartzen dituen predikatuak. Oro har, *-a(k)* atzizkia ezaugarri iraunkorrek adierazten dituzten predikatuetan agertu ohi da (izaki mailako predikatuetan) eta $\emptyset$ eta *-rik* atzizkiak, aldiz, behin-behineko ezaugarriak adierazten dituzten predikatuetan (egoera mailako predikatuetan):

²⁷ Bigarren mailako predikatuak ez dira ezinbestekoak perpausa gramatikala izateko, horrexegatik analizatzen dira adjuntu modura: *Ameriketara artzain joan da. / Ameriketara joan da. / Joan da.*

²⁸ Izenki-predikatu osagarriak ezinbestekoak dira perpausa gramatikala izateko: **bihurtu zuen, *dago*.

*Jone haurdun dago / *Jone haurduna dago*
*Jone azkarra da / *Jone azkar da*
Jone lodi dago / Jone lodia dago
Jone lodia da / Jone lodi da
*Jonek zintzo jokatu du / *Jonek zintzoa jokatu du*
*Jone zintzoa da / *Jone zintzo da*

Tesi-lan honetan aztertu ditugun corpusetan aurkitutako determinatzaile-errore mota guztien eskuzko etiketatzea egin da; tratamendu automatikoari dagokionez, berriz, lehen urrats honetan argumentu guneeetan azaltzen diren determinatzaile-sintagmetan egindako errore batzuen tratamendu automatikoa bideratu dugu (VII. kapitulua).

VI.3. Determinatzaile-erroreak etiketatzeko irizpideak

Erroreen eskuzko etiketatzea ezinbesteko urratsa da adibide erroredunen bildumak osatzeko, erroreen tratamendu automatikoa bideratzeko eta eskuzko lana eta detekzio automatikoa alderatu ahal izateko. Erroreak etiketatzeko irizpideak ongi definitzea, beraz, garrantzitsua da. Kalitatezko etiketatzea egin nahi baldin bada, finkatutako irizpide horiek eskuliburu batean jasotzea komeni da, zalantzen aurrean errore-etiketatzaleek kontsultatu ahalko duten eskuliburu batean.

Asko dira erroreen eta desbideratzeen analisiaren inguruan egin eta argitaratu diren lanak (Fernández, 1997; de Mönnink, 2000; Granger, 2003; Díaz-Negrillo eta Fernández-Domínguez, 2006). Lan horietan guztietan erroreen analisirako corpusez, erroreen sailkapenez, errore-editoreez eta erroreen etiketatze-prozesuez hitz egiten da, oro har. Baina ez dugu aurkitu erroreak etiketatzeko errore-eremuak nola definitu (erroretzat zein eremu hartu) behar diren argi zehazten duen artikulurik. Joera nagusia errorea duen hitza bakarrik etiketatzea da, eta errorea bi hitzetan baldin badago, bi hitzak markatzea. Ez dago, ordea, erroreak etiketatzeko eredu estandarrik, eta ondorioz, ikerketa bakoitzaren helburuen arabera irizpideak finkatu ohi dira.

Euskarazko erroreen eta desbideratzeen azterketa eta prozesamenduaren alorrean, irizpide batzuk zehaztu ditugu gure helburuak eta gure hizkuntzaren ezaugarriak aintzat hartuta. Irizpide nagusietako bat errore-eremuari dagokio,

erroretzat zein eremu hartuko den finkatzeari. Errore mota bakoitzak bere ezaugarriak dituenez, ezin dira denak modu berean tratatu. Horren arabera, erroretzat hartuko den eremua aldatuz doa. Errore ortografikoen kasuan, esaterako, ez dago zalantzarik: hitz-mailako erroreak dira eta hitz-mailan tratatzen dira (**leihaketa irabazi du*, **tsakurrak gustatzen zaizkit...*). Ortografikoak ez diren erroreen eremuak, aldiz, hitz bakarreko sintagma-mailakoak (**gizon ikusi dut*), hitz bat baino gehiagoko sintagma-mailakoak (**gizona hori da zure aita?*) edota perpaus-mailakoak (**etorri diren gizona ikusi dut*, **kotxea dutenak ekar dezatela*) izan daitezke. Oronozen (2009) tesi-lanean, adibidez, hiru errore-kategorien tratamendu automatikoa landu da: komunztadura, datak eta postposizio-lokuzioak. Lehenengoak esaldi-mailan etiketatzen dira, hau da, errore-eremutzat esaldi osoa hartzen da (**zentral nuklearrak zakar erradiaktiboa eratzen dute*); data-egituretan egiten diren erroreen kasuan, data-egitura osoa markatzen da (**2007ko abuztuaren 3*); eta postposizio-lokuzioetan egindako erroreetan postposizio-sintagma osatzen duten elementuak etiketatzen dira (**basoaren zehar joango gara amarekin*).

Determinatzaile-erroreen eremuak mugatzeko, hainbat aukera aurreikusi genituen:

- Batetik, errorea duen hitza bakarrik, bere horretan, marka genezake. Errorea non dagoen edo non ikusten den adieraziko genuke honela: **etorri den* <DETK>*gizon*<DETK> *zure aita da?*, **auto* <DETK>*berri*<DETK> *erostera goaz*, **berehala* <DETE>*araua*<DETE> *batzuk idatzi zituzten...* Aukera hau, baina, ez zaigu zuzena iruditzen; batetik, determinatzaile-errorea sintagmako hainbat osagairi dagokiolako eta ez markatutako hitzari bakarrik; bestetik, etiketatutako hitzak bakarrik ez digulako nahikoa informazio ematen zein errore mota den esateko.
- Beste aukera bat litzateke determinatzaile-erroreak, sintagma-mailakoak direnez, sintagma-mailan etiketatzea, sintagma osoak markatuz: **<DETK>etorri den gizon*<DETK> *zure aita da?*, **<DETK>lapurtutako material*<DETK> *ez du ekarri*, **berehala* <DETE>*araua batzuk*<DETE> *idatzi zituzten...* Sintagma osoa markatzeak, ordea, askotan ez du zentzurik, izenaren (buruaren) ezkerretara dauden izenlagun edota menpeko perpaus modifikatzaileek ez baitute eraginik determinatzaile-erroreen gainean.

- Determinatzaile-erroreen eremuan aditza ere sar genezake, hau da, determinatzaileari dagokion elementua eta aditza marka genitzake: **bere* <DETK>**lanbide praktikatu**<DETK> *ahal du*, **brehala* <DETE>**araua batzuk idatzi**<DETE> *zituzten*, **<DETK>auto berri erostera*<DETK> *goaz...* Baina determinatzaile-erroreen eremuan aditza ere sartzea ez zaigu irizpide zuzena iruditzen, aditza ez baitagokio, kasu gehienetan, determinatzaile-erroreari (izenki-predikatuen eta aditz-esapideen kasuan salbu).
- Determinatzaile-errorea dagoela esateko behar den hitz kopuru minimoa etiketatzea da beste aukera bat, hau da, sintagmako burua eta eskuineko elementu guztiak markatzea, segida horrek sintagma bat osatu behar duela jakinda: **etorri den* <DETK>**gizon**<DETK> *zure aita da?*, **<DETK>auto berri*<DETK> *erostera goaz*, **brehala* <DETE>**araua batzuk**<DETE> *idatzi zituzten*, **nire*<DETE>**laguna handia**<DETE> *etorriko da...* Eta guk irizpide horren alde egin dugu. Bada:
 - o ISaren buruak ezkerretara izan ditzakeen izenlagunak edota modifikatzaileak ez ditugu errore-eremuaren barruan sartzen: **lapurtutako* <DETK>**material**<DETK> *ez du ekarri*, **etorri den* <DETK>**gizon**<DETK> *zure aita da?*, **bere* <DETE>**laguna handia**<DETE> *etorriko da...*
 - o ISaren buruaren eskuinaldean dauden hitzak edota hitz-segidak (adjektiboak, erakusleak, zenbatzaile mugagabeak...) errore-eremuan sartzen ditugu, erroretzat burua eta buruaren eskuinaldeko sekuentzia osoa hartuz: **<DETK>auto berri*<DETK> *erostera goaz*, **<DETE>haurra hauek*<DETE> *ailegatu direnean*, **<DETE>etxea handiago bat*<DETE> *erosi dute...*; izan ere, izenaren eskuinaldean dauden osagaiek, ezkerrean daudenek ez bezala, eragina dute sintagmaren osaeran eta, beraz, determinatzaile-erroreen gainean.
 - o Mugagabea eskatzen duen zenbatzaile zehaztugabearen ondoren determinatzaile atzizkidun hitza edo hitz-segida badator, sintagma osoa etiketatuko dugu: **<DETE>hainbat gizonak*<DETE> *ikusi ditut*, **<DETE>hainbat liburu interesgarriak*<DETE> *idatzi ditu...*
 - o Sintagma osatzen duten elementuen arteko ordena okerra denean, ezinbestean markatu behar dira gaizki ordenatutako elementuak:

*<DETO>**asko** *posibilitate*<DETO> *dituzu*, *<DETO>**gutxi** *pertsona*<DETO> *ginen...*

Izen-sintagmaren barruan determinatzaile-errorea adierazten duten elementu minimoak etiketatzea da, beraz, determinatzaile-erroreen eremua definitzeko finkatu dugun irizpidea.

Badaude, ordea, aipatu beharreko hainbat kasu berezi: izenki-predikatuak (**barazkizale da nire lagun hau...*) eta aditz-esapide batzuk (**gauean loa egin dugu...*). Adibide horien errore-eremuari dagokionez, izen-sintagmaz gain ezinbestekoa da aditza ere etiketatzea. Izan ere, izenki-predikatuen kasuan aditzaren araberakoa izan ohi da determinatzaile-atzizkiaren erabilera, eta aditz-esapideak HAUL bezala tratatu behar dira. Beraz, izenki-predikatuen kasuan, errore-eremua predikatu-osagarriak eta aditzak osatuko dute; aditz-esapideen kasuan, berriz, izenak eta aditzak.

Errore motaren arabera, etiketatu beharreko errore-eremua aldatuz joan ohi da, kasuan kasu. Horregatik, garrantzitsua da errore mota bakoitzari zein eremu dagokion argi zehaztea; batetik, etiketatzaile guztiek modu berean etiketa dezaten, eta bestetik, eskuzko etiketatzea eta ondoren datorren detekzio automatikoan errore-eremuak bat etor daitezen.

Errore-eremuari dagozkion irizpideez gain, badaude erroreen etiketatzeari dagozkion beste hainbat zehaztapen:

Errorearen hasiera (HAS) eta bukaera (BUK) zehazten ditugu erroreen sailkapeneko kategoriei edota azpikategoriei dagozkien errore-kodeetan. Bada, **Euskal Herria nazioa bat da* adibidean, esaterako, errore-eremua *nazioa bat* izango da. Sintagma berean bi determinatzaile erabili direnez (*-a* eta *bat*), <DETE> (DETerminatzailea Errepikatu) kodea esleituko diogu adibide erroredunari; *nazioa* izenean ipiniko dugu errore-hasierari dagokion marka (HAS) eta *bat* determinatzailean bukaerari (BUK) dagokiona:

Euskal Herria <DETE_HAS>**nazioa bat**<DETE_BUK> *da*

Bestalde, gerta liteke esaldi edota sintagma berean errore bat baino gehiago egotea. Hitzaldi eman eta gero Miren Azkarate Kultura Sailburua komentatu du

hauzia hau guztion artean konpondu behar dugun kontua bat dela esaldian, esaterako, bost errore desberdin daude: determinatzailearik gabeko ISa (*Hitzaldi*); komunztadura eza subjektua eta aditzaren artean, kasuari dagokionean (*Kultura Sailburua komentatu du*); errore ortografikoa (*hauzia*); determinatzaile bi dituzten ISak (*hauzia hau* eta *kontua bat*). Horrelakoetan, errore guztiak etiketatzen ditugu, errore-kodeen ordena errespetatuz:

<DETK_HAS>*Hitzaldi*<DETK_BUK> *eman eta gero Miren Azkarate*
 <KOMPAS-KAS_HAS>*Kultura Sailburua komentatu du*<KOMPAS-KAS_BUK>
 <DETE_HAS><LEGHH_HAS>*hauzia*<LEGHH_BUK> *hau*<DETE_BUK>
guztionean artean konpondu behar dugun <DETE_HAS>*kontua bat*<DETE_BUK>
dela.

Badira, halaber, kategoria batean baino gehiagotan sailka daitezkeen erroreak, hau da, bi interpretazio posible dituzten errore-adibideak: **trampa* hitza, esaterako, errore ortografikoen definizioaren arabera, *Ortografikoa_Letra ordezkatzeari* (LEOENM) azpikategorian sailkatu ahalko genuke, *n* beharrean *m* hizkia erabili delako; baina *Lexikoa_Mailegua* (LEX_MAIL) azpikategorian ere sailkagarria da, gaztelaniatik mailegatutako hitz gisa. Kasu horietan, etiketatzaileari errorea azpikategoria posible guztietan sailkatzeko aukera ematea iruditu zaigu egokiena; azken finean, errore batek bi interpretazio (edo gehiago) izan baititzake:

<LEOENM_HAS><LEX_MAIL_HAS>*trampa*<LEX_MAIL_BUK><LEOENM_BUK>

Posible da, era berean, errore-eremu berean errore bat baino gehiago egotea. Esaterako, **asko bero egin zuen* esaldian, bi errore ditugu: determinatzaile-sintagmako osagaien ordena okerra (DETO) (*bero asko egin zuen* baliokide zuzenarekin) eta *asko* zenbatzailea erabili izana *handia* adjektiboa erabili beharrean (DETN) (*bero handia egin zuen* baliokide zuzenarekin):

<DETO_HAS><DETN_HAS>*asko bero*<DETO_BUK><DETN_BUK> *egin zuen*

Kategoria batean baino gehiagotan sailka daitezkeen errore-adibideek ez dute diagnosi linguistiko bakarra izango; errorea sailkatu den kategorien arabera diagnosi linguistikoak izango ditu.

Bestalde, gerta liteke etiketatzailerak errore bati zein azpikategoria dagokion ez jakitea. Zalantzazko kasuen aurrean, bi aukera ditu etiketatzailerak: azpikategoria gehiegi zehaztu gabe, adibide erroredunari dagokion kategoria nagusienean sailkatzea edo kategoria nagusi bakoitzaren barruan dagoen *Bestelakoak* azpikategorian sailkatzea. Azpiatal hori, hain zuzen, gainerako kategorietan lekurik ez duten adibideak sailkatzeko sortu da, eta bertan sailkatzen diren adibideak aztertu eta ikusiko da egitura berdinak dituzten adibideekin azpikategoria berriak era daitezkeen ala ez.

Beraz, kalitatezko etiketatzea egin nahi baldin bada, garrantzitsua da, batetik, landuko diren erroreen eremuak ongi zehaztea eta, bestetik, erroreak etiketatzeko irizpide objektibo batzuk finkatu eta etiketatzailerari lehenengo mementotik jakinaraztea, etiketatzeko-lana ahal den homogeneousena eta sistematikoena izan dadin.

VI.4. Determinatzaile-erroreen sailkapena

Tesi-lan honen helburuetako bat determinatzaile-errore batzuen azterketa linguistikoa egin eta horien detekzio automatikoa bideratzea izan da. Horretarako, euskara-ikasleen corpusean detektatutako adibideak etiketatu ahala determinatzaile-errorei dagozkien kategoriak eta azpikategoriak zehazten joan gara. Batik bat euskara-ikasleen corpusean etiketatutako errore-adibide errealak hartu ditugu oinarri. Baina gramatika-liburuetan determinatzaile-erroreen inguruan ematen diren arauak, azalpenak, oharrak eta adibideak ere kontsultatu ditugu testu errealeko adibideekin alderatu eta datuak osatzeko.

Determinatzaile-erroreen kasuistika nahiko konplexua da eta hasiera batean aurreikusitakoak baino determinatzaile-errore mota gehiago aurkitu ditugu corpusa etiketatu ahala. Sailkapena, baina, dinamikoa da eta azpikategoria berriak kentzeko, gehitzeko edota aldatzeko aukera ematen digu.

Euskarazko erroreen eta desbideratzeen sailkapen orokorrean bezala, determinatzaile-erroreen kategorian ere irizpide linguistikoa eta deskribatzailea izan ditugu kontuan: errore motaren gramatika-kategoria definitu (determinatzaile-errorea) eta irizpide deskribatzailearen arabera osatu ditugu gainerako azpikategoriak, errore mota bakoitzaren azaleko egituraren deskribazioan oinarrituta (determinatzailea(k) kentzea, gehitzea, errepikatzea...). Bederatzi

azpikategoria nagusik osatzen dute determinatzaile-erroreen sailkapena (20. irudia). Azpikategoria nagusi bakoitzaren azpian, halaber, beste hainbat azpikategoria daude, determinatzaile-erroreen ezaugarriak zehatz-mehatz jasotzen dituztenak (C eranskina). Sailkapenean, azpikategoria bakoitzari dagokion kodea, kodearen azalpena eta errore mota bakoitzaren adibideak zehazten dira:

3.3. Determinatzaile-erroreak:

- 3.3.1. **DETK** (DETerminatzailea Kendu): **txokolate nahi dut*
- 3.3.2. **DETG** (DETerminatzaile atzizkia Gehitu): **nahi adina dirua dauka*
- 3.3.3. **DETE** (DETerminatzailea Errepikatu): **mendia bat ikusten dut*
- 3.3.4. **DETO** (DSko osagaien Ordena okerra): **gutxi pertsona ginen*
- 3.3.5. **DETMUG** (MUGatasun kontuak): **edozein aldizkarietan irtetzen da*
- 3.3.6. **DETN** (DETerminatzailea/Zenbatzailea Nahastu): **jende osoak daki*
- 3.3.7. **DETGK** (DETG eta DETK sintagma berean): **mozkorra handi hartu genuen*
- 3.3.8. **DET?** (esaldi-mailan zuzenak, testu-mailan okerrak): **hamar garagardoak erosi nituen*
- 3.3.9. **DET** (zalantzazko kasuak)

20. irudia. Determinatzaile-erroreen sailkapena.

Determinatzailea kendu (DETK) azpikategorian determinatzailea behar denean ez erabiltzeagatik egiten diren erroreak sailkatu ditugu. Atal honetan zazpi azpikategoria bereizi ditugu, sintagma erroredunen azaleko egituraren arabera²⁹:

- *Izena + determinatzailea kendu* motako sintagma (3.3.1.1. azpikategoria):
 - *biok <DETK_HAS>perretxiko<DETK_BUK> jan ditugu
 - *<DETK_HAS>klase<DETK_BUK> amaitu ondoren esan zigun
 - *<DETK_HAS>txupinazo<DETK_BUK> eta gero hasi zen
 - *<DETK_HAS>albiste<DETK_BUK> atzo eman zigun
 - *bere <DETK_HAS>amodio<DETK_BUK> aitortu zidan
- *Izena + adjektiboa + determinatzailea kendu* motako sintagma (3.3.1.2. azpikategoria):

²⁹ Adibideetan, aztergai ditugun erroreez gain, bestelako errore batzuk ere ager daitezke. Errore horiek bere horretan utzi ditugu, zuzendu gabe, jatorrizko testutik hartuta bezala.

*<DETK_HAS>ile kizkur<DETK_BUK> daukala
*nuen <DETK_HAS>erresaka handi<DETK_BUK>

- *Izena + zenbatzaile orokorra + determinatzailea kendu* motako sintagma (3.3.1.3. azpikategoria):

*<DETK_HAS>gau oso<DETK_BUK> pasatu ginen farrez
*odolez beteta geratu zen <DETK_HAS>gela oso<DETK_BUK>

- *Aditz-esapideetan determinatzaile atzizkia kendu* motako sintagma (3.3.1.4. azpikategoria):

eta <DETK_HAUL_HAS>larru jo<DETK_HAUL_BUK> egin genuen

- *Postposizio-egituretan determinatzaile atzizkia kendu* motako sintagma (3.3.1.5. azpikategoria):

*<DETK_POS_HAS>"pinguino" bezala<DETK_POS_BUK> nenbiela esan dit
*Nahiz eta Daniel haundiena izan <DETK_POS_HAS>ume bezala<DETK_POS_BUK> da

- *Izenki-predikatuei determinatzaile atzizkia kendu* motako sintagma (3.3.1.6. azpikategoria):

*badirudi oso <DETK_IP_HAS>kristau dela<DETK_IP_BUK>
*oso <DETK_IP_HAS>mendizale naiz<DETK_IP_BUK>

- *Hitza deklinatzean determinatzaile atzizkia kendu* motako sintagma (3.3.1.7. azpikategoria):

*Lagunak <DETK_DEKL_HAS>perretxikok<DETK_DEKL_BUK> hartu ditu
*<DETK_DEKL_HAS>itxason<DETK_DEKL_BUK> sartu ginen

Determinatzaile atzizkia gehitu (DETG) azpikategorian sailkatu ditugu behar ez denean determinatzaile atzizkia erabiltzeagatik egiten diren erroreak. Atal honetan bost azpikategoria bereizi ditugu, sintagma erroredunen egiturak kontuan hartuta:

- *Determinatzaile atzizkia erabiltzea behar ez denean* motako sintagma (3.3.2.1. azpikategoria):

*<DETG_HAS>nahi adina dirua<DETG_BUK> dauka

- *Aditz-esapideetan determinatzaile atzizkia gehitu* motako sintagma (3.3.2.2. azpikategoria):

*<DETG_HAUL_HAS>ametsa egiten<DETG_HAUL_BUK> dut batzutan egun horrekin
*bihar eta etzi <DETG_HAUL_HAS>jaia daukat<DETG_HAUL_BUK>

- *Postposizio-egituretan determinatzaile atzizkia gehitu* motako sintagma (3.3.2.3. azpikategoria):

*<DETG_POS_HAS>mahaia inguruan<DETG_POS_BUK> ikasle pila bat biltzen ohi zen
*<DETG_POS_HAS>hiria erdian<DETG_POS_BUK> bazkaldu gara

- *Izenki-predikatuei determinatzaile atzizkia gehitu* motako sintagma (3.3.2.4. azpikategoria):

*etxe barruan <DETG_IP_HAS>iluna zegoen<DETG_IP_BUK>
*oso <DETG_IP_HAS>indartsua nago<DETG_IP_BUK>

- *Hitza deklinatzean determinatzaile atzizkia gehitu* motako sintagma (3.3.2.5. azpikategoria):

*<DETG_DEKL_HAS>herriara<DETG_DEKL_BUK> hotza eta gosea ekarri zituen
*<DETG_DEKL_HAS>herriatik<DETG_DEKL_BUK> kampinera joan ginen

Determinatzailea erreplikatu (DETE) azpikategorian sartu ditugu sintagma berean elementu bati baino gehiagori (adibidez, izenari eta adjektiboari) determinatzailea gehitzeagatik egiten diren erroreak. Kategoria honetan sartzen dira, halaber, ISaren elementu bati bi determinatzaile jartzeagatik egiten diren erroreak. Azkenik, Ø determinatzailea behar duen zenbatzailea duen sintagmari determinatzaile ageria gaineratzeagatik egiten diren erroreak ere sartu ditugu kategoria honetan. Hala, hogeita bi azpikategoria barne hartzen ditu atal honek:

- *Izena + determinatzaile atzizkia + adjektiboa + determinatzaile atzizkia* motako sintagma (3.3.3.1. azpikategoria):

*<DETE_HAS>mina haundia<DETE_BUK> sentitzen nuen
*oso <DETE_HAS>etxea polita<DETE_BUK> da
*<DETE_HAS>gauzak berriak<DETE_BUK> ezagutu nahi nituen

- *Izena + determinatzaile atzizkia + artikulua partitiboa* motako sintagma (3.3.3.2. azpikategoria):

*ez duzu <DETE_PART_HAS>soinuarik<DETE_PART_BUK> egin
*munduko <DETE_PART_HAS>gizonarik<DETE_PART_BUK> alaitsuena

- *Izena + determinatzaile atzizkia + artikulua zehaztugabea* motako sintagma (3.3.3.3. azpikategoria):

*Euskal Herria <DETE_HAS>nazioa bat<DETE_BUK> izan dela
*<DETE_HAS>araua batzuk<DETE_BUK> idatzi zituzten
*kriston giro dugu <DETE_HAS>sagardotegia batean<DETE_BUK>

- *Izena + determinatzaile atzizkia + adjektiboa + artikulua zehaztugabea* motako sintagma (3.3.3.4. azpikategoria):

*<DETE_HAS>ametsa polit bat<DETE_BUK> egin dut
*<DETE_HAS>albiste on batzuk<DETE_BUK> ekarri zituen
*eman diot <DETE_HAS>laguna on bati<DETE_BUK>

- *Izena + adjektiboa + determinatzaile atzizkia + artikulua zehaztugabea* motako sintagma (3.3.3.5. azpikategoria):

*irakurri dut <DETE_HAS>liburu polita bat<DETE_BUK>
*<DETE_HAS>jai handia batzuk<DETE_BUK> antolatu nituen

- *Izena + determinatzaile atzizkia + adjektiboa + determinatzaile atzizkia + artikulua zehaztugabea* motako sintagma (3.3.3.6. azpikategoria):

*<DETE_HAS>afaria ederra bat<DETE_BUK> prestatu zigun
*esan nizkion <DETE_HAS>gauzak politak batzuk<DETE_BUK>
*<DETE_HAS>laguna ona batzuei<DETE_BUK> eman diet

- *Izena + determinatzaile atzizkia + erakuslea* motako sintagma (3.3.3.7. azpikategoria):

*nik idatzi dut <DETE_HAS>idazlana hau<DETE_BUK>
*<DETE_HAS>eskutitza honekin<DETE_BUK> dena egiaztatuko duzu

- *Izena + determinatzaile atzizkia + adjektiboa + erakuslea* motako sintagma (3.3.3.8. azpikategoria):

*etorriko da <DETE_HAS>laguna handi hura<DETE_BUK>
*<DETE_HAS>kontraesana garbi honetan<DETE_BUK> erori dira

- *Izena + adjektiboa + determinatzaile atzizkia + erakuslea* motako sintagma (3.3.3.9. azpikategoria):

*<DETE_HAS>irudi gogorra hori<DETE_BUK> guri bota diezaguketela
*<DETE_HAS>leku polita horretara<DETE_BUK> goaz

- *Izena + determinatzaile atzizkia + adjektiboa + determinatzaile atzizkia + erakuslea* motako sintagma (3.3.3.10. azpikategoria):

*<DETE_HAS>mendia altua hau<DETE_BUK> igo dugu

*<DETE_HAS>laguna ona hark<DETE_BUK> esan dit

- *Izena + determinatzaile atzizkia + zenbatzaile orokorra + determinatzaile atzizkia* motako sintagma (3.3.3.11. azpikategoria):

*eta <DETE_HAS>gerrak guztiak<DETE_BUK> bukatuta dira

*<DETE_HAS>azalak guztiak<DETE_BUK> erori ziren

- *Izena + determinatzaile atzizkia + zenbatzaile zehaztugabea* motako sintagma (3.3.3.12. azpikategoria):

*<DETE_HAS>argazkia asko<DETE_BUK> atera genituen

*egun honetan <DETE_HAS>frogak asko<DETE_BUK> egin dizkizugu

- *Zenbatzaile zehaztugabea + izena + determinatzaile atzizkia* motako sintagma (3.3.3.13. azpikategoria):

*<DETE_HAS>zenbait gizona<DETE_BUK> etorri da

*amestutako <DETE_HAS>zenbait gauzak<DETE_BUK> egia bihurtu dira

- *Zenbatzaile zehaztugabea + izena + determinatzaile atzizkia + adjektiboa* motako sintagma (3.3.3.14. azpikategoria):

*<DETE_HAS>hainbat gizona zahar<DETE_BUK> etorri da

*udan <DETE_HAS>zenbait laguna berri<DETE_BUK> egin nituen

- *Zenbatzaile zehaztugabea + izena + adjektiboa + determinatzaile atzizkia* motako sintagma (3.3.3.15. azpikategoria):

*<DETE_HAS>zenbait teknika berriak<DETE_BUK> debekatu egin zituztelako

- *Zenbatzaile zehaztugabea + izena + determinatzaile atzizkia + adjektiboa + determinatzaile atzizkia* motako sintagma (3.3.3.16. azpikategoria):

*<DETE_HAS>zenbait mendia altua<DETE_BUK> igo ditugu

*<DETE_HAS>hainbeste gauzak ezberdinak<DETE_BUK> ikusteagatik

- *Determinatzaile zehaztugabea, nolakotzailea edo galdetzailea + izena + determinatzaile atzizkia* motako sintagma (3.3.3.17. azpikategoria):

*Vignemal <DETE_HAS>edozein mendia<DETE_BUK> baino politagoa da

*<DETE_HAS>edozein lanpostua<DETE_BUK> hartuko luke

- *Determinatzaile zehaztugabea, nolakotzailea edo galdetzailea + izena + determinatzaile atzizkia + adjektiboa + determinatzaile atzizkia* motako sintagma (3.3.3.18. azpikategoria):

*<DETE_HAS>zein plana polita<DETE_BUK> egin genuen

- *Zenbakia + izena + determinatzaile atzizkia* motako sintagma (3.3.3.19. azpikategoria):

*<DETE_HAS>Bi posibilitatea<DETE_BUK> dauzkate

*<DETE_HAS>Bost gola<DETE_BUK> sartu genuen eta irabazi genuen

- *Zenbakia + izena + determinatzaile atzizkia + adjektiboa* motako sintagma (3.3.3.20. azpikategoria):

*aurkitu nituen <DETE_HAS>bi txakurra polit<DETE_BUK>

*<DETE_HAS>bi herria desberdin<DETE_BUK> bisitatu ditu

- *Zenbakia + izena + adjektiboa + determinatzaile atzizkia* motako sintagma (3.3.3.21. azpikategoria):

*aurkitu nituen <DETE_HAS>bi txakur polita<DETE_BUK>

*<DETE_HAS>lau herri desberdina<DETE_BUK> pasatu genituen

- *Zenbakia + izena + determinatzaile atzizkia + adjektiboa + determinatzaile atzizkia* motako sintagma (3.3.3.22. azpikategoria):

*aurkitu nituen <DETE_HAS>bi txakurra polita<DETE_BUK>

*nire <DETE_HAS>hiru lagunak onak<DETE_BUK> etorriko dira

Determinatzaile-sintagmako osagaien ordena okerra (DETO) azpikategoriari dagozkio sintagma bateko buruaren eta determinatzailearen arteko ordena aldrebesa duten egiturak. Atal honetan bi azpikategoria sortu ditugu, orain artean:

- *Zenbatzaile zehaztugabea + izena* motako sintagma (3.3.4.1. azpikategoria):

*<DETO_HAS>asko posibilitate<DETO_BUK> dituzu

*<DETO_HAS>gehiago liburu<DETO_BUK> irakurri behar ditut

- *Artikulu zehaztugabea + izena* motako sintagma (3.3.4.2. azpikategoria):

*eta <DETO_HAS>batzuk pintxo<DETO_BUK> jan ditugu

Mugatasun-erroreak (DETMUG) azpikategorian sailkatu ditugu zenbakien, zenbatzaile zehaztugabeen edota determinatzaile zehaztugabeen atzetik datozen elementuak mugagabeen erabili ez izanagatik sortzen diren erroreak. Hiru azpikategoria daude hemen:

- *Zenbatzaile zehaztugabea + izena + determinatzaile atzizkia* motako sintagma (3.3.5.1. azpikategoria):

*<DETMUG_HAS>zenbait kasuetan<DETMUG_BUK> izango du baliotasuna

*<DETMUG_HAS>hainbat gizonarekin<DETMUG_BUK> etorri da

- *Determinatzaile zehaztugabea + izena + determinatzaile atzizkia* motako sintagma (3.3.5.2. azpikategoria):

*<DETMUG_HAS>Edozein aldizkarietan<DETMUG_BUK> ia astero irtetzen da

*<DETMUG_HAS>Edozein pertsonen<DETMUG_BUK> minak gure gupida merezi du

- *Zenbakia + izena + determinatzaile atzizkia* motako sintagma (3.3.5.3. azpikategoria):

*<DETMUG_HAS>hiru kalimotxoekin<DETMUG_BUK> kristona harrapatzen genuen

*<DETMUG_HAS>bost lagunarekin<DETMUG_BUK> etorriko da

Determinatzailea edo zenbatzailea nahastu (DETN) azpikategorian (3.3.6. azpikategoria) sailkatu ditugu determinatzaile edo zenbatzaile bat determinatzaile, zenbatzaile edota beste elementuren batekin nahasteagatik egiten diren errore motak:

*Itsasoko <DETN_HAS>ur osoa<DETN_BUK> edango nuke (*ur guztia* esan beharrean)

*han <DETN_HAS>bero asko<DETN_BUK> egin zuen (*bero handia* esan beharrean)

Sintagma berean determinatzailea gehitu eta kendu (DETGK) azpikategoriari dagozkio sintagmako buruari determinatzailea gehitu eta jarraian datorren adjektiboari kentzeagatik egiten diren erroreak (3.3.7. azpikategoria):

*eta <DETGK_HAS>mozkorra haundi<DETGK_BUK> harrapatu genuen

Esaldi-mailan zuzentzat eman daitezkeen arren testuinguruan erroredunak diren egiturak (DET?) 3.3.8. azpikategorian sailkatu ditugu. Galdera ikurrarekin (?) adierazten dugu, hain zuzen, esaldi-mailan gramatikalak diren arren, testu-mailan erroredunak diren adibideak direla:

- *nik <DET?_HAS>lau txuletak<DET?_BUK> egin nituen
- *<DET?_HAS>10 garagardoak<DET?_BUK> erosi nituen
- *burura etortzen zitzaizkion <DET?_HAS>hamaika irudiak<DET?_BUK>

DET? azpikategorian sailkatu ditugu, halaber, zalantzazkoak edota anbiguoak izan daitezkeen Hitz Anitzeko Unitate Lexikal (HAUL) batzuk³⁰. Adibidez, *asko negar egiten dut* esaldia gramatikala da; baina okertzat ere har liteke testuingurua kontuan hartuta *negar asko egiten dut* baliokide zuzena dagokiola ikusten bada. Beraz, eskuzko etiketatzean, erroreen testuingurua kontuan hartuta, *asko negar egiten dut* esatea baino *negar asko egiten dut* esatea egokiagoa dela erabaki dezake etiketatzailerak. Horrelako adibideak DET? azpikategorian sailkatuko ditu. Automatikoki testuinguru osoa kontuan hartzea, ordea, ezinezkoa zaigu oraindik, eta analizatzaile sintaktikoaren analisiaren arabera hartuko dira erroretzat edo zuzentzat *asko negar egiten dut* moduko esaldiak. Analizatzaile sintaktikoak ematen digun analisiak, beraz, badu eraginik horrelako adibideen tratamendu automatikoan, batzuetan HAUL gisa markatzen baitira eta beste batzuetan ez. Horren arabera, *asko negar egiten dut* esaldia zuzentzat hartuko da *negar* eta *egin* (izena eta aditza) HAUL bezala analizatuta badatoz (*negar_egin*), *asko* adberbio duela; *negar* eta *egin* bereiz analizatuta badatoz, aldiz, determinatzaile-sintagmako osagaien ordena okerrari dagokion errorea (DETO) detektatuko da (*<DETO_HAS>*asko negar<DETO_BUK> egiten dut*), erregeletan aurkakorik zehazten ez bada.

Beraz, erroreen eskuzko etiketatzean galdera ikurrarekin (?) markatu ditugu esaldi-mailan zuzentzat eman daitezkeen arren, testu-mailan erroredunak izan daitezkeen adibideak. Horrelako esaldien tratamendu automatikoa oraindik bidera ezina zaigun arren, adibideak etiketatu eta bildu egin ditugu, etorkizunean testu mailako azterketak egiteko baliagarriak izango zaizkigulakoan.

Determinatzaile-erroreen sailkapenean *bestelakoak* (DET) azpikategoria (3.3.9. azpikategoria) ere gehitu dugu, gainontzeko azpikategoriei ez dagozkien adibideak etiketatzeko. Azpikategoria honetan sailkatutako erroreak aztertu egingo

³⁰ *negar egin, hitz egin, lan egin, lo egin...* bezalako egiturak.

dira ondoren, eta ikusiko da adibide horiekin azpikategoria berriak sortzea merezi duen:

- *guk ez ditugu <DET_HAS>dirurik asko<DET_BUK>
- *hamabietatik ordu bitara <DET_HAS>potek asko<DET_BUK> hartu genituen

Azkenik, determinatzaile-erroreekin batera -A organikoa kentzeagatik egin diren erroreak ere etiketatu ditugu eskuz, AORG kodepean:

- *Karibetik <AORG_HAS>bidai<AORG_BUK> bat egin nahi genuen
- *eta <AORG_HAS>gauz<AORG_BUK> batzuk gehiago
- *saiatu gara <AORG_HAS>animali<AORG_BUK> bat edukitzen

Adibide horiek errore ortografikoen kategoriakoak (sailkapeneko 1.5. azpikategoriakoak) diren arren, determinatzaile-erroreekin nolabaiteko lotura badutela ikusten dugu: -A organikoa -a determinatzaile atzizkiarekin nahastu izana. Hitz-mailako erroreak direnez, zuzentzaile ortografikoari dagozkion adibideak dira. Baina horrelako hitz batzuek -A organikoa inongo testuingurutan galdu ezin dezaketen bitartean (*gauza*, *pertsona...*), -ia amaiera dutenek (*animalia*, *familia...*), esaterako, -A galtzeko aukera dute hitz-elkarketan. Ordea, hitz bat hitz elkartu gisa erabili den ala ez detektatzeko, ezinbestekoa da testuingurua kontuan hartzea, eta zuzentzaile ortografikoak ez du halakorik egiten. Beraz, inongo kasutan -A organikoa gal ezin dezaketen hitzak erroretzat hartzen ditu XUXEN-ek, baina -A gabeko formak ere onartzen dituzten hitzak, deklinatuta ez daudenean, ez ditu detektatzen oraindik (**animali asko ikusi genituen*)³¹. Adibide horien tratamendu automatikoa aurreragoko urratsetan egin beharko da. Hala ere, interesgarria iruditua zaigu horrelako adibideak etiketatzea, batik bat hizkuntzen i(ra)kaskuntzaren alorrean dauden arazo lexikoak lantzeko.

Bada, batez ere testuetan aurkitutako errore-adibideak kontuan hartuta zehaztu eta osatu dugu determinatzaile-erroreen sailkapena (C eranskina). Azpikategoria bakoitzari errore mota bat dagokio eta errore mota bakoitzaren ezaugarriak oinarri

³¹ Hitz horiek deklinatuta baldin badaude, ordea, detektatzen ditu zuzentzaileak (**bidaiko*, **animaliren...*).

hartuta bideratu dugu ondoren determinatzaile-errore batzuen detekzio automatikoa (VII. kapitulua).

VI.5. Determinatzaile-erroreen eskuzko etiketatzea: datuak

Askotan, eskura dauden bitartekoak medio, ezinezkoa da corpusean errore mota guztiak aldi berean etiketatzea. Ondorioz, erroreen analisia urratsez urrats egin behar izaten da, errore motaka. Tesi-lan honetan, hain zuzen, determinatzaile-erroreak ditugu aztergai, eta kategoria horretako errore batzuen etiketatzea, azterketa linguistikoa eta tratamendu automatikoa bideratzeko egin dugun lana aurkezten dugu.

Eskuzko etiketatzean, 17 determinatzaile-errore mota eta -A organikoa kentzeagatik egindako erroreak³² etiketatu ditugu, adibide bakoitzari dagokion kodea esleituz (13. taula):

Errore mota	Errore-kodea
3.3.1. Determinatzailea kendu	DETK
3.3.1.1. Determinatzaile atzizkia kendu Hitz Anitzeko Unitate Lexikaletan	DETK_HAUL
3.3.1.2. Determinatzaile atzizkia kendu postposizio-egituretan	DETK_POS
3.3.1.3. Determinatzaile atzizkia kendu izenki-predikatueta	DETK_IP
3.3.1.4. Determinatzaile atzizkia kendu hitza deklinatzean	DETK_DEKL
3.3.2. Determinatzaile atzizkia gehitu	DETG
3.3.2.1. Determinatzaile atzizkia erabiltzea behar ez denean	DETG
3.3.2.2. Determinatzaile atzizkia gehitu Hitz Anitzeko Unitate Lexikaletan	DETG_HAUL
3.3.2.3. Determinatzaile atzizkia gehitu postposizio-egituretan	DETG_POS
3.3.2.4. Determinatzaile atzizkia gehitu izenki-predikatueta	DETG_IP
3.3.2.5. Determinatzaile atzizkia gehitu hitza deklinatzean	DETG_DEKL
3.3.3. Determinatzailea errepikatu	DETE
3.3.4. Determinatzaile-sintagmako osagaien ordena okerra	DETO
3.3.5. Determinatzailearekin zerikusia duten mugatasun-kontuak	DETMUG
3.3.6. Determinatzailea edo zenbatzailea nahastu	DETN
3.3.7. Sintagma berean determinatzailea gehitu eta kendu	DETGK
3.3.8. Esaldi-mailan zuzenak baina testuinguruan errore direnak	DET?
3.3.9. Bestelakoak	DET
-A organikoa	AORG

13. taula. Etiketatutako errore motak eta bakoitzari dagokion errore-kodea³³.

³² -A organikoa falta duten izenak errore ortografikoak dira berez, ez determinatzaile-erroreak. Baina -A organikoa -a determinatzaile atzizki singularrarekin nahastu ohi denez, determinatzaile-erroreekin batera etiketatu dugu errore mota hori, eta eskuzko etiketatzeari dagozkion datuetan kontuan izan dugu.

Determinatzaile-erroreen eskuzko etiketatze-lana bi corpus-motatan egin dugu: euskara-ikasleen corpusean eta espezialitate-euskarako ikasleen testuetan. Hizkuntza-ikasleek zenbat eta zein motatako determinatzaile-errore egiten dituzten ikusi nahi izan dugu beraien testuetan; euskara-ikasleek egindako determinatzaile-erroreen kopurua alderatu nahi izan dugu, ondoren, espezialitate-euskarako ikasleek egindakoekin. Corpus motaren arabera aurkeztuko ditugu jarraian eskuzko etiketatzeari dagozkion zenbait datu. Determinatzaile-erroreak sintagma-mailako erroreak direnez, datuak corpuseko izen-sintagma (IS) kopuruaren arabera kalkulatu ditugu³⁴.

Euskara-ikasleen corpora hiru hizkuntza-mailatan banatu dugu³⁵: 1. mailako testuak, 2. mailakoak eta 3. mailakoak (14. taula), maila batetik bestera dauden aldaketak edota desberdintasunak ikusteko:

Hizkuntza-maila	Hitz kopurua	IS kopurua
1. maila (behe-maila)	39.117	13.682
2. maila (erdi-maila)	42.219	15.078
3. maila (goi-maila)	31.954	9.362
Guztira	113.290	38.122

14. taula. Eskuz etiketatu den euskara-ikasleen corpora.

Euskara-ikasleen corpusean eskuz etiketatutako erroreen portzentajea, IS kopuruarekiko, % 2,07koa da: % 2,73koa behe-mailako testuetan, % 1,61ekoa erdi-mailakoetan eta % 1,81ekoa goi-mailakoetan (15. taula):

³³ Letra lodiz azpimarratutakoak determinatzaile-erroreen sailkapenari dagozkion kategoria nagusiak dira, gainontzekoak azpikategoriak. Hainbat azpikategoria dituzten DETK eta DETG kategorien kasuan, datuak bateratuta emango ditugu, azpikategoria bakoitzeko errore-kopurua ez baita oso handia. Hala ere, D eranskinean ikusgai daude azpikategoria bakoitzari dagozkion adibideak.

³⁴ Corpuseko IS kopurua automatikoki kalkulatu da.

³⁵ HEOKek (HABE, 1999) definitutako mailaketa (IV.5.2. atala).

Hizkuntza-maila	Errore kop.	IS kop.	%
1. maila (behe-maila)	374	13.682	2,73
2. maila (erdi-maila)	244	15.078	1,61
3. maila (goi-maila)	170	9.362	1,81
Guztira	788	38.122	2,07

15. taula. Hizkuntza-maila bakoitzean etiketatutako determinatzaile-erroreak, corpuseko IS kopuruarekiko.

Hizkuntza-maila batetik bestera, errore kopuruak eta motak aldatu egin ohi dira, oro har: badaude behe-mailan asko egin arren hizkuntza menderatu ahala desagertzen diren erroreak eta, alderantziz, behe-mailan egiten ez diren arren hizkuntza menderatu ahala (hizkuntzaren egitura berri eta konplexuagoak ikasteen) agertzen hasten diren errore motak. Determinatzaile-erroreen kasuan, 1. mailako ikasleen testuetan etiketatu da errore gehien (% 2,73); 2. mailako ikasleen testuetan erroreen kopurua jaitsi egiten da (% 1,61) eta 3. mailakoetan, berriz, pixka bat igotzen da (% 1,81).

Guztira, 788 determinatzaile-errore etiketatu dira eskuz euskara-ikasleen corpusean; horietatik 374 (% 47,46) behe-mailako testuetan, 244 (% 30,97) erdi-mailako testuetan eta 170 (% 21,57) goi-mailakoetan (16. taula).

Hizkuntza-maila	Errore kop.	%
1. maila (behe-maila)	374	47,46
2. maila (erdi-maila)	244	30,97
3. maila (goi-maila)	170	21,57
Guztira	788	100

16. taula. Hizkuntza-maila bakoitzean etiketatutako determinatzaile-erroreak, guztira etiketatutako errore kopuruarekiko.

Etiketatatutako errore motei dagokienez, oro har gehien egin diren erroreak DETK (determinatzailea kentzea) eta DETE (sintagma berean determinatzailea errepikatzea) dira. 17. taulan ikus daitekeen moduan, 788 adibideetatik 327 (% 41,50) dira DETK kategoriakoak eta 208 (% 26,39) DETE kategoriakoak. 101 errore (% 12,82) dagozkio AORG (-A organikoa kentzea) kategoriari eta 58 (% 7,36) DETG

(determinatzaile atzizkia gehitu) kategoriari³⁶. DET? kodepean (esaldi mailan zuzenak diruditen arren testu-mailan okerrak diren kasuak) 40 adibide (% 5,08) etiketatu dira eskuz eta DETO (determinatzaile-sintagmako osagaien ordena okerra) kodepean 27 (% 3,43):

Errore mota	Errore kop.	%
DETK	327	41,50
DETG	58	7,36
DETE	208	26,39
DETO	27	3,43
DETMUG	11	1,39
DETN	10	1,27
DETGK	3	0,38
DET?	40	5,08
DET	3	0,38
AORG	101	12,82
	788	100

17. taula. Etiketatutako determinatzaile-errore motak, guztira etiketatutako errore kopuruarekiko³⁷.

Bestalde, hizkuntza-maila batetik bestera determinatzaile-erroreen kopurua eta egiten diren errore motak nola aldatzen diren ikusi nahi izan dugu. Esan bezala, behe-mailako testuetan etiketatu da determinatzaile-erroreen kopuru handiena (% 2,73); erdi- eta goi-mailako testuetan, berriz, behe-mailakoetan baino errore gutxiago etiketatu da (% 1,61 eta % 1,81, hurrenez hurren). Beraz, hizkuntza-maila baxuetan erdi- eta goi-mailetan baino determinatzaile-errore gehiago egiten direla ikusten dugu. Errore motei dagokienez, datu orokorrekin bat eginez, DETK eta DETE erroreen kopurua jaitsi egiten da 1. mailako testuetatik (% 1,09 eta % 0,77)

³⁶ DETK eta DETG errore-kategoriek hainbat azpikategoria barne hartzen dituzte. DETK barruko azpikategorien errore kopuruak hauek dira (guztira etiketatutako errore kopuruarekiko): DETK azpikategoria nagusiari dagozkio 224 errore (% 28,42), DETK_HAUL azpikategoriari 2 (% 0,25), DETK_POS azpikategoriari 4 (% 0,50), DETK_IP azpikategoriari 59 (% 7,49) eta DETK_DEKL azpikategoriari 38 (% 4,82). DETG barruko azpikategorien errore kopuruak, berriz, hauek dira: DETG azpikategoria nagusiari dagozkio 6 errore (% 0,76), DETG_HAUL azpikategoriari 4 (% 0,50), DETG_POS azpikategoriari 4 (% 0,50), DETG_IP azpikategoriari 23 (% 2,92) eta DETG_DEKL azpikategoriari 21 (% 2,66).

³⁷ Letra lodiz nabarmendu ditugu gehien etiketatu diren errore moten kopuruak.

2. mailakoetara (% 0,68 eta % 0,34), baina 3. mailako testuetan 2. mailakoetan baino DETK eta DETE motako errore gehiago etiketatu dira (% 0,80 eta % 0,53). DETG eta DET? kodepean etiketatutako erroreen kopurua gutxituz doa hizkuntza-mailan aurrera egin ahala. DETO eta DETN kategoriako erroreak ere beheko mailan egiten dira batik bat eta hizkuntza-mailan aurrera egin ahala desagertuz doaz. AORG kategoriako errore gehienak erdi-mailako testuetan etiketatu dira (18. taula).

Errore mota	Hizkuntza-maila eta errore kopuruak					
	Behe	%	Erdi	%	Goi	%
DETK	149	1,09	103	0,68	75	0,80
DETG	28	0,20	16	0,11	14	0,15
DETE	106	0,77	52	0,34	50	0,53
DETO	23	0,17	4	0,03	0	0
DETMUG	2	0,01	4	0,02	5	0,05
DETN	8	0,06	2	0,01	0	0
DETGK	3	0,02	0	0	0	0
DET?	22	0,16	13	0,09	5	0,05
DET	3	0,02	0	0	0	0
AORG	30	0,22	50	0,33	21	0,22
	374	2,73	244	1,61	170	1,81

18. taula. Hizkuntza-maila bakoitzean etiketatutako errore motak, maila bakoitzeko IS kopuruarekiko.

Maila bakoitzean etiketatutako errore kopuruari dagokionez, behe-mailako testuetan, etiketatu diren 374 erroreetatik 149 (% 39,85) DETK kategoriari dagozkio, 106 (% 28,34) DETE kategoriari, 28 (% 7,49) DETG kategoriari, 30 (% 8,02) AORG kategoriari, 23 (% 6,15) DETO kategoriari eta 22 (% 5,88) DET? kategoriari. Erdi-mailako testuetan, berriz, etiketatu diren 244 errore-adibideetatik, gehien egin diren erroreak DETK kategoriakoak dira, 103 agerpenekin (% 42,21); 52 (% 21,31) dira DETE kategoriakoak, 50 (% 20,49) AORG kategoriakoak eta 16 (% 6,55) DETG kategoriakoak. Eta goi-mailako testuei dagokienez, DETK motako erroreak dira ugariak: 170 adibideetatik 75 (% 44,12); 50 adibide (% 29,41) dagozkio DETE kategoriari, 21 (% 12,36) AORG kategoriari eta 14 (% 8,23) DETG kategoriari. Hizkuntza-ikasleen artean, beraz, determinatzailea kentzea (DETK) eta errepikatzea (DETE) dira errore mota ohikoenak, maila guztietan; eta horien atzetik, determinatzaile atzizkia gehitzeagatik (DETG) eta izenari -A organikoa kentzeagatik (AORG) egindako erroreak. 19. taulak erakusten dizkigu hizkuntza-maila bakoitzean etiketatutako errore moten kopuruak eta portzentajeak:

Errore mota	Hizkuntza-maila eta errore kopuruak					
	Behe	%	Erdi	%	Goi	%
DETK	149	39,85	103	42,21	75	44,12
DETG	28	7,49	16	6,55	14	8,23
DETE	106	28,34	52	21,31	50	29,41
DETO	23	6,15	4	1,64	0	0
DETMUG	2	0,53	4	1,64	5	2,94
DETN	8	2,14	2	0,82	0	0
DETGK	3	0,80	0	0	0	0
DET?	22	5,88	13	5,33	5	2,94
DET	3	0,80	0	0	0	0
AORG	30	8,02	50	20,49	21	12,36
	374	100	244	100	170	100

19. taula. Hizkuntza-maila bakoitzean etiketatutako errore motak, maila bakoitzeko errore kopuruarekiko.

Bestalde, sei euskaltegitatik (A, B, C, D, E eta F euskaltegiak) jasotako testuak etiketatu dira eskuz. A euskaltegiko testuetan etiketatu da errore gehien (% 2,72), euskaltegi horretako testu gehienak behe-mailako ikasleek idatzitakoak baitira; eta D euskaltegiko testuetan daude errore gutxien (% 1,42), euskaltegi horri erdi- eta goi-mailako testuak baitagozkio (20. taula).

Euskaltegia	Hitz kop.	IS kop.	Errore kop.	%
A	40.325	13.907	379	2,72
B	28.092	10.301	158	1,53
C	23.449	6.416	125	1,94
D	10.311	3.580	51	1,42
E	7.956	2.846	57	2
F	3.157	1.072	18	1,67
Guztira	113.290	38.122	788	

20. taula. Euskaltegi bakoitzeko testuetan etiketatutako determinatzaile-erroreak, maila bakoitzeko IS kopuruarekiko.

Etiketatuak errore kopuruei eta datuei dagokienez, 788 erroreetatik A euskaltegiko testuetan etiketatu da erroreen % 48,10; B euskaltegiko testuetan % 20,05; C-ko testuetan % 15,86; D-koetan % 6,47, E-koetan % 7,23 eta F euskaltegiko testuetan, azkenik, % 2,29 (21. taula).

Euskaltegia	Errore kop.	%
A	379	48,10
B	158	20,05
C	125	15,86
D	51	6,47
E	57	7,23
F	18	2,29
Guztira	788	100

21. taula. Euskaltegi bakoitzeko testuetan etiketatutako determinatzaile-erroreak, guztira etiketatutako determinatzaile-errore kopuruarekiko.

Errore mota	A		B		C		D		E		F	
	Kop.	%	Kop.	%	Kop.	%	Kop.	%	Kop.	%	Kop.	%
DETK	150	39,58	65	41,14	66	52,8	14	27,46	27	47,37	5	27,78
DETG	28	7,38	9	5,70	11	8,8	4	7,84	6	10,53	0	0
DETE	104	27,44	40	25,31	28	22,4	19	37,25	10	17,54	7	38,89
DETO	25	6,60	0	0	2	1,6	0	0	0	0	0	0
DETMUG	2	0,54	2	1,26	1	0,8	2	3,92	2	3,52	2	11,11
DETN	9	2,37	1	0,63	0	0	0	0	0	0	0	0
DETGK	3	0,79	0	0	0	0	0	0	0	0	0	0
DET?	21	5,54	8	5,07	3	2,4	3	5,88	3	5,26	2	11,11
DET	3	0,79	0	0	0	0	0	0	0	0	0	0
AORG	34	8,97	33	20,89	14	11,2	9	17,65	9	15,78	2	11,11
	379	100	158	100	125	100	51	100	57	100	18	100

22. taula. Euskaltegi bakoitzeko testuetan etiketatutako errore motak, euskaltegi bakoitzean egindako determinatzaile-errore kopuruarekiko.

Euskaltegien arteko erroreak aztertzea interesgarria izan liteke eskuratutako datuetan desberdintasun nabarmenak antzemanaz gero, euskaltegietan jarraitzen diren ikas-estrategiek erroreetan eraginik baduten ikertzeko. Izan ere, gerta liteke errore mota batzuk euskaltegi batzuetan beste batzuetan baino gehiago egitea. Eta hori gertatuz gero, ondoriozta liteke euskaltegi bakoitzean jarraitutako metodologiak eragina izan lezakeela hizkuntzaren i(r)a kaskuntzan. Gure kasuan, determinatzaile-erroreen datu orokorrak ikusita, euskaltegien arteko kopuruak alderatzeak ez digu ondorio zehatzetara heltzeko aukerarik eman. Horren arrazoi bat izan liteke euskaltegietatik jasotako testu guztiak neurri eta hizkuntza-maila berekoak ez izatea.

Euskara-ikasleen testuak etiketatzeaz gain, espezialitate-euskarako ikasleek idatzitako hainbat testu ere etiketatu ditugu eskuz, Euskara Teknikoko ikasleen

lanak, hain zuzen. Ikasle horien artean euskara-maila desberdina duten hiztunak daude, baina denak ari dira menderatzen ez duten hizkera berezitu bat ikasten. Euskara Teknikoko ikasleen testuetan etiketatutako erroreen kopurua txikia da, 59 adibide baino ez dira etiketatu (% 0,98) (23. taula); euskara-ikasleek egindako erroreen kopurua (% 2,07) baino are baxuagoa.

Testu mota	Hitz kop.	IS kop.	Errore kop.	%
Euskara Teknikoko ikasleen testuak	19.391	5.963	59	0,98

23. taula. Euskara Teknikoko ikasleen testuetan etiketatutako determinatzaile-erroreak, corpuseko IS kopuruarekiko.

Errore motei dagokienez, DETK motakoak dira, nabarmenki, Euskara Teknikoko ikasleen testuetan etiketatutako errore ohikoenak: 59 erroreetatik 41 (% 69,49) dira, hain zuzen, DETK kategoriakoak³⁸. 5 adibide erroredun (% 8,47) etiketatu dira DETMUG kategoriakoak, 4 adibidetan (% 6,78) gehitu da determinatzaile atzizkia behar ez denean (DETG) eta 3 adibidetan (% 5,08) errepikatu da determinatzailea sintagman (DETE). Bestalde, DETO, DET?, DET eta AORG errore moten kopuruak agerpen batekoak edo bikoak dira, eta Euskara Teknikoko ikasleen testuetan ez da detektatu DETN eta DETGK kodeei dagokien errorerik. 24. taulan zehazten ditugu etiketatutako errore mota bakoitzari dagozkion kopuruak eta portzentajeak.

³⁸ DETK barruko azpikategorien errore kopuruak hauek dira: DETK azpikategoria nagusiari dagozkio 16 errore (% 27,11), DETK_HAUL azpikategoriari errore bat (% 1,69), DETK_IP azpikategoriari 7 (% 11,86) eta DETK_DEKL azpikategoriari 17 (% 28,81).

Errore mota	Errore kop.	%
DETK	41	69,49
DETG	4	6,78
DETE	3	5,09
DETO	1	1,69
DETMUG	5	8,48
DETN	0	0
DETGK	0	0
DET?	2	3,39
DET	1	1,69
AORG	2	3,39
	59	100

24. taula. Euskara Teknikoko ikasleen testuetan etiketatutako determinatzaile-erroreak, guztira etiketatutako determinatzaile-errore kopuruarekiko.

Oro har, beraz, etiketatu den determinatzaile-erroreen kopurua, corpuseko IS kopuruarekiko, ez da handia: % 2,07 euskara-ikasleen corpusean eta % 0,98 Euskara Teknikoko ikasleen testuetan. Hala ere, euskara-ikasleen testuetan eskuz etiketatutako 788 determinatzaile-erroreak (D eranskina) eta Euskara Teknikoko ikasleen testuetan etiketatutako 59 adibideak (D eranskina) ezinbestekoak izan zaizkigu elementu linguistiko honen inguruan egin ohi diren erroreen bilduma osatu eta sortzen diren arazo nagusiak edota usukoenak ezagutzen hasteko. Euskara-ikasleen corpusean etiketatutako errore-adibideak abiapuntutzat hartuta, determinatzaile-erroreen sailkapena fintzen joan gara, azpikategoria berriak sortuz. Eta eskuz etiketatutako eta sailkatutako adibide erroredunen arabera sortu ditugu, ondoren, errore mota batzuen detekzio automatikorako erregelak; alegia, adibide errealak oinarri hartuta egin ditugu determinatzaile-errore jakin batzuk automatikoki detektatzeko erregelak.

Determinatzaile-erroreen kopurua oso handia ez den arren, aztertu beharreko errore mota da hau, gramatika-errore larritzat har baitaitezke hizkuntza-maila eta testu mota guztietan. Hala, maiztasunagatik baino larritasunagatik hautatu dugu errore mota honen eskuzko etiketatzea, azterketa linguistikoa eta ondorengo tratamendu automatikoa, euskararako garatzen ari garen gramatika-zuzentzaileak kontuan hartu beharreko egitura edo elementu linguistikoa delako.

VI.6. Laburbilpena

Hizkuntza-ikasleen corpusean askotariko erroreak daude, kopuru eta mota guztietakoak. Errore guztiak aldi berean etiketatzea, baina, ezinezkoa zaigu eskura ditugun baliabideekin. Ondorioz, errorez errore ekin diegu euskarazko erroreen eta desbideratzeen azterketa linguistikoari eta tratamendu automatikoari.

Kapitulu honetan euskara-ikasleen corpusean eskuz etiketatu ditugun determinatzaile-erroreen analisia aurkeztu dugu: erroreak etiketatzeko finkatu ditugun irizpideak azaldu eta azterketa horretatik atera ditugun hainbat datu ezagutarazi ditugu. Gure ikerketen abiapuntua euskara-ikasleen corpora da batik bat, baina hainbat lanetan aipatzen diren oharrak eta erroreak ere kontuan izan ditugu determinatzaile-erroreen azterketa linguistikoa egiteko.

Erroreen eskuzko etiketatzea egiteko, bi corpus-mota baliatu ditugu: euskara-ikasleen corpora (ahalik eta errore-adibide gehien eskuratzearen) eta Euskara Teknikoko ikasleek idatzitako testuak. Guztira 17 determinatzaile-errore mota etiketatu dira, baita -A organikoa kentzeagatik egindako erroreak ere. Euskara-ikasleen corpusean 788 adibide etiketatu dira (% 2,07) eta Euskara Teknikoko ikasleen lanetan 59 (% 0,98). Espero bezala, determinatzaile-errore gehiago egiten da, beraz, hizkuntza-ikasleen testuetan.

Euskara-ikasleen corpusari dagokionez, 1. mailako testuetan etiketatu da errore gehien (% 2,73); 2. mailako testuetan erroreen kopurua jaitsi egiten da (% 1,61) eta 3. mailakoetan, berriz, pixka bat igotzen da (% 1,81). Errore motari dagokionez, determinatzailea kentzea (DETK) eta errepikatzea (DETE) dira, nabarmenki, hiru hizkuntza-mailetan eta euskaltegi guztietan gehien egiten diren errore motak (% 41,50 eta % 26,39, hurrenez hurren); eta horien atzetik -A organikoa kentzea (AORG) eta determinatzaile atzizkia gehitzea (DETG) (% 12,82 eta % 7,36, hurrenez hurren). Euskara Teknikoko ikasleen testuetan ere DETK errore mota da ohikoena, baina hiztun horien testuetan DETE motako errore gutxi etiketatu dira (% 5,09), hizkuntza-ikasleen testuetan etiketatutakoen aldean (% 26,39). Badirudi, beraz, DETE errorea euskara-ikasleei dagokiela gehiago.

Aztertutako testuetan determinatzaile-erroreen kopurua oso handia izan ez arren, erroreen azterketa eta tratamendu automatikoaren alorrean landu beharreko errore

mota da hau. Izan ere, determinatzaile-erroreak gramatika-erroreak dira. Arauen urraketak ez-gramatikaltasunera garamatza eta, beraz, garrantzitsua da euskararako garatzen ari garen gramatika-zuzentzaileak determinatzaile-erroreak barne hartzea. Hizkuntzen i(r)a(k)askuntzaren alorrean ere interesgarria da eskuz etiketatutako erroreen azterketa linguistikoa egitea, hizkuntza-ikasleen zailtasunak kontuan hartzen dituzten laguntza-tresnak edota ikas-estrategiak garatu ahal izateko. Esaterako, bildutako errore-adibideak baliagarriak dira euskara-ikasleekin lantzeko edota adibide horiek oinarri hartuta determinatzailearen erabileraren inguruko ariketak prestatzeko.

Eskuzko etiketatzea eta azterketa linguistikoa izan dira determinatzaile-erroreen tratamendu automatikoa bideratzeko oinarria eta abiapuntua. Tesi-lan honetan eskuz etiketatutako determinatzaile-errore batzuen detekzio eta diagnosi linguistiko automatikoa bideratu dugu eta horretaz hitz egingo dugu, hain zuzen, hurrengo kapituluan.

Euskarazko errorearen eta desbideratzearen tratamendu automatikoaren xede nagusiak dira ortografia-, gramatika- eta estilo-zuzentzailea edota aldaki dialektalen markatzailea bezalako tresnak garatzea eta ordenagailuz lagundutako hizkuntzen i(r)a-kaskuntzaren alorrean erabiltzaileen beharretara egokitutako laguntza-tresnak prestatzea. Tesi-lan honetan determinatzaile-errorearen detekzio automatikoa bideratzeko lehen urratsak egin ditugu.

VII. Determinatzaile-errorearen tratamendu automatikoa

VII.1. Sarrera

Euskarazko errorearen eta desbideratzearen analisia eta prozesamendua dira tesi-lan honi dagozkion ikerlerro nagusiak. Kapitulu honetan, eskuz etiketatutako determinatzaile-errorearen detekzio automatikoa dugu mintzagai.

IXA taldean errore ortografikoen detekzio eta zuzenketa automatikoari ekin zitzaion lehenik eta behin. Horrela garatu zen XUXEN zuzentzaile ortografikoa (Agirre *et al.*, 1992), euskararen morfologiaren tratamendu automatikoan oinarrituta. Euskararako gramatika-zuzentzailea garatzea da orain dugun erronka.

Hainbat teknika edo metodo baliatzen dira, oro har, errorearen detekzio automatikorako. Horien artean, teknika enpirikoak edota sinbolikoak nagusitzen dira:

- *Teknika enpirikoak* corpusetan oinarritutako teknikak dira, ezagutza testuetatik erauzten dutenak. Talde honetakoak dira metodo estatistikoak eta *machine learning* edo ikasketa automatikoa, hau da, datu-multzo batetik ezagutza ateratzen saiatzen diren metodoak. Euskarazko erroreen eta desbideratzeen detekzio automatikoa bideratzeko, teknika enpirikoak ez ditugu asko erabili: batetik, emaitza onak lortzeko corpus erroredun handiak beharko genituzkeelako; are gehiago, kontuan izanik corpus erroredunetan, normalean, errore baino egitura zuzen gehiago egon ohi dela; eta bestetik, metodo estatistikoetan oinarritutako ikasketan sortzen den ezagutza ez delako gizakiarentzat ulerterraza.
- *Teknika sinbolikoak* hizkuntza-ezagutzan oinarritzen dira eta ezagutza hori erregeletan kodetu ohi da. Teknika sinbolikoak egokiak dira bai hizkuntzaren azterketa eta prozesamendurako eta bai erroreen detekzio automatikorako, erregela bidez deskriba baitaitezke hizkuntza-egiturak, zuzenak edota erroredunak. Erroreen tratamendu automatikoari dagokionez, teknika sinbolikoak baliatuz alarma faltsu gutxiago sortu ohi dira. Gainera, erabiltzailearekiko *feedbacka* edo atzeraelikadura lantzeko diagnosi-mezu egokiak idazteko aukera eskaintzen duten teknikak dira, eta erroreen detekzioarekin batera diagnosi linguistikoa egiteko aukera izatea ekarpen interesgarria da gure ustez.

Analisi morfosintaktikoan edota sintaktikoan oinarrituta “*erroreak detektatzeko hainbat hurbilpen daude: erlaxazioa, chart-aren erabilpena, erroreen testuinguruak deskribatzen dituzten patroiak, gramatikei errore jakin batzuei buruzko erregelak txertatzea, etab.*” (Gojenola, 2000). Egitura erroredun guztiek, ordea, ezaugarri desberdinak dituzte eta landu nahi den errore motaren arabera, teknika bat edo beste baliatzea komeni da, ez baitago errore mota guztientzako baliagarria den metodorik. IXA taldean, esaterako, euskarazko gramatika-erroreen tratamendu automatikoa bideratzeko hainbat teknika erabili dira: determinatzaileen eta postposizio-lokuzioen kategoriako erroreak Murriztapen Gramatika (MG) formalismoan oinarritutako erregela bidez detektatzen dira (Uria *et al.*, 2009; Díaz de Ilarraza *et al.*, 2008), data-egituretan egiten diren erroreak XFST (Xerox Finite State Tool) tresnarekin (Díaz de Ilarraza *et al.*, 2007), koma-erroreak ikasketa automatikoko tekniken bidez (Alegria *et al.*, 2006) eta Saroi tresna garatu da (Oronoz, 2009)

mendekotasun-zuhaitzetan informazioa kontsultatuz esaldi-mailako erroreak detektatzeko (komunztadura-erroreak, esaterako).

Euskararako gramatika-zuzentzailearen oinarriak ezarri eta hainbat errore motaren tratamendu automatikoa bideratuz, XUXENg-ren lehen bertsio bat garatu dugu. Zuzentzaileak, oraingoz, gramatika-errore hauek tratatzen ditu, besteak beste: ‘harik eta... arte’ esapide indargarriaren erabilera okerra (**ez mugitu harik eta nik esaten dizut arte*); nozioen inguruko esapide okerrak (**gosea/beroa/beldurra... dut*); zenbatasun-adierazpen okerrak (**zazpi litrotako botila da*); bihurkarien erabilera okerra (**ikusiko gara*); ‘guztia’ren erabilera okerra (**liburu guzti horiek irakurri ditut*); konparazioak egiteko ‘baina’ erabiltzea ‘baino’ren ordez (**ni zu baina azkarragoa naiz*); perpaus berean bi ezeztapen erabiltzea (**ez dut ezer ez egin*)... Tesi-lan honetan egin dugun determinatzaile-erroreen azterketa ekarpen garrantzitsua da euskararako garatzen ari garen gramatika-zuzentzailea osatze bidean.

Kapitulu honetan determinatzaile-erroreen tratamendu automatikoz hitz egingo dugu. Lehenik eta behin, IXA taldean garatu den euskararako analizatzaile sintaktikoa aurkeztuko dugu (VII.2. atala), analisi-katearen urratsak laburbilduz (VII.2.1. atala); izan ere, analizatzaile sintaktikoaren analisisa da determinatzaile-erroreen tratamendu automatikoaren abiapuntua. Determinatzaile-erroreak detektatzeko erregelak Murritzapen Gramatika (MG) formalismoa baliatuz idatzi ditugu eta formalismoaren nondik norakoez hitz egingo dugu VII.3. atalean. Jarraian, VII.4. puntuan, determinatzaile-erroreak automatikoki detektatzeko sortu dugun erroreen gramatikaz mintzatuko gara, eta VII.5. atalean aipatuko ditugu erroreen detekzio automatikoan aurkitu ditugun hainbat arazo. Erregelen ebaluazioari dagokio VII.6. puntua: euskara-ikasleen testuetan eskuratutako emaitzak (VII.6.1.) eta testu zuzenetan lortutakoak (VII.6.2.) aurkeztuko ditugu. MG bidez tratatzen ez diren determinatzaile-errore mota batzuen tratamenduaz hitz egingo dugu ondoren, VII.7. atalean; hitzak deklinatzean determinatzaile atzizkia ongi ez erabiltzeagatik sortutako erroreen tratamenduaz, hain zuzen. Amaitzeko, kapituluaren laburbilpena egingo dugu VII.8. atalean.

VII.2. Euskararako analizatzaile sintaktikoa

Euskararen azterketa eta prozesamendurako IXA taldean garatu den analizatzaile sintaktikoak edo *parse*rrak ematen digun analisia hainbat azterketa linguistikoren abiapuntua da. Tesi-lan honetan aztergai ditugun determinatzaile-erroreen detekzio automatikoa egiteko erregelak, esaterako, analisi-kateak ematen digun irteeran oinarrituta idatzi ditugu. Horregatik, analizatzaile sintaktikoa azalduko dugu lehenik eta behin, erroreen tratamendu automatikoaren abiapuntua zein den ezagutzeko.

Analizatzailea hainbat moduluk osatzen dute. Modulu bakoitza aurreko urratsean definitutako informazioan oinarritzen da eta, era berean, hurrengo moduluetakoa informazioa aberastuz doa. Analisi morfosintaktikoa egiten da lehenik: “*morfemetatik abiatuta hitz osoaren analisia lortzeko mekanismoen deskribapena eta inplementazioa*” (Gojenola, 2000). Izan ere, lema eta morfema bakoitzari dagokion informazio morfosintaktikoaren arabera osatuko dira hitzak. Analisi morfosintaktikoan oinarrituta egiten da ondoren analisi sintaktikoa, hitzen arteko erlazioak zehaztuz. Beraz, analisi-katea osatzen duten modulu batzuk analisi morfosintaktikoari dagozkio eta beste batzuk analisi sintaktikoari. *Parse*rraren nolakotasunari dagokionez, Gojenolak (2000), Aranzabek (2008) eta Oronozek (2009) ezaugarri nagusi hauek azpimarratzen dituzte:

- Analizatzaile sendoa da, testu errealeko elementu guztiei gutxienez analisi zuzen bat emateko gai dena.
- Hizkuntzaren ezagutzan oinarritzen da eta egoera finituko mekanismoen bidez definitu da batez ere, bi teknika hauek erabilia: Murritzapen Gramatika (MG) (Karlsson *et al.*, 1995) eta XFST tresna (Aït-Mokhtar eta Chanod, 1997; Karttunen *et al.*, 1997). Baina geruza edo fase batzuetan teknika estatistikoak ere aplikatzen dira (desanbiguazio morfosintaktikorako, adibidez).
- Hurbilpen murriztailea erabiltzen du: elementu bakoitzaren interpretazio posible guztiak ematen ditu lehenik eta ondoren, testuinguruaren arabera, batzuk baztertzen ditu, horretarako desanbiguazio-teknikak baliatuz.
- Informazioa mendekotasun-egiturak erabiliz errepresentatzen du, ez osagai-egiturak erabilia.
- Analisi sintaktiko partziala egiten du, analisi gramatikal ez osoa: “*analisi partzialak fidagarritasuna eta sendotasuna ditu helburu, sakontasuna eta osotasuna neurri batean galduaz*” (Oronoz, 2009).

VII.2.1. Analisi-katea: urratsak

Euskararako analizatzaileak ondoko urrats hauek egiten ditu: analizatzaile morfosintaktikoaren (MORFEUS) barruan, tokenizazioa, segmentazioa, morfosintaxia eta Hitz Anitzeko Unitate Lexikalen (HAUL) tratamendua egiten da lehenik; ondoren, analisisien desanbigrazio morfosintaktikoa (EUSTAGGER); jarraian, zatitzailea (IXATI) aplikatzen da: sintaktikoki erlazionaturik dauden hitz-segidak (entitate izendunak, postposizio-lokuzioak, sintagmak eta aditz-kateak) identifikatzen dira eta funtzio sintaktikoak desanbiguatzen dira; bukatzeko, mendekotasun sintaktikoak markatzen dira.

Urrats batetik bestera egiten diren aldaketak ikusteko eta determinatzaile-erroreen detekzio automatikoa bideratzeko idatzi ditugun erregelen abiapuntua zein den ulertzeko, ondoko esaldi erroreduna erabiliko dugu:

***Hitzaldi** eman eta gero Miren Azkarate Kultura Sailburuak komentatu du **auzia hau** guztion artean konpondu behar dugun **kontua bat** dela.

Esaldian hiru sintagma erroredun daude: **hitzaldi** (determinatzaile gabeko sintagma), **auzia hau** (determinatzaile atzizki singularra eta erakuslea dituen sintagma) eta **kontua bat** (determinatzaile atzizki singularra eta zenbatzaile mugagabea dituen sintagma). Esaldi hori adibidetzat hartuta azalduko ditugu, jarraian, testuen analisi-prozesuan egiten diren urratsak:

MORFEUS: Analizatzaile morfosintaktikoa. Analizatzaile morfosintaktikoaren lehen urratsa analisi morfosintaktikoari dagokion moduluak osatzen du eta MORFEUS analizatzaile morfologikoaren bitartez egiten da (Alegria, 1995; Aduriz *et al.*, 1998; Ezeiza, 2002). Modulu hau lau geruza nagusik osatzen dute:

- Tokenizazioa: jatorrizko testua *token*etan banatzen da, *item*etan. Token horiek deklinatu gabeko edota deklinatutako hitzak, zenbakiak, laburdurak, puntuazio-markak... izango dira, analisi morfologikoan sarrera gisa erabiliko diren unitateak, hain zuzen. Banaketa hori *tokenizatzailearen* (*token* edo *item* ezagutzailearen) bitartez egiten da.

- Segmentazioa (segmentazio morfologikoa): hitz-forma bakoitza lema eta morfemetan banatu eta bakoitzari dagozkion interpretazio morfologiko posible guztiak esleitzen zaizkio: lema, kategoria, azpikategoria, deklinabide-atzizkia, numeroa edota mugatasun-kontuak, funtzio sintaktikoak (baldin badagozkio) eta, aditzen kasuan, modua, denbora eta aspektua. Segmentazioa edo analisi morfologikoa EDBLko informazioa oinarri hartuta egiten da eta helburua segmentatzaile morfologikotik analizatzaile morfologikoa eta sintaktikoa bideratzea da (Aldezabal *et al.*, 2007a).
- Morfosintaxia: segmentazioan eskuratzen den informazioa elaboratzen da, datuak bilduz eta txukunduz, ondorengo azterketa linguistikoetarako baliagarria izango den informazioa aukeratuz eta goratuz (Arriola *et al.*, 2005), testuingururik gabeko gramatika baten bidez (Aduriz, 2000; Gojenola, 2000). Analisi morfologikoaren emaitza ezinbestekoa da informazio morfosintaktikoaren irteera ahalbidetzeko.
- Hitz Anitzeko Unitate Lexikalen (HAUL) tratamendua: unitate batek baino gehiagok osatzen dituen hitz konbinatuak (esate baterako, *hain zuzen ere* edota *behar izan*) elkarrekin lematizatu eta analizatu behar dira, hitz-konbinazioko osagai bakoitzaren funtzioa eta konbinazio osoarena ez baitira beti berdinak. HAULak identifikatu eta dagokien informazio morfosintaktikoa esleitzeko, HABIL tresna sortu zen (Ezeiza, 2002).

Aipatu lau geruzek osatzen duten MORFEUS analizatzaile morfosintaktikoaren irteera, desanbiguatu aurrekoa, erakusten digu 21. irudiak, adibide gisa hartu dugun esaldian³⁹. Hitz bakoitzaren interpretazio posible guztiak ikusten ditugu (hitz bakoitzari dagokion informazio morfologikoa) lehen urrats honetan: *hau* hitzak, esaterako, erakuslearen, aditz laguntzailearen eta aditz trinkoaren analisiak hartzen ditu eta *bat*, berriz, determinatzaile, izen eta aditz bezala analizatu da. Urrats honetan, bestalde, HAUL ziurrak tratatuta datoz jada (*behar_izan*, esaterako, unitate gisa analizatuta dator).

³⁹ Letra lodiz nabarmendu ditugu determinatzaile-erroreei dagozkien analisiak.

```

"<Hitzaldi>"<HAS_MAI>"
  "hitzaldi" IZE ARR ZERO HAS_MAI @KM>
  "hitzaldi" IZE ARR ABS MG HAS_MAI @OBJ @PRED @SUBJ
"<eman>"
  "eman" ADT PNT MDNC NOR_NORK NR_HURA NK_HIK-NO @+JADNAG
  "eman" ADI SIN ADOIN NOTDEK @-JADNAG
  "eman" ADI SIN PART ABS MG @-JADNAG_MP_OBJ @-JADNAG_MP_PRED @-JADNAG_MP_SUBJ
  "eman" IZE ARR ZERO @KM>
  "eman" IZE ARR ABS MG @OBJ @PRED @SUBJ
"<eta>"
  "eta" LOT JNT EMEN AORG @PJ
  "eta" LOT MEN KAUS AM AORG @+JADLAG_MP @+JADNAG_MP
"<gero>"
  "gero" ADB ARR ZERO @ADLG
  "gero" IZE ARR ZERO @KM>
  "gero" IZE ARR ABS MG @OBJ @PRED @SUBJ
"<Miren>"<HAS_MAI>"
  "mira" IZE ARR GEN NUMP MUGM ZERO AORG HAS_MAI @<IZLG @IZLG>
  "mira" IZE ARR ABS MG AORG HAS_MAI @OBJ @PRED @SUBJ
  "Miren" IZE IZB ZERO HAS_MAI @KM>
  "Miren" IZE IZB ABS NUMS MUGM HAS_MAI @OBJ @PRED @SUBJ
  "mi" IZE ARR ABS MG HAS_MAI @OBJ @PRED @SUBJ
"<Azkarate>"<HAS_MAI>"
  "Azkarate" IZE IZB ZERO HAS_MAI @KM>
  "Azkarate" IZE IZB ABS NUMS MUGM HAS_MAI @OBJ @PRED @SUBJ
  "Azkarate" IZE LIB ZERO HAS_MAI @KM>
  "Azkarate" IZE LIB ABS NUMS MUGM HAS_MAI @OBJ @PRED @SUBJ
"<Kultura>"<HAS_MAI>"
  "kulturatu" ADI SIN ADOIN NOTDEK HAS_MAI @-JADNAG
  "kultura" IZE ARR ZERO AORG HAS_MAI @KM>
  "kultura" IZE ARR ABS MG AORG HAS_MAI @OBJ @PRED @SUBJ
  "kultura" IZE ARR ABS NUMS MUGM AORG HAS_MAI @OBJ @PRED @SUBJ
"<Sailburuak>"<HAS_MAI>"
  "sailburu" IZE ARR ABS NUMP MUGM HAS_MAI @OBJ @PRED @SUBJ
  "sailburu" IZE ARR ERG NUMS MUGM HAS_MAI @SUBJ
"<komentatu>"
  "komentatu" ADI SIN PART NOTDEK @-JADNAG
  "komentatu" ADI SIN PART BURU NOTDEK @-JADNAG
"<du>"
  "*edun" ADL A1 NOR_NORK NR_HURA NK_HARK @+JADLAG
  "ukan" ADT PNT A1 NOR_NORK NR_HURA NK_HARK @+JADNAG
"<auzia>"
  "auzi" IZE ARR ABS NUMS MUGM @OBJ @PRED @SUBJ
"<hau>"
  "*edun" ADL A1 NOR_NORK NR_HI NK_HARK @+JADLAG
  "ukan" ADT PNT A1 NOR_NORK NR_HI NK_HARK @+JADNAG
  "hau" DET ERKARR ABS NUMS MUGM @OBJ @PRED @SUBJ
"<guztion>"
  "guzti" DET ORO GEN PH MUGM @<IZLG @IZLG>
  "guzti" DET ORO ABS MG @OBJ @PRED @SUBJ
"<artean>"
  "artean" ADB ARR ZERO @ADLG
  "arte" IZE ARR INE NUMS MUGM @ADLG
"<konpondu>"
  "konpondu" ADI SIN PART NOTDEK @-JADNAG
  "konpondu" ADI SIN PART BURU NOTDEK @-JADNAG
  "konpondu" ADI SIN PART ABS MG @-JADNAG_MP_OBJ @-JADNAG_MP_PRED @-JADNAG_MP_SUBJ
"<behar_dugun>"
  "behar_izan" ADI ADK PNT ERLT A1 NR_HURA NK_GUK NOTDEK @+JADNAG_MP_IZLG>
  "behar_izan" ADI ADK PNT ZHG A1 NR_HURA NK_GUK NOTDEK @+JADNAG_MP_OBJ
  "behar_izan" ADI ADK PNT MOS A1 NOR_NORK NR_HURA NK_GUK NOTDEK @+JADNAG_MP_ADLG
"<kontua>"
  "kontu" IZE ARR ABS NUMS MUGM @OBJ @PRED @SUBJ
"<bat>"
  "bat" DET DZH NMGS ZERO @ID>
  "bat" DET DZH NMGS ABS MG @OBJ @PRED @SUBJ
  "bat" IZE ZKI ZERO @KM>
  "bat" IZE ZKI ABS MG @OBJ @PRED @SUBJ
  "batu" ADI SIN ADOIN NOTDEK @-JADNAG
"<dela>"
  "izan" ADL KONPL A1 NOR NR_HURA @+JADLAG_MP_SUBJ @+JADLAG_MP_OBJ
  "izan" ADT PNT KONPL A1 NOR NR_HURA @+JADNAG_MP_SUBJ @+JADNAG_MP_OBJ
"<$.>"<PUNT_PUNT>"

```

• **EUSTAGGER: Desanbiguazio morfosintaktikoa.** EUSTAGGER lematizatzaileak/etiketatzaileak (Aduriz *et al.*, 2001; Alegria *et al.*, 2003) MORFEUS analizatzaile morfosintaktikoak emandako interpretazioak edota etiketa morfosintaktikoak desanbiguatzea du helburu, hau da, testuinguruan egokiak ez diren interpretazioak ezabatzea. EUSTAGGERek lau desanbiguazio-maila ditu eta bi teknika erabiltzen dira desanbiguazio morfosintaktikoa egiteko: lehenengo hiru mailetan MG formalismoan (ezagutza linguistikoan) oinarritutako erregelak eta corpusetan oinarritzen diren teknika estatistikoak edo enpirikoak baliatzen dira kategoria, azpikategoria edota kasua desanbiguatzeko (Ezeiza *et al.*, 1998); laugarren mailan, berriz, MG bakarrik erabiltzen da eta, beraz, gutxiago desanbiguaten da. Gehien desanbiguaten duen maila hirugarrena da (25. taula):

Metodoa	Maila	Desanbiguaten diren ezaugarriak
MG + estatistika	1. maila	Kategoria
	2. maila	Kategoria, azpikategoria
	3. maila	Kategoria, azpikategoria, kasua
MG	4. maila	Ezaugarri morfologikoak

25. taula. Desanbiguazio-mailak.

Gramatika-erroreen tratamendu automatikorako desanbiguazio-maila bat ala beste oinarri hartuta, emaitza desberdinak lortzen dira. Oronozek (2009), esaterako, ikusi du postposizio-lokuzio erroredun batzuetan eta data-egituretan egindako erroreen kasuan, desanbiguatu gabeko analisiak baliatuz lortzen direla emaitzarik onenak; eta komunztadura-erroreen kasuan, berriz, gehien desanbiguatutako analisiak erabilia emaitzak hobeak direla. Euskararako XUXENg gramatika-zuzentzailea garatzeko, hirugarren mailan desanbiguatutako analisisetan oinarritutako erregelak idazteko erabakia hartu zen, gramatika-erroreen detekzioa gehien desanbiguatutako analisiak oinarri hartuta bideratzeko erabakia. Irizpide bera jarraitu dugu guk determinatzaile-erroreak detektatzeko erregelak egiteko.

Aztergai dugun esaldi erroredunari EUSTAGGER pasa ondoren, aurreko moduluak emandako informazio morfosintaktikoa desanbiguatu eta hainbat datu edo interpretazio desagertu egin dira: *hau* determinatzaileak, adibidez, hiru analisi zituen lehenengo urratsean (**edun* aditz-laguntzailearena, *ukan* aditz trinkoarena eta *hau* erakuslearena) eta determinatzaileari dagokion analisisia bakarrik geratzen da

desanbiguazioaren ostean. Gauza bera gertatzen da gainerako elementuekin ere (22. irudia).

```
"<HitzaIdi>"<HAS_MAI>"
  "hitzaIdi" IZE ARR ABS MG HAS_MAI @OBJ @PRED @SUBJ
"<eman>"
  "eman" ADI SIN PART NOTDEK @-JADNAG
"<eta>"
  "eta" LOT JNT EMEN AORG @PJ
"<gero>"
  "gero" ADB ARR ZERO @ADLG
"<Miren>"<HAS_MAI>"
  "Miren" IZE IZB ZERO HAS_MAI @KM>
"<Azkarate>"<HAS_MAI>"
  "Azkarate" IZE IZB ABS NUMS MUGM HAS_MAI @OBJ @PRED @SUBJ
"<Kultura>"<HAS_MAI>"
  "kultura" IZE ARR ZERO AORG HAS_MAI @KM>
"<Sailburuak>"<HAS_MAI>"
  "sailburu" IZE ARR ERG NUMS MUGM HAS_MAI @SUBJ
"<komentatu>"
  "komentatu" ADI SIN PART BURU NOTDEK @-JADNAG
"<du>"
  "*edun" ADL A1 NOR_NORK NR_HURA NK_HARK @+JADLAG
"<auzia>"
  "auzi" IZE ARR ABS NUMS MUGM @OBJ @PRED @SUBJ
"<hau>"
  "hau" DET ERKARR ABS NUMS MUGM @OBJ @PRED @SUBJ
"<guztion>"
  "guzti" DET ORO GEN PH MUGM @<IZLG
  "guzti" DET ORO GEN PH MUGM @IZLG>
"<artean>"
  "arte" IZE ARR INE NUMS MUGM @ADLG
"<konpondu>"
  "konpondu" ADI SIN PART NOTDEK @-JADNAG
"<behar_dugun>"
  "behar_izan" ADI ADK PNT ERLT A1 NR_HURA NK_GUK NOTDEK @+JADNAG_MP_IZLG>
  "behar_izan" ADI ADK PNT ZHG A1 NR_HURA NK_GUK NOTDEK @+JADNAG_MP_OBJ
  "behar_izan" ADI ADK PNT MOS A1 NOR_NORK NR_HURA NK_GUK NOTDEK @+JADNAG_MP_ADLG
"<kontua>"
  "kontu" IZE ARR ABS NUMS MUGM @OBJ @PRED @SUBJ
"<bat>"
  "bat" DET DZH NMGS ABS MG @OBJ @PRED @SUBJ
"<delat>"
  "izan" ADT PNT KONPL A1 NOR NR_HURA @+JADNAG_MP_SUBJ
  "izan" ADT PNT KONPL A1 NOR NR_HURA @+JADNAG_MP_OBJ
  "izan" ADT PNT KONPL A1 NOR NR_HURA @+JADNAG_MP_PRED
  "izan" ADT PNT MOD/DENB A1 NOR NR_HURA @+JADNAG_MP_ADLG
"<$.>"<PUNT_PUNT>"
  PUNT_PUNT
```

22. irudia. EUSTAGGER desanbiguatzaile morfosintaktikoaren irteera (3. mailan desanbiguatuta).

IXATI: Zatitzailea edo chunker-a. IXATI zatitzailearen bidez (Aduriz *et al.*, 2006b), desanbiguatutako analisien gainean entitate izendunak, postposizio-lokuzioak eta kate sintaktikoak (sintagmak eta aditz-kateak) zehazten dira, hau da, sintaktikoki erlazionaturik dauden hitz-segidak identifikatzen dira. IXATIk, gainera, funtzio sintaktikoen desanbiguazioa ere egiten du.

Entitate izendunak mugatzeko eta sailkatzeko EIHERA tresna (Alegria *et al.*, 2003) erabiltzen da; entitatearen hasiera/bukaera (HAS/BUK) eta mota (pertsona,

tokia edota erakundea) zehazten dira horrela. Aztergai dugun adibidean, esaterako, *Miren Azkarate* pertsona gisa analizatzen da (ENTI_HAS_PER eta ENTI_BUK_PER etiketekin) eta *Kultura Sailburua* erakunde gisa (ENTI_HAS_ORG eta ENTI_BUK_ORG etiketen bidez). MGn oinarritutako hainbat gramatika sortu dira, bestalde, postposizio-lokuzioak eta kate sintaktikoak zehazteko edota funtzio sintaktikoak desanbiguatzeke. Gure adibidean ikusten dugunez (23. irudia), *guztion artean* postposizio-lokuzioari, esaterako, POS-HAS eta POS-BUK etiketak esleitu zaizkio; eta zuzenak diren sintagmen hasierak eta bukaerak ere mugatu dira %SIH eta %SIB etiketekin, edota %SINT etiketarekin, hitz bakarreko sintagmetan.

```
"<Hitzaldi>"<HAS_MAI>"
    "hitzaldi" IZE ARR ABS MG HAS_MAI @OBJ @PRED %SINT
"<eman>"
    "eman" ADI SIN PART NOTDEK @-JADNAG %ADIKAT
"<eta>"
    "eta" LOT JNT EMEN AORG AORG @PJ
"<gero>"
    "gero" ADB ARR ZERO @ADLG %SINT
"<Miren>"<HAS_MAI>"
    "Miren" IZE IZB ENTI_HAS_PER HAS_MAI @KM> %SIH
"<Azkarate>"<HAS_MAI>"
    "Azkarate" IZE IZB ABS NUMS MUGM ENTI_BUK_PER HAS_MAI @OBJ @PRED %SIB
"<Kultura>"<HAS_MAI>"
    "kultura" IZE ARR ENTI_HAS_ORG AORG HAS_MAI AORG @KM> %SIH
"<Sailburuak>"<HAS_MAI>"
    "sailburu" IZE ARR ERG NUMS MUGM ENTI_BUK_ORG HAS_MAI @SUBJ %SIB
"<komentatu>"
    "komentatu" ADI SIN PART BURU NOTDEK @-JADNAG %ADIKATHAS
"<du>"
    "*edun" ADL A1 NOR_NORK NOR_NORK NR_HURA NK_HARK @+JADLAG %ADIKATBU
"<auzia>"
    "auzi" IZE ARR ABS NUMS MUGM @OBJ @PRED %SINT
"<hau>"
    "hau" DET ERKARR ABS NUMS MUGM @OBJ @PRED
"<guztion>"
    "guzti" DET ORO GEN PH MUGM @IZLG> {POS-HAS12 %SIH
"<artean>"
    "arte" IZE ARR INE NUMS MUGM @ADLG {POS-BUK12 %SIB
"<konpondu>"
    "konpondu" ADI SIN PART NOTDEK @-JADNAG %ADIKAT
"<behar_dugun>"
    "behar_izan" ADI ADK PNT ERLT A1 NR_HURA NK_GUK NOTDEK @+JADNAG_MP_IZLG> %ADIKAT
    "behar_izan" ADI ADK PNT MOS A1 NOR_NORK NOR_NORK NR_HURA NK_GUK NOTDEK
        @+JADNAG_MP_ADLG %ADIKAT
"<kontua>"
    "kontu" IZE ARR ABS NUMS MUGM @SUBJ %SINT
"<bat>"
    "bat" DET DZH NMGS ABS MG @SUBJ
"<de la>"
    "izan" ADT PNT KONPL A1 NOR NOR NR_HURA @+JADNAG_MP_SUBJ %ADIKAT
    "izan" ADT PNT KONPL A1 NOR NOR NR_HURA @+JADNAG_MP_OBJ %ADIKAT
    "izan" ADT PNT KONPL A1 NOR NOR NR_HURA @+JADNAG_MP_PRED %ADIKAT
"<$.>"<PUNT_PUNT>"
    PUNT_PUNT
```

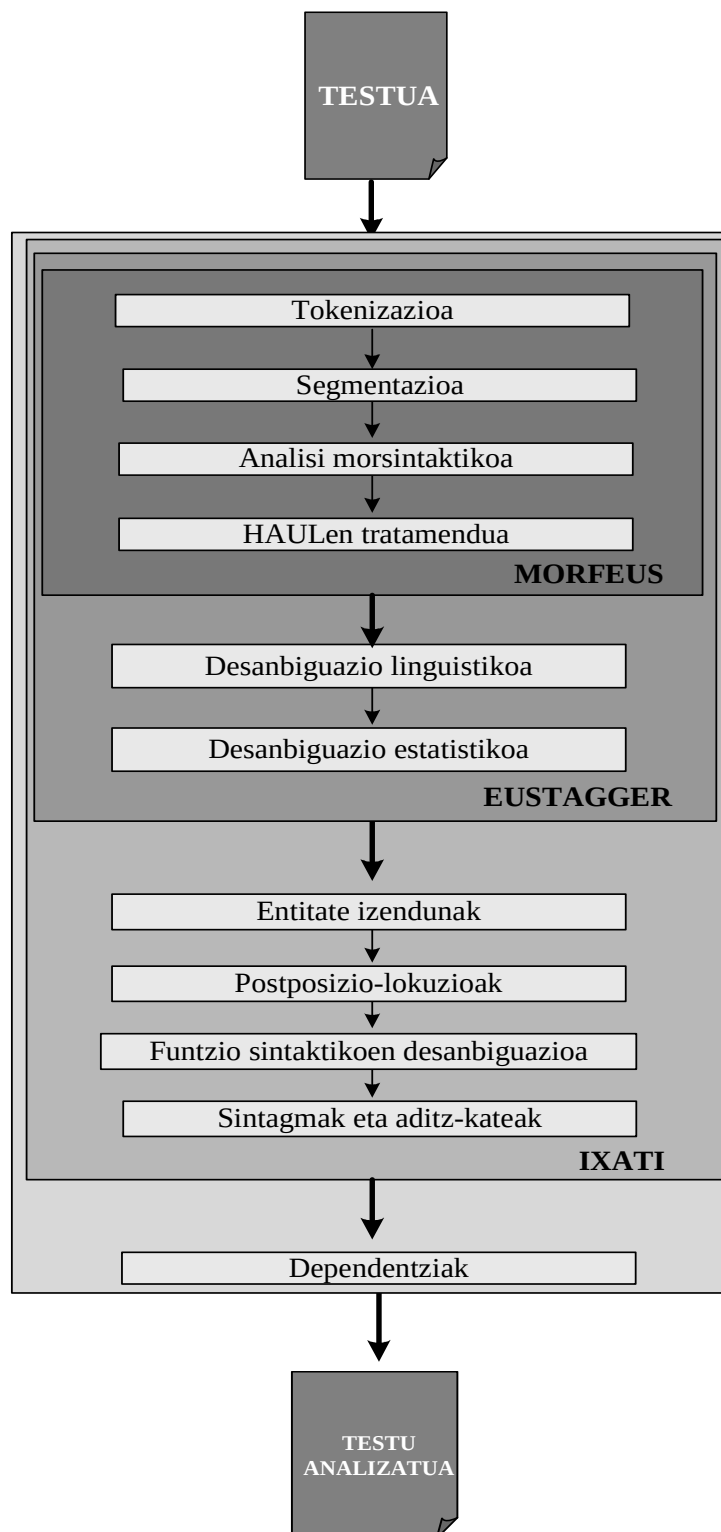
23. irudia. IXATI zatitzailearen analisiak.

• **Mendekotasun sintaktikoak: Mendekotasun-erlazioen etiketatzailerak.**

Azkenik, IXATIk ematen digun irteeran oinarrituta esleitzen dira mendekotasun sintaktikoak. Euskararako Dependentsia Gramatika Konputazionalaren (EDGK) (Aranzabe, 2008) bitartez ezartzen dira, & ikurrarekin, esaldia osatzen duten hitzen arteko mendekotasun-erlazioak edota dependentziak (24. irudia), eta horrela eskuratzen ditugu testuen analisi sintaktiko partzialak. Analisi-kateko azken modulu hau urrats garrantzitsua da, besteak beste, erroreen tratamendu automatikoaren alorrean komunztadura-erroreak detektatzeko garatu den *Saroi* tresnarentzako (Oronoz, 2009).

```
"<Hitzaldi>"<HAS_MAI>"
    "hitzaldi" IZE ARR BIZ- ABS MG HAS_MAI @OBJ %SINT &NCOBJ>
    "hitzaldi" IZE ARR BIZ- ABS MG HAS_MAI @PRED %SINT
"<eman>"
    "eman" ADI SIN PART NOTDEK @-JADNAG %ADIKAT &ADITZ_EMEN>
"<eta>"
    "eta" LOT JNT EMEN AORG @PJ
"<gero>"
    "gero" ADB ARR ZERO w4,L-A-ADB-ARR-3,lsfi6 @ADLG %SINT &NCMOD>
"<Miren>"<HAS_MAI>"
    "Miren" IZE IZB PLU- ZERO HAS_MAI @KM> %SIH &NCMOD>
"<Azkarate>"<HAS_MAI>"
    "Azkarate" IZE IZB PLU- ABS NUMS MUGM HAS_MAI @OBJ %SIB &NCOBJ>
    "Azkarate" IZE IZB PLU- ABS NUMS MUGM HAS_MAI @PRED %SIB
"<Kultura>"<HAS_MAI>"
    "kultura" IZE ARR BIZ- ZERO AORG HAS_MAI @KM> %SIH
"<Sailburuak>"<HAS_MAI>"
    "sailburu" IZE ARR ERG NUMS MUGM HAS_MAI @SUBJ %SIB
"<komentatu>"
    "komentatu" ADI SIN PART BURU NOTDEK @-JADNAG %ADIKATHAS
"<du>"
    "*edun" ADL A1 NOR_NORK NR_HURA NK_HARK @+JADLAG %ADIKATBU &<AUXMOD>
"<auzia>"
    "auzi" IZE ARR BIZ- ABS NUMS MUGM @OBJ %SINT &NCOBJ>
    "auzi" IZE ARR BIZ- ABS NUMS MUGM @PRED %SINT
"<hau>"
    "hau" DET ERKARR ABS NUMS MUGM @OBJ @PRED
"<guztion>"
    "guzti" DET ORO GEN PH MUGM @IZLG> %SIH &NCMOD>
"<artean>"
    "arte" IZE ARR BIZ- INE NUMS MUGM @ADLG %SIB &NCMOD>
"<konpondu>"
    "konpondu" ADI SIN PART NOTDEK @-JADNAG %ADIKAT
"<behar_dugun>"
    "behar_izan" ADI ADK PNT ERLT A1 NR_HURA NK_GUK NOTDEK @+JADNAG_MP_IZLG> %ADIKAT
    "behar_izan" ADI ADK PNT ZHG A1 NR_HURA NK_GUK NOTDEK @+JADNAG_MP_OBJ %ADIKAT
    "behar_izan" ADI ADK PNT MOS A1 NOR_NORK NR_HURA NK_GUK NOTDEK @+JADNAG_MP_ADLG %ADIKAT
"<kontua>"
    "kontu" IZE ARR BIZ- ABS NUMS MUGM @SUBJ %SIB
"<bat>"
    "bat" DET DZH NMGS ABS MG @SUBJ
"<de la>"
    "izan" ADT PNT KONPL A1 NOR NR_HURA @+JADNAG_MP_SUBJ %ADIKAT &<CCOMP>
    "izan" ADT PNT KONPL A1 NOR NR_HURA @+JADNAG_MP_OBJ %ADIKAT &<CCOMP>
    "izan" ADT PNT KONPL A1 NOR NR_HURA @+JADNAG_MP_PRED %ADIKAT
    "izan" ADT PNT MOD/DENB A1 NOR NR_HURA @+JADNAG_MP_ADLG %ADIKAT &<CMOD>
"<$.>"<PUNT_PUNT>"
    PUNT_PUNT
```

25. irudiak erakusten dizkigu euskararako analisi-katearen modulu eta geruza nagusiak:



25. irudia. Analisi-katearen urratsak.

Determinatzaile-erroreak detektatzeko erregelak IXATI zatitzailearen analisietan oinarrituta idatzi ditugu, hitzen arteko mendekotasun-erlazioak edo dependentziak kontuan hartu gabe; eta zehazki, hirugarren mailan desanbiguatutako analisiak hartu ditugu oinarri.

VII.3. Baliatutako formalismoa: Murriztapen Gramatika

Murriztapen Gramatika (MG) (Karlsson *et al.*, 1995) egoera finituko mekanismoetan oinarritutako formalismoa da. Erregelen bidez, analizatutako testuei informazio linguistikoa gehitzeko edota desanbiguatzeko baliatu ohi da. IXA taldean, esaterako, MGn oinarritutako erregelen bidez markatzen dira postposizio-lokuzioen eta kate sintaktikoen (sintagmen eta aditz-kateen) hasierak eta bukaerak edota mendekotasun sintaktikoak; MG bidez egiten da, halaber, desanbiguazio morfosintaktikoa eta funtzio sintaktikoen desanbiguazioa; eta errore batzuen detekziorako ere baliatzen da MG. Determinatzaile-erroreak detektatzeko erregelak, esaterako, MG formalismoa erabilia idatzi ditugu.

Determinatzaile-erroreak errore lokalak dira, hau da, komunikazioa eteten ez duten sintagma-mailako erroreak; eta MGk aukera ematen digu, ezagutza linguistikoan oinarritutako erregelen bidez, sintagma-mailako hainbat errore detektatzeko. Erroreak detektatzeaz gain, erroreen diagnosi linguistikoa ere egin daiteke MGrekin. Horretarako, erroreen detekziorako idazten diren erregelak euskarazko erroreen eta desbideratzeen sailkapeneko kategoriei edota azpikategoriei lotuta daude. Detekzioaz gain erroreen diagnosi linguistikoa egitea ekarpen interesgarria da guretzat, aztergai ditugun bi ikerketa-alorretan.

Oro har, sintagma-mailako erroreen detekzio automatikoa bideratzeko hainbat hizkuntzatan erabili izan den formalismoa da MG (26. taula):

- Suedierarako gramatika-zuzentzailea (Grammatifix) garatzeko (Arppe, 2000; Birn, 2000) hainbat teknika baliatu dituzte. Errore sintaktikoen detekziorako, adibidez, MG formalismoa erabili dute. 26 errore mota landu dituzte; batez ere, sintagma barruko komunztadura eza eta aditz-kateetan gertatzen den koherentzia falta motako erroreak. Guztira, 650 erregela inguru idatzi dituzte errore sintaktikoen detekziorako eta lortutako emaitzak onak dira: % 70eko

doitasuna eta % 85eko estaldura⁴⁰. Errore gutxi eta sinpleak detektatuz emaitza onak lortzearen aldeko apustua egin dute ikertzaile horiek, gramatika-zuzentzailearen erabiltzaileei begira.

- Johannessen *et al.*-ek (2002) garatu dute Norvegierarako gramatika-zuzentzailea (Grammar Checker for Norwegian, NGC), suedierarako zuzentzailea oinarri hartuta. Analizatzaile morfologiko batek, MG bidezko desanbiguatzaile morfologiko batek eta erroreak detektatzeko modulu batek osatzen dute zuzentzailea. 700 erregela egin dira sintagma barruko komunztadura eza, ezezeko egitura okerrak edota hitzen ordena okerra motako erroreak detektatzeko, besteak beste. Erregelen doitasuna % 75ekoa da, eta ez daukagu estalduraren inguruko daturik.
- Katalanerako ortografia- eta gramatika-zuzentzailearen (Badia *et al.*, 2004) oinarrian dagoen analizatzaile morfosintaktikoa MG formalismo bidez inplementatutako hainbat moduluk eta erroreen detekzioarako propio garatu diren beste zenbait moduluk osatzen dute. Sintagma barruko komunztadura eza, aditzaren eta subjektuaren arteko komunztadura eza, hitzen ordena okerra edota aditzen azpikategoriazioa motako erroreen detekzio automatikoa bideratu dute, besteak beste. Ez dute, ordea, datu zehatzik ematen, ez baitute ebaluazio sendorik egin.

Gramatika-zuzentzaileak	Landutako fenomenoak	Emaitzak
Suediera (Arppe, 2000)(Birn, 2000)	• Sintagma barruko komunztadura eza • Koherentzia falta aditz katean • ...	Doitasuna: % 70 Estaldura: % 85
Norvegiera (Johannessen <i>et al.</i> , 2002)	• Sintagma barruko komunztadura eza • Hitzen ordena	Doitasuna: % 70 Estaldura: - Alarma Faltsuak: % 30
Katalana (Badia <i>et al.</i> , 2004)	• Sintagma barruko komunztadura eza • Subj-Adi arteko komunztadura eza • Hitzen ordena • Aditzen azpikategoriazioa	-

26. taula. MG formalismoa erabili duten zenbait gramatika-zuzentzaile.

⁴⁰ Doitasunak adierazten du automatikoki detektatu diren errore guztien artean zenbat errore detektatu diren ongi eta estaldurak, berriz, detektatu beharko liratekeen erroreetatik zenbat errore detektatu diren.

Zaila da MG baliatzen duten azterketetan lortutako emaitzak konparatzea; izan ere, hizkuntzak desberdinak dira eta landutako errore motak ere bai. Baina, oro har, formalismo horren bidez nahiko emaitza onak lortu ohi dira. Hala ere, ez da posible errore lokal guztiak MGrekin detektatzea; eta esaldi-mailako erroreak, errore semantikoak edota errore pragmatikoak ere ezin dira formalismo horrekin detektatu.

VII.4. Determinatzaile-erroreen gramatika

Tesi-lan honetan eskuz etiketatutako determinatzaile-errore batzuk automatikoki detektatzeko erregela multzo bat sortu dugu. Esan bezala, erregelak MG formalismoa baliatuz idatzi ditugu, IXA taldean garatu den euskararako analizatzaile sintaktikoaren irteera, hirugarren mailan desanbiguatutakoa, oinarri hartuta.

Analisi-katea azaltzean (VII.2.1. atala) erabili dugun esaldian (****Hitzaldi** eman eta gero Miren Azkarate Kultura Sailburuak komentatu du **auzia hau** guztion artean konpondu behar dugun **kontua bat** dela*) hiru determinatzaile-errore daude: **hitzaldi*, **auzia hau* eta **kontua bat*. Lehenengo adibideari (****Hitzaldi** eman eta gero*), esaterako, &OKER_DETK1_0 kodedun erregela dagokio. Erregela honela irakurriko genuke⁴¹:

⁴¹ Erregelak nola egiten ditugun erakusteko adibideak baino ez dira hauek, hurbilpen bat. Benetako erregelak osatuagoak dira, osagai bakoitzak eta inguruko elementuek bete beharreko baldintza gehiago zehaztu behar izaten baitira normalean. Erregela guztiak, osorik, E eranskinean daude ikusgai.

Esleitu (ADD) &OKER_DETK1_0 errore kodea izenari (TARGET IZE) baldin eta (IF):

- 0. posizioan (0) dagoen izena (IZE) arrunta (ARR), absolutiboa (ABS) eta mugagabea (MG) bada;
- Izena A itsatsiduna ez bada (NOT 0 AORG);
- Izenaren eskuinetara 1. posizioan (1) aditza (ADI OR ADT) badago;

Erregela:

```
ADD (&OKER_DETK1_1) TARGET IZE IF (0 IZE + ARR + ABS + MG)
                                     (NOT 0 AORG)
                                     (1 ADI OR ADT);
```

Errorearen detekzioa:

```
"<Hitzaldi>"<HAS_MAI>"
  "hitzaldi" IZE ARR ABS MG HAS_MAI @OBJ @PRED %SINT &OKER_DETK1_0
"<eman>"
  "eman" ADI SIN PART NOTDEK @-JADNAG %ADIKAT
```

Edo **auzia hau* motako sintagma erroredunak detektatzeko, sintagma berean determinatzailea errepikatzen dela adierazi behar da erregelari:

Esleitu (ADD) &OKER_DETE4A_1 errore kodea izenari (TARGET IZE) baldin eta (IF):

- 0. posizioan (0) dagoen izena (IZE), arrunta (ARR), absolutiboa (ABS) eta singularra (NUMS) bada;
- Izena A itsatsiduna ez bada (NOT 0 AORG);
- Izenaren eskuinetara 1. posizioan (1) determinatzaile erakusle (DET-ERKARR) singularra edo plurala (SING-PLUR) eta absolutiboa (ABS) badator;
- Izenaren eskuinetara 2. posizioan (2) *izan* aditza ez badator (NOT 2 IZAN);

Erregela:

```
ADD (&OKER_DETE4A_1) TARGET IZE IF (0 IZE + ARR + ABS + NUMS)
                                     (NOT 0 AORG)
                                     (1 DET-ERKARR + SING-PLUR + ABS);
                                     (NOT 2 IZAN);
```

Errorearen detekzioa:

```
"<auzia>"
  "auzi" IZE ARR ABS NUMS MUGM @OBJ @PRED %SINT &OKER_DETE4A_1
"<hau>"
  "hau" DET ERKARR ABS NUMS MUGM @OBJ @PRED
```

Analizatzaile sintaktikoak emandako analisiak oinarri hartuta idazten ditugu, beraz, determinatzaile-erroreak detektatzeko erregelak, MG formalismoa baliatuz. Errore mota bakoitzari dagokion erregela aplikatzen zaio, eta aplikatutako errore-kodeen bidez egiten da erroreen diagnosi linguistikoa. Horretarako, erregeletan zehaztutako etiketa bakoitza V.3. atalean azaldu dugun euskarazko erroreen eta desbideratzeen sailkapeneko kategoria bati lotuta dago: &OKER_DETK1_0 errore-kodea sailkapeneko 3.3.1.1. azpikategoriari dagokio (*izena + determinatzailea kendu* motako sintagmari), &OKER_DETE4A_1 kodea 3.3.3.7. azpikategoriari (*izena + determinatzaile atzizkia + erakuslea* motako sintagmari) eta &OKER_DETE2A_1 kodea 3.3.3.3. azpikategoriari (*izena + determinatzaile atzizkia + artikulu zehaztugabea* motako sintagmari) (26. irudia).

```
"<Hitzaldi>"<HAS_MAI>"
  "hitzaldi" IZE ARR ABS MG HAS_MAI @OBJ @PRED %SINT &OKER_DETK1_0
"<eman>"
  "eman" ADI SIN PART NOTDEK @-JADNAG %ADIKAT
"<eta>"
  "eta" LOT JNT EMEN AORG @PJ
"<gero>"
  "gero" ADB ARR ZERO @ADLG %SINT
"<Miren>"<HAS_MAI>"
  "Miren" IZE IZB ENTI_HAS_PER HAS_MAI @KM> %SIH
"<Azkarate>"<HAS_MAI>"
  "Azkarate" IZE IZB ABS NUMS MUGM ENTI_BUK_PER HAS_MAI @OBJ @PRED %SIB
"<Kultura>"<HAS_MAI>"
  "kultura" IZE ARR ENTI_HAS_ORG AORG HAS_MAI @KM> %SIH
"<Sailburuak>"<HAS_MAI>"
  "sailburu" IZE ARR ERG NUMS MUGM ENTI_BUK_ORG HAS_MAI @SUBJ %SIB
"<komentatu>"
  "komentatu" ADI SIN PART BURU NOTDEK @-JADNAG %ADIKATHAS
"<du>"
  "*edun" ADL A1 NOR_NORK NR_HURA NK_HARK @+JADLAG %ADIKATBU
"<auzia>"
  "auzi" IZE ARR ABS NUMS MUGM @OBJ @PRED %SINT &OKER_DETE4A_1
"<hau>"
  "hau" DET ERKARR ABS NUMS MUGM @OBJ @PRED
"<guztion>"
  "guzti" DET ORO GEN PH MUGM @IZLG> {POS-HAS12 %SIH
"<artean>"
  "arte" IZE ARR INE NUMS MUGM @ADLG {POS-BUK12 %SIB
"<konpondu>"
  "konpondu" ADI SIN PART NOTDEK @-JADNAG %ADIKAT
"<behar_dugun>"
  "behar_izan" ADI ADK PNT ERLT A1 NR_HURA NK_GUK NOTDEK @+JADNAG_MP_IZLG> %ADIKAT
  "behar_izan" ADI ADK PNT MOS A1 NOR_NORK NR_HURA NK_GUK NOTDEK @+JADNAG_MP_ADLG %ADIKAT
"<kontua>"
  "kontu" IZE ARR ABS NUMS MUGM @SUBJ %SINT &OKER_DETE2A_1
"<bat>"
  "bat" DET DZH NMGS ABS MG @SUBJ
"<dela>"
  "izan" ADT PNT KONPL A1 NOR NR_HURA @+JADNAG_MP_SUBJ %ADIKAT
  "izan" ADT PNT KONPL A1 NOR NR_HURA @+JADNAG_MP_OBJ %ADIKAT
  "izan" ADT PNT KONPL A1 NOR NR_HURA @+JADNAG_MP_PRED %ADIKAT
"<$.>"<PUNT_PUNT>"
  PUNT_PUNT
```

26. irudia. Determinatzaile-erroreak detektatzeko idatzi diren erregelen aplikazioa.

Erregeletan errore-eremuaren hasierak bakarrik zehazten ditugu eta errore-eremuen amaiera-etiketak postprozesu baten bidez ezartzen dira ondoren, automatikoki. Horretarako, errore-etiketaren eskuin muturreko zenbakiak (&OKER_DETK1_0) adierazten digu errore-eremuaren amaiera zenbat hitzetara dagoen: **hitzaldi eman eta gero* adibidean, esaterako, (0) zenbakiak esaten digu errore-eremuaren hasiera eta bukaera hitz berean daudela, *hitzaldi* hitzean; **auzia hau* motako sintagma erroredunak detektatzeko erregelari (&OKER_DETE4A_1), berriz, 1 zenbakiak adierazten du errore-eremuaren bukaera hitz batera dagoela: 0. posizioan dagoen hitzari (*auzia* izenari) dagokio errorearen hasiera eta eskuinerantz 1. posizioan dagoen hitzari (*hau* determinatzaileari) bukaera. Gramatika-zuzentzaileak markatuko dituen errore-eremuak eta eskuz etiketatutakoak bat datoz, hau da, irizpide berdinak jarraitzen ditugu eskuzko etiketatzean eta detekzio automatikoan.

Bestalde, errore mota bakoitzari dagokion eremuak (ikus VI.3. atalean zehaztutako irizpideak) eta erroreen detekzio automatikorako erregeletan baliatzen den informazioak ez dute zertan bat etorri: **hitzaldi eman eta gero* adibidean, esaterako, **hitzaldi* da sintagma erroreduna, markatu beharreko eremua, baina errorea detektatzeko erregelari inguruko hitzei dagozkien ezaugarriak ere zehaztu behar izaten dira. Izan ere, determinatzailearik gabeko *hitzaldi* izen-sintagmak askotariko hitzak izan ditzake inguruan: **hitzaldi eman zuen*, **beste hitzaldi bihar emango du*, **hitzaldi eta gero...* eta horrelako kasu bakoitzeko erregela independenteak egin behar izaten dira askotan, sintagma erroredunaren inguruan dauden hitzak eta hitz horien ezaugarriak zehaztuz. Horregatik, kasu batzuetan, errore mota bat detektatzeko erregela bat baino gehiago idatzi behar izan da, hau da, azpikategoria bakoitzari erregela bat baino gehiago dagokio.

Erregela bakoitzean ondoko informazioa zehazten da: (1) errore mota, sailkapeneko kategoria edota azpikategoriaren arabera; (2) errorearen deskripzioa; (3) errore motari dagokion adibide pare bat; (4) errore motari dago(z)kion zuzenketa(k); (5) errorearen jatorria edo iturria (errore-egitura hori non aurkitu den); (6) oharrak (baldin badaude); eta (7) erregela bera. Adibide gisa

[illegible]

27. taula. Erregela-adibidea.

Kategoria	Kodea	Azpikategorien kopurua
Determinatzailea kendu	DETK	7
Determinatzaile atzizkia gehitu	DETG	5
Determinatzailea errepikatu	DETE	22
DSko osagaien ordena okerra	DETO	2
Mugatasun kontuak	DETMUG	3

28. taula. Landutako errore-kategoria bakoitzean dauden azpikategorien kopurua (VI.4. atala).

detektatzeko eta 57 erregela DETE kategoriako erroreentzako. Halaber, 4 erregela daude DETG kategoriako errore batzuk detektatzeko, erregela bat DETO kategoriako errore mota baterako eta 3 erregela DETMUG kategoriako hainbat errore detektatzeko (29. taula):

Errore mota	Erregela kop.
DETK	20
DETG	4
DETE	57
DETO	1
DETMUG	3
	85

29. taula. Errore mota bakoitzeko erregelaren kopurua.

Lehen hurbilpen honetan ez dugu eskuz etiketatutako errore mota guztien detekzio automatikoa bideratu, eta bigarren urrats batean helduko zaio aztertu gabe geratu diren erroreen tratamendu automatikoari.

Determinatzaile-erroreak detektatzeko erregelak testu errealean etiketatutako adibideetan oinarrituta egin ditugu batik bat, baina hainbat gramatika-liburutan zehaztutako oharrak eta ohiko akatsak ere kontuan izan ditugu. Idatzi ditugun erregelak euskararako garatzen ari garen XUXENg gramatika-zuzentzailean integratzeko asmoa dago (Otegi, 2003; Ansa *et al.*, 2004; Otegi, 2006).

Bestalde, guk baliatutako teknika (MG formalismoa) beste metodo batekin alderatu dugu; izan ere, proba batzuk egin dira determinatzaile-erroreak *machine learning* edo ikasketa automatikoko teknikarekin detektatzeko (Cermeño, 2008). Ikasketa automatikoko teknikarekin lortutako emaitzak, baina, ez dira onak; batetik, erabilitako corpusa ez delako oso handia; bestetik, detektatutako erroreen kopurua egitura zuzenena baino askoz ere txikiagoa delako (erroreen proportzioa % 1,5ekoa da, errore-gabekoen % 98,5aren parean). Adibide erroredun gutxi etiketatuta dituen corpus txikia baliatzea izan liteke, beraz, lortutako emaitza eskasen arrazoietakoa bat. Horregatik, corpus handiagoekin eta errore-adibide gehiagorekin probatu beharko litzateke *machine learning* teknikarekin emaitza hobeak eskuratzea posible ote den. Hala ere, ikasketa automatikoa errore mota batzuen detekziorako baliagarria izan daitekeen arren, teknika horrek ez du erroreen diagnosi linguistikorako erraztasunik eskaintzen, eta erroreen detekzioa eta diagnosia aldi berean egiteko aukera izatea

interesatzen zaigu guri. Bestalde, ikasketa automatikoaren helburua hizkuntza-corpusetatik ezagutza ateratzea da eta horretarako ezinbestekoa da corpusa aldeztetik etiketatuta izatea. MG bidez erregelak egiteko, berriz, ez dago corpus etiketatuen beharrik, eta hori MG formalismoak eskaintzen digun abantaila handi bat da.

Oro har, beraz, ez dago errore mota guztien detekzioarako baliagarria den tekniarik eta errore motaren arabera bide bat edo beste jarraitu beharko da.

VII.5. Detekzio automatikoan aurkitutako arazo batzuk

Determinatzaile-erroreen tratamendu automatikoa bideratzeko prozesuan, errore batzuen detekzioa oztopatu duten edota alarma faltsuak eragin dituzten hainbat arazo azaldu zaizkigu. Arazo horietaz hitz egingo dugu atal honetan.

VII.5.1. Errore ortografikoak

Hizkuntza-ikasleen testuetan ohikoak dira errore ortografikoak eta lexikoak (**boleto hartu nuen*, **euskare ikasi dut*, **zuzentzen naiz saila hortara*, **arratxaldion bostetan altxatu naiz*, **mensaje bat bidali zidan*, **flotador apurtu zen...*). MORFEUS analizatzaile morfosintaktikoak ezezaguna zaion hitza lexikorik gabeko hitz bezala aztertzen du, EZEZAG etiketa esleituz. Hitzaz ezezagunak ez duenez, interpretazio posible batzuk sortuko ditu analizatzaileak eta desanbiguatzaileak, ondoren, analisi batzuk hautatu eta beste batzuk baztertuko ditu. Baina analizatzaileak emandako interpretazioa bat etor liteke, edo ez, erroreari berez legokiokeen analisiarekin; alegia, lexikorik gabeko hitzei esleitutako analisiek bete ditzakete, edo ez, gramatika-errore batzuk detektatzeko definitu diren ezaugarri linguistikoak edota baldintzak. Analizatzailearen analisia erregelan zehaztutakoarekin bat baldin badator, errorea detektatuko da; bestela ez.

**Boleto hartu nuen* esaldi erroredunean, esaterako, *boleto* lexikorik gabeko hitza da eta EZEZAG gisa analizatu da. Esaldi horren egitura **tiket hartu nuen* edo **ikastaro egingo luke* esaldi erroredunen parekoa da eta kasu honetan horrelako adibideei dagokien &OKER_DET_K1_0 erregela ondo aplikatu da. Errore-etiketa egokia aplikatu da, hain zuzen, **boleto* hitz ezezagunari *tiket* eta *ikastaro* izenak

duten analisi berbera esleitu diolako analizatzaileak eta inguruko hitzen analisiak ere erregelan zehaztutako baldintzekin bat datozelako (27. irudia):

```
"<Boleto>"<HAS_MAI>"
  EZEZAG "boleto" IZE ARR ABS MG <HAS_MAI> @OBJ @PRED %SINT &OKER_DETK1_0
"<hartu>"
  "hartu" ADI SIN PART BURU NOTDEK @-JADNAG %ADIKATHAS
"<nuen>"
  "*edun" ADL B1 NOR_NORK NR_HURA NK_NIK @+JADLAG %ADIKATBU
  "*edun" ADL ZHG B1 NOR_NORK NR_HURA NK_NIK @+JADLAG_MP_OBJ %ADIKATBU
"<$.>"<PUNT_PUNT>"
```

27. irudia. EZEZAG etiketa duen lexikorik gabeko hitzean erregela aplikatzea, aplikatu behar denean.

Beste batzuetan, ordea, lexikorik gabeko hitzei analizatzaileak ez die analisi egokia esleitzen. Ondorioz, alarma faltsuak sortzen dira edo erroreak detektatu gabe geratzen dira. Esaterako, 28. irudiko adibidean, *parserrak* ez du *euskare* hitza ulertu, ezezaguna zaio (EZEZAG) eta izen arrunt, absolutibo eta mugagabe gisa analizatu du, *euskara* hitz zuzenari legokiokkeen analisisa hori ez den arren. Hitz horren eta ingurukoen analisiak &OKER_DETK1_0 erregelan zehaztutako baldintzekin bat datozenez, erregela behar ez den lekuan aplikatu da, hau da, alarma faltsua sortu da.

```
"<euskare>"
  EZEZAG "euskare" IZE ARR ABS MG @OBJ @PRED %SINT &OKER_DETK1_0
"<ikasi>"
  "ikasi" ADI SIN PART BURU NOTDEK @-JADNAG %ADIKATHAS
"<dut>"
  "*edun" ADL A1 NOR_NORK NR_HURA NK_NIK @+JADLAG %ADIKATBU
```

28. irudia. EZEZAG etiketa duen lexikorik gabeko hitzean erregela aplikatzea, aplikatu behar ez denean.

*Zuzentzen naiz **saila hortara** esaldian, berriz, *determinatzailea errepikatzea* (DETE) motako errorea daukagu, determinatzaile atzizki singularra duen *saila* izenaren atzetik erakuslea baitator, sintagma berean. Erakuslea, ordea, gaizki idatzita dago (**hortara*) eta ez da determinatzaile gisa desanbiguatu; analizatzaileari **hortara* hitza ezezaguna zaio (EZEZAG) eta adjektibo bati dagokion analisisa (ADJ) esleitu dio. Analisi horrek ez ditu mota horretako erroreari dagokien erregelan zehaztutako baldintzak betetzen eta, beraz, determinatzaile-errorea ez da detektatzen (29. irudia):

```
"<zuzentzen>"
  "zuzendu" ADI SIN AMM ADOIN ASP EZBU NOTDEK @-JADNAG %ADIKATHAS
"<naiz>"
  "izan" ADL A1 NOR NR_NI @+JADLAG %ADIKATBU
"<saila>"
  "sail" IZE ARR DEK ABS NUMS MUGM @OBJ @SUBJ @PRED %SINT
"<hortara>"
  EZEZAG "hortar" ADJ IZO DEK ABS NUMS MUGM AR_TAR @OBJ @SUBJ @PRED
"<$.>"<PUNT_PUNT>"
```

29. irudia. EZEZAG etiketa duen lexikorik gabeko hitzean erregela ez aplikatzea, aplikatu behar denean.

Baina errore ortografikoa zuzenduko balitz (*horretara*), determinatzaile-errorea ongi detektatuko litzateke (30. irudia):

```
"<zuzentzen>"
  "zuzendu" ADI SIN AMM ADOIN ASP EZBU NOTDEK @-JADNAG %ADIKATHAS
"<naiz>"
  "izan" ADL A1 NOR NR_NI @+JADLAG %ADIKATBU
"<saila>"
  "sail" IZE ARR DEK ABS NUMS MUGM @OBJ @SUBJ @PRED %SINT &OKER_DETE4B_1
"<horretara>"
  "hori" DET ERKARR ALA NUMS MUGM @ADLG
"<$.>"<PUNT_PUNT>"
```

30. irudia. Erregela aplikatzea, aplikatu behar denean.

Errore ortografikoak edota lexikoak, beraz, oztopo dira gramatika-erroreak detektatzeko. Ondorioz, gramatika-erroreen detekzio automatikoa egin baino lehen errore ortografikoak zuzentzea komeni dela aurreikusten dugu, gramatika-erroreen detekziorako oztopo den ‘zarata’ neurri handi batean jaitsiko bailitzateke horrela: **euskare* zuzenduko bagenu (*euskara*), ez litzateke alarma faltsurik sortuko; eta **hortara* zuzenduz, berriz, **saila horretara* sintagman dagoen determinatzaile-errorea ongi detektatzen da. Horrelako errore ortografiko batzuen zuzenketa automatikoa, baina, zaila edota ezinezkoa da kasu batzuetan.

VII.5.2. Hitz-segida ez-gramatikalak

Hizkuntza-ikasleen testuetan badira, halaber, ortografikoki zuzenak diren hitzekin osatutako hitz-segida ez-gramatikalak: **eta gero sabela mina dugu*, **txoznetan asko dantza nintzen*, **dantza bere modu bizitzeko da...* Horrelako egituretan ere

determinatzaile-erroreak detektatzeko erregelek alarma faltsuak sor ditzakete (31. irudia):

```
"<eta>"
  "eta" LOT JNT EMEN AORG @PJ
"<gero>"
  "gero" ADB ARR ZERO @ADLG %SINT
"<sabela>"
  "sabel" IZE ARR ABS NUMS MUGM @OBJ @PRED %SINT &OKER_DETE1A1_1
"<mina>"
  "min" ADJ ARR ABS NUMS MUGM @OBJ @PRED %SINT
"<dugu>"
  "ukan" ADT PNT A1 NOR_NORK NR_HURA NK_GUK @+JADNAG %ADIKAT
"<$.>"<PUNT_PUNT>"
```

31. irudia. Hitz-segida ez-gramatikaletan sortutako alarma faltsuak.

Horrelako hitz-segida ez-gramatikalak automatikoki zuzentzea ezinezkoa da eta, beraz, oztopo dira gramatika-erroreen detekzio automatikoan.

VII.5.3. Aurreko urratsetan gertatutako akatsak

Determinatzaile-erroreak detektatzeko garaian aurkitutako arazoen artean ditugu errore-erregelak aplikatu aurreko urratsetan gertatu ahal izan diren erroreak ere: esaterako, desanbiguatzaileak utzitako analisia hitz bati testuinguru jakin batean dagokiona ez izatea. Desanbiguatutako analisia egokia ez denean, determinatzaile-errore batzuk detektatu gabe gera daitezke edota errorerik ez dagoen esaldietan alarma faltsuak sortu ahal dira. Esaterako, 32. irudiko adibideetan letra lodiz markatuta dauden hitzek ez dute esaldi horietan berez dagokien analisia: *merkatuan* hitzak aditzaren (ADI) analisia dauka izenarena eduki beharrean eta *hori* erakusleak adjektiboarena (ADJ). Ondorioz, alarma faltsuak sortu dira zuzenak diren esaldietan:

```

"<sartu>"
    "sartu" ADI SIN PART BURU NOTDEK @-JADNAG %ADIKATHAS
"<dira>"
    "izan" ADL A1 NOR NR_HAIEK @+JADLAG %ADIKATBU
"<indarrez>"
    "indarrez" ADB ARR ZERO @ADLG %SINT
"<lan>"
    "lan" IZE ARR ABS MG @SUBJ %SINT  &OKER_DETK1_0
"<merkatuan>"
    "merkatu" ADI SIN PART INE NUMS MUGM @-JADNAG_MP_ADLG %ADIKAT
"<$.>"<PUNT_PUNT>"

"<etorkizun>"
    "etorkizun" IZE ARR ZERO @KM> %SIH  &OKER_DETK2_1
"<hori>"
    "hori" ADJ ARR ABS MG @OBJ @PRED %SIB
"<iristear>"
    "iritsi" ADI SIN ADOIN EZBU NOTDEK @-JADNAG %ADIKATHAS
"<dago>"
    "egon" ADT PNT A1 NOR NR_HURA @+JADNAG %ADIKAT

```

32. irudia. Desanbiguatzaileak utzitako analisisia egokia ez denean sortutako alarma faltsuak.

Beraz, aurreko urratsetan gertatu ahal izan diren akatsak egon daitezke erregelak idazteko oinarri hartzen ditugun analisisetan eta akats horiek, noski, eragina dute erroreen detekzio automatikoan. Analisi-kateko urratsetan gertatu ahal izan diren akatsak konpontzea zaila izango da kasu batzuetan, baina garrantzitsua da horrelako adibideak detektatzea eta dokumentatzea, ahal den neurrian konpontzen joateko.

VII.5.4. HAULen tratamendua

MORFEUS analizatzaile morfosintaktikoak utzitako analisisiak ere eragina izango du hainbat HAULetan erroreak ongi detektatzeko edota alarma faltsuak sortzeko garaian. HAUL guztiak ez dira oraindik automatikoki elkartuta markatzen, seguruak direnak bakarrik analizatzen dira unitate gisa (Alegria *et al.*, 2004). Ondorioz, *gehiago ikasteko **ikastaro** egingo luke esaldi erroredunean &OKER_DETK1_0 erregela aplikatzen den moduan (33. irudia), ez dut **hitz** egingo esaldi zuzenean ere

aplikatu egingo litzateke erregela bera (34. irudia), *hitz egingo* unitate gisa analizatuta ez datorrenez, bi esaldietako izenek (*ikastaro* eta *hitz*) eta aditzek (*egin*) analisi berdinak dituztelako⁴²:

```
"<gehiago>"
  "gehiago" DET DZG MG ABS MG @OBJ @PRED %SINT
"<ikasteko>"
  "ikasi" ADI SIN ADIZE GEL ZERO @-JADNAG_MP_IZLG> %ADIKAT
"<ikastaro>"
  "ikastaro" IZE ARR ABS MG @OBJ @PRED %SIB &OKER_DETK1_0
"<egingo>"
  "egin" ADI SIN PART GERO NOTDEK @-JADNAG %ADIKATHAS
"<luke>"
  "*edun" ADL B2 NOR_NORK NR_HURA NK_HARK @+JADLAG %ADIKATBU
```

33. irudia. Ongi detektatzen den determinatzaile-errorea.

```
"<ez>"
  "ez" PRT EGI @PRT %ADIKATHAS
"<dut>"
  "*edun" ADL A1 NOR_NORK NR_HURA NK_NIK @+JADLAG %ADIKATETEN
"<hitz>"
  "hitz" IZE ARR ABS MG @OBJ @PRED %SINT &OKER_DETK1_0
"<egingo>"
  "egin" ADI SIN PART GERO NOTDEK @-JADNAG %ADIKAT
```

34. irudia. 'hitz egin' HAUL gisa analizatuta ez egoteagatik sortuko litzatekeen alarma faltsua.

Edo *asko hitz egin* zuen adibidean, *hitz egin* unitate gisa analizatuta ez datorrenez, bi errore-erregela aplikatuko lirateke: determinatzailea kentzeagatik eta Dsko osagaien ordena aldatzeagatik sortzen diren erroreak detektatzeko erregelak (&OKER_DETK1_0 eta &OKER_DETO1_1, hurrenez hurren) (35. irudia):

⁴² Badaude dagoeneko unitate gisa analizatzen diren hainbat aditz-esapide (*nahi_izan*, *behar_izan*, *lo_egin*...). Kasu horietan ez dago lehenengo osagaiaren gainean (izenaren gainean) alarma faltsuak sortzeko arriskurik.


```

"<asko>"
  "asko" ADB ARR GRAD @ADLG &OKER_DET01_1
"<hitz>"
  "hitz" IZE ARR ABS MG @OBJ @PRED %SINT &OKER_DETK1_0
"<egin>"
  "egin" ADI SIN PART BURU NOTDEK @-JADNAG %ADIKATHAS
"<zuen>"

```

35. irudia. 'hitz egin' HAUL gisa analizatuta ez egoteagatik sortuko lirakeen alarma faltsuak.

Erregeletan aurkakorik adierazi ezean, horrelako adibideetan alarma faltsuak izango genituzke. Horrelakoak ekiditeko, lehen urrats honetan corpusetan agertu zaizkigun adibide jakin batzuei dagozkien salbuespeneko baldintzak zehaztu ditugu erregeletan, *ad hoc*. Euskararen azterketa eta prozesamenduaren alorrean aurrera egin ahala, erroreen gramatiketan ere aldaketak, hobekuntzak, egokitzapenak edota zuzenketak egiten joan beharko dugu.

VII.5.5. Hizkuntzaren konplexutasuna

Hizkuntzaren beraren konplexutasunak edota anbiguotasunak zaildu egiten du erroreen detekzio automatikoa. Esaldi batzuk, esaterako, zuzenak edota erroredunak izan daitezke, testuinguruaren arabera.

Gerta liteke, adibidez, esaldi-mailan zuzentzat hartzen diren sintagmak testu-mailan erroredunak izatea:

Bapatean <DET?_HAS>hiru neskek<DET?_BUK> agertu ziren primeran adibidea, esaterako, zuzena da esaldi-mailan. Testu osoa irakurrita, ordea, etiketatzaileak badaki ezezagunak diren edota aurretik aipatu ez diren hiru neskei egiten zaiela erreferentzia eta, hortaz, *neska* hitzak ez duela determinatzaile atzizki pluralaren beharrik. Esaldi-mailan zuzentzat hartzen den sintagma, beraz, erroreduna izan daiteke testuingurua kontuan hartuta.

Gerta liteke, halaber, sintagma-mailan erroreduntzat hartzen diren egiturak, esaldi-mailan zuzenak izatea:

<DET?_HAS>Laguna hura<DET?_BUK> agurtu eta gero etorri zen esaldian etiketatzaileak badaki, testua irakurrita, *agurtu* aditzaren osagarri

zuzena dela **laguna hura*. Sintagma erroreduna da, sintagma batek ezin baititu bi determinatzaile hartu (determinatzaile atzizki singularra eta erakuslea, kasu honetan). Esaldi hau, ordea, zuzena ere izan zitekeen *laguna* balitz *etorri* aditzaren subjektua eta *hura* balitz *agurtu* aditzaren osagarri zuzena (*laguna (gizon) hura agurtu eta gero etorri zen*). Beraz, sintagma-mailan erroretzat hartzen diren egitura batzuk zuzenak izan daitezke esaldi-mailan.

Eskuzko etiketatzean DET? etiketa esleitzen diegu horrelako adibideei, galdera ikurrarekin adieraziz testuinguruaren arabera zuzentzat edo erroretzat hartuko diren adibideak direla. Horrelako adibideen tratamendu automatikoa, ordea, oso zaila da eta horretarako euskararen azterketa eta prozesamenduaren alorrean hainbat ekarpen egin beharra dago: beharrezkoak izango dira, besteak beste, testu-mailako azterketa automatikoa, galdera-erantzunen sistemen garapena edota anaforen tratamendua.

VII.5.6. Hizkuntza-ekoizpen gramatikalak aztertzeko baliabideak erabiltzea

Euskararen azterketa eta prozesamenduaren alorrean garatu ditugun oinarritzko baliabideak (analizatzaile sintaktikoa edota EDBL datu-basea, esaterako) hizkuntza-ekoizpen gramatikalen tratamendu automatikoa egiteko sortu dira, ez erroreak tratatzeko. Ondorioz, egitura erroredunei esleitzen zaien analisia askotan ez da berez erroreei dagokiena. IXATI zatitzaileak, esaterako, sintagmen hasierak (%SIH) eta bukaerak (%SIB) zehazten ditu (edota %SINT etiketa esleitzen die hitz bakarreko sintagmei) sintagmak zuzenak direnean. Sintagma erroredunen kasuan, ordea, sintagmen mugak ez dira behar bezala zehazten. Adibide gisa erabili dugun esaldira bueltatuz (**Hitzaldi eman eta gero Miren Azkarate Kultura Sailburua komentatu du auzia hau guztion artean konpondu behar dugun kontua bat dela*), **auzia hau* sintagma erroredunean, esaterako, *auzia* izenak (IZE) absolutibo (ABS) singular (NUMS) mugatuaren (MUGM) analisia dauka, eta *auzi hau* sintagma zuzenean, berriz, ZERO marka (kasu-markarik ez duela adierazten duen analisia). Hala, sintagma zuzenean sintagmaren hasiera (%SIH) eta bukaera (%SIB) ondo mugatu dira, baina sintagma erroredunean ez: IXATIk absolutibo marka duen *auzia* izen-sintagmatzat hartu du eta %SINT etiketa esleitu dio (36. irudia):

```

"<auzia>"
    "auzi" IZE ARR ABS NUMS MUGM @OBJ @PRED %SINT
"<hau>"
    "hau" DET ERKARR ABS NUMS MUGM @OBJ @PRED

"<auzi>"
    "auzi" IZE ARR ZERO @KM> %SIH
"<hau>"
    "hau" DET ERKARR ABS NUMS MUGM @OBJ @PRED %SIB

```

36. irudia. *auzia hau eta auzi hau sintagmen analisiak.

Beraz, sintagma erroredunetan eta horiei dagozkien egitura zuzenetan osagaiek ezaugarri desberdinak dituzte, analizatzaileak hitz bakoitzari esleitzen dion analisia desberdina da. Euskararako analizatzaile sintaktikoak hizkuntza-ekoizpen gramatikalak ditu aztergai eta helburu (ez testu erroredunak). Ondorioz, sintagmen mugak eta dependentziak esleitzeko edota funtzio sintaktikoak desanbiguatzeko gramatikak, esaterako, erlaxatu egin beharko lirатеke egitura erroredunetan etiketa horiek behar diren lekuetan jartzeko. Eta horixe egiten du Oronozek (2009), bere tesi-lanean, komunztadura-erroreak edota data-egituretan egiten diren erroreak detektatzeko: komunztadura-errore batzuk detektatu ahal izateko, mendekotasun-erlazioak esleitzen dituen gramatikan zehaztutako baldintzak erlaxatu behar izan ditu; eta data-egituretan egindako erroreak detektatzeko ere “mailaz mailako erlaxazio” moduko bat erabili du: data-egituretan gerta daitezkeen errore motak aurreikusi eta errore posible horiek kontuan izanik idatzi ditu erroreen detekziorako erregelak. Determinatzaile-erroreen detekziorako, berriz, ez dugu erlaxazioaren beharrik ikusi eta postposizio-erroreak detektatzeko ere Oronozek (2009) ez du erlaxaziorik erabiltzen. Errore mota batzuekin ‘erlaxazioak’ erabili behar izaten dira, beraz, erroreen detekzioa helburu duen analizatzaile sintaktikorik ez dagoenean.

Maritxalarrek (1999), bere aldetik, egitura linguistiko ez-gramatikalak analizatzea helburu duen *parser* baten garapena proposatzen du bere tesi-lanean, erroreen tratamendu automatikorako analisi-kate zuzena soilik erabiltzea ez baita egokia bere ustetan. Gainera, euskara-ikasleen hizkuntza-maila bakoitzari dagokion tarte-hizkuntza finkatu eta informazio hori analisi-katean txertatzeko beharra planteatzen du; alegia, analizatzaile, desanbiguatzaile eta lematizatzaileak

hizkuntza-maila bakoitzari dagozkion ezaugarriak kontuan hartu eta hizkuntzaren analisi linguistikoa egitean informazio hori baliatzea ekarpen garrantzitsua dela uste du, erabiltzaileen mailara egokitutako analisiak egingo dituzten tresnak garatu ahal izateko.

Guk ere ikusi dugu euskara estandarren azterketa eta prozesamendurako sortu diren tresnek mugak dituztela erroreen tratamendu automatikorako, ez baitira helburu horrekin garatu. Euskarazko erroreen analisiari ekin zitzaionean, hain zuzen, planteatu zen hizkuntza-ekoizpen gramatikalen azterketa eta prozesamendurako garatu zen *parserra* egokia ote zen egitura erroredunen tratamendurako; ez ote zen hobe hizkuntza ez-gramatikala aztertzeko adibide erroredunetan oinarritutako baliabideak garatzea. Lan hori oso garestia da, ordea, eta orain artean ez dugu eskura izan hori horrela dela frogatzeko azterketak egiteko bitartekorik. Ondorioz, euskara gramatikala eta estandarra aztertzeko sortu dugun analizatzaile sintaktikoak ematen digun informazioan oinarritu behar izan dugu testu ez-gramatikalak analizatu eta erroreen tratamendu automatikoa bideratzeko, horrek hainbat muga dituela badakigun arren.

VII.6. Erregelen ebaluazioa

Eskuz etiketatutako determinatzaile-erroreen detekzio automatikoa bideratzeko, MG formalismoan oinarritutako 85 erregela egin eta ebaluatu ditugu. Atal honetan erregela horien ebaluazioan lortutako emaitzak aurkeztuko ditugu.

Hizkuntzaren azterketa eta tratamendu automatikoaren alorrean garatzen diren tresnak adibide errealetan oinarrituta egon ohi dira. Erroreak automatikoki tratatzeko teknikak, sistemak edota erregelak egiteko ere ezinbestekoa da adibideak eskura izatea. Hiru informazio-iturri edo bitarteko baliatu ohi dira, oro har, errore-adibideak biltzeko: testu errealak, ikertzaileek asmatutako adibideak eta automatikoki sortutakoak. Guk testu errealetan etiketatutako adibideetan oinarrituta egin ditugu, batik bat, determinatzaile-erroreak detektatzeko erregelak, hau da, corpora hartu dugu euskarazko erroreen edota desbideratzeen azterketarako eta ebaluaziorako abiapuntutzat.

Corpusa bi zatitan banatu dugu: *garapenerako corpusa* eta *probarako corpusa*. Garapenerako corpusa esaten zaio erregelak egiteko oinarri hartzen denari:

corpuseko adibideetan oinarrituta erregelak idazten dira, ondo aplikatzen diren aztertu eta beharrezko zuzenketak edo aldaketak egiten dira erregelak gehiago findu ezin direla ikusten den arte. Probarako corpora, berriz, azken ebaluazioa egiteko baliatzen den corpora da: erregelak behin bakarrik aplikatzen dira testuan eta emaitzak aztertzen dira.

Erregelak ebaluatzeko, lau egoera hauek hartzen ditugu kontuan:

- *Gezurrezko Negatiboak*: testuan dauden egitura erroredunak ez detektatzea.
- *Gezurrezko Positiboak*: egitura zuzenak okerrak balira bezala markatzea, alarma faltsuak.
- *Egiazko Negatiboak*: egitura zuzenak ez markatzea (zuzenak zuzentzat hartzea).
- *Egiazko Positiboak*: testuan dauden erroreak ongi detektatzea.

Egoera edo parametro horiek erabili ditugu, hain zuzen, determinatzaile-erroreen detekzio automatikorako baliatu dugun errore-gramatikaren doitasuna (*precision*) eta estaldura (*recall*) neurtzeko. Detektaturiko erroreen zuzentasuna neurtzen du doitasunak, hau da, automatikoki detektatu diren errore guztien artean zenbat errore detektatu diren ongi adierazten du. Estaldurak, berriz, detektatu beharko liratekeen errore guztietatik zenbat errore detektatu diren esango digu. Bi parametro horiek ondoko ekuazio hauen bidez kalkulatzen dira:

$$\text{DOITASUNA} = \frac{\text{egiazko positiboak}}{\text{egiazko positiboak} + \text{gezurrezko positiboak}}$$

$$\text{ESTALDURA} = \frac{\text{egiazko positiboak}}{\text{egiazko positiboak} + \text{gezurrezko negatiboak}}$$

Erroreak detektatzeko egiten diren erregelak ebaluatzeko euskara-ikasleen corpora eta testu zuzenak erabili ohi dira. Euskara-ikasleen corpusean ongi aplikatu diren erregelen kopurua neurtzen da bereziki; testu zuzenetan, berriz, sor daitezkeen alarma faltsuena. Hala ere, litekeena da hizkuntza-ikasleen corpusean alarma faltsuak

aurkitzea eta testu zuzenetan erroreren bat edo beste detektatzea. Determinatzaile-erroreen erregelak egiteko, euskara-ikasleen testuak eta Euskara Teknikoko ikasleenak etiketatu ditugu eskuz (VI.5. atala). Erroreen detekzio automatikorako idatzi ditugun erregelak ebaluatzeko, berriz, eskuz etiketatutako hizkuntza-ikasleen corpora eta euskara-ikasker ez diren hiztunen testuak (Euskara Teknikoko ikasleen lanak eta *Euskaldunon Egunkaria* corpora) baliatu ditugu.

VII.6.1. Erroreen detekzioa euskara-ikasleen testuetan

Euskara-ikasleen corpusari dagokionez, atal bi bereizi ditugu: erregelak egiteko erabili dugun *garapenerako corpora* (% 75) eta erregelak ebaluatzeko *probarako corpora* (% 25) (30. taula):

	Guztira	Garapenerako corpora (% 75)	Probarako corpora (% 25)
Hitz kopurua	113.290	84.976	28.314
Sintagma kopurua	38.122	28.591	9.531

30. taula. Garapenerako eta probarako baliatu dugun euskara-ikasleen corpora.

Corpus osoan eskuz etiketatutako 788 adibideetatik (VI.5. atala) errore mota jakin batzuen tratamendu automatikoa bideratu dugu lehen urrats honetan, 85 erregelaren bidez. Erregelak euskara-ikasleen probarako corpus-laginean ebaluatu ditugu (9.531 sintagmako corpora). Corpus-lagin horretan, 134 adibide etiketatu dira eskuz; automatikoki, berriz, 60 errore detektatu dira ongi, 72 gaizki eta beste 74 errore ez dira detektatu. Beraz, erregelen doitasuna % 45,45 da eta estaldura % 44,77 (31. taula).

	Ongi detektatuak	Gaizki detektatuak	Ez detektatuak	Guztira	Doitasuna	Estaldura
Detekzio automatikoa	60	72	74	132	% 45,45	% 44,77
Eskuzko etiketatzea	134					

31. taula. Euskara-ikasleen probarako corpusaren gainean egindako ebaluazioaren emaitzak.

Errore motei dagokienez, DETK, DETG, DETE, DETO eta DETMUG azpikategoriako determinatzaile-errore batzuk detektatzeko erregelak egin dira.

Errore mota bakoitzaren barruan dauden azpikategorien kopurua desberdina denez, errore mota bakoitzeko idatzi den erregela-kopurua ere diferentea da (32. taula):

Errore mota	Erregela kopurua
DETK	20
DETG	4
DETE	57
DETO	1
DETMUG	3

32. taula. Errore mota bakoitzeko erregelen kopurua.

Doitasunik handiena (% 83,3) DETO motako erroreak detektatzen dituen erregelak du, seguruenik ere kategoria horretako errore mota bakar bat landu dugulako. Gainontzeko kategorietan aztertutako kasuistika, aldiz, zabalagoa izan da. DETE motako erroreak detektatzeko erregeletan % 50eko doitasuna lortu da, DETK motako erroreak detektatzeko erregeletan % 37,8koa; eta DETMUG motako erroreei dagokien erregeletan % 30ekoa. Estaldurari dagokionez, berriz, DETMUG motako erroreak detektatzeko erregelek dute estaldura handiena (% 60) eta datuak antzekoak dira DETK, DETE eta DETO motei dagozkien erregelen kasuan (% 43, % 42,8 eta % 45,4 hurrenez hurren). DETG kategoriako errorerik ez dago probarako corpusean (33. taula).

Errore mota	Erregela kopurua	Doitasuna (%)	Estaldura (%)
DETK	20	37,8	43
DETG	4	-	-
DETE	57	50	42,8
DETO	1	83,3	45,4
DETMUG	3	30	60

33. taula. Probarako corpusaren gainean egindako ebaluazioaren emaitzak, errore mota bakoitzeko.

Euskara-ikasleen probarako corpusean, beraz, automatikoki detektatu diren 132 determinatzaile-erroreetatik 60 (% 45,45) ongi detektatu dira. Ongi detektatutako hiru adibide erroredun erakusten dizkigu 37. irudiak:

```

"<eta>"
    "eta" LOT JNT EMEN AORG @PJ
"<eguzki>"
    "eguzki" IZE ARR ABS MG @OBJ @PRED %SINT &OKER_DETK1_0
"<hartu>"
    "hartu" ADI SIN PART BURU NOTDEK @-JADNAG %ADIKATHAS
"<nuen>"
    "*edun" ADL B1 NOR_NORK NR_HURA NK_NIK @+JADLAG %ADIKATBU
    "*edun" ADL ZHG B1 NOR_NORK NR_HURA NK_NIK @+JADLAG_MP_OBJ %ADIKATBU

"<asko>"
    "asko" ADB ARR GRAD @ADLG &OKER_DET01_1
"<odola>"
    "odol" IZE ARR ABS NUMS MUGM @OBJ @PRED
"<kampora>"
    "kanpora" "kampora" ADB ARR ZERO AORG @ADLG %SINT
"<atera>"
    "atera" ADI SIN PART BURU NOTDEK AORG @-JADNAG %ADIKATHAS
"<zuen>"
    "*edun" ADL B1 NOR_NORK NR_HURA NK_HARK @+JADLAG %ADIKATBU
    "*edun" ADL ZHG B1 NOR_NORK NR_HURA NK_HARK @+JADLAG_MP_OBJ %ADIKATBU

"<Urrengo>"<HAS_MAI>"
    "hurrengo" "urrengo" ADJ ARR ZERO HAS_MAI @IA> @KM> %SIH
"<goizan>"
    EZEZAG "goiza" IZE ARR INE NUMS MUGM AORG @ADLG %SIB
"<erresaka>"
    "erresaka" IZE ARR ZERO AORG @KM> %SIH &OKER_DETK2_1
"<haundi>"
    "handi" "haundi" ADJ ARR ABS MG @OBJ @PRED %SIB
"<nuen>"
    "ukan" ADT PNT B1 NOR_NORK NR_HURA NK_NIK @+JADNAG @-JADNAG %ADIKAT
    "ukan" ADT PNT ZHG B1 NOR_NORK NR_HURA NK_NIK @+JADNAG_MP_OBJ %ADIKAT
"<$.>"<PUNT_PUNT>"

```

37. irudia. Ongi detektatutako errore-adibideak.

Eta 72 alarma faltsu egon dira (% 54,55). Alarma faltsuak eragin dituzten kausak VII.5. atalean azaldu ditugu xehetasunez. 72 alarma faltsuetatik 25 (% 34,72) EZEZAG etiketa duten lexikorik gabeko hitzetan gertatu dira; 16 alarma faltsu (% 22,23) hitz-segida ez-gramatikaletan; 14 (% 19,44) desanbiguatzaileak utzitako analisi desegokietan; eta 2 alarma faltsu (% 2,78) eskuzko etiketatzean ahaztu diren adibideei dagozkie (ebaluazioak, beraz, eskuz etiketatu ez diren erroreak detektatzeko ere balio digu). Benetako alarma faltsuak 15 dira (% 20,83).

34. taulak erakusten dizkigu euskara-ikasleen probarako corpusean sortu diren alarma faltsuen kopuruak:

	Kop.	%
Lexikorik gabeko hitzak	25	34,72
Esaldi-egitura arraroak	16	22,23
Desanbiguatzailearen analisi desegokia	14	19,44
Eskuzko etiketatzean ahaztutakoak	2	2,78
Benetako alarma faltsuak	15	20,83
Guztira	72	100

34. taula. Alarma faltsuen kopuruak.

Alarma faltsuen artean, beraz, 57 (% 79,17) ez dagozkien erroreak detektatzeko idatzi diren erregelari. Ondorioz, horiek kontuan hartzen ez badira, idatzitako erregelen doitasuna nabarmen hobetzen da, % 44,45etik % 80ra igoaz (35. taula).

	Ongi detektatuak	Gaizki detektatuak	Guztira	Ez detektatuak	Doitasuna	Estaldura
Detekzio automatikoa	60	15	75	74	% 80	% 44,77

35. taula. Probarako corpusean egindako ebaluazioaren emaitzak, erregelari ez dagozkien alarma faltsuak kontuan hartu gabe.

Interesgarria iruditu zaigu doitasuna bi modutara ebaluatzea: egoera erreala (testuetan dauden bestelako erroreak, desanbiguatzaileak utzitako analisi desegokiak, eskuzko etiketatzean ahaztutako adibideak...) kontuan hartuz eta testuak, desanbiguatzailearen analisiak eta etiketatzailearen lana zuzenak direla suposatuz. Izan ere, horrela ikus daiteke zein den erregelen berezko doitasuna (% 80) eta erregelak XUXENg gramatika-zuzentzailean integratzen direnean izango dutena (% 45,45). Alarma faltsu batzuk ekiditea, bestalde, bideragarria izan daiteke (errore ortografiko batzuk gramatika-erroreen detekzioa baino lehen zuzenduz, aurreko urratsetan gertatu ahal izan diren akatsak konponduz...), baina beste batzuk saihestea (hitz-segida ez-gramatikalen zuzenketa automatikoa, esaterako) oso lan zaila da.

VII.6.2. Alarma faltsuak testu zuzenetan

Determinatzaile-erroreak detektatzeko idatzi ditugun erregelak euskara-ikasleen probarako corpusean ebaluatzeaz gain, euskara-ikasleak ez diren hiztunen testuetan

ere ebaluatu ditugu: espezialitate-euskarako ikasleen lanak (5.963 sintagmako corpus-lagina) eta *Euskaldunon Egunkaria* corpusaren lagin bat (16.434 sintagma) baliatu ditugu horretarako. Corpus horiek probarako corpus gisa erabili ditugu soilik, ongi detektatutako erroreen (egiazko positiboen) eta alarma faltsuen (gezurrezko positiboen) kopuruak ezagutzeko, idatzitako erregelen doitasuna eta estaldura kalkulatu gabe.

Testu horietan automatikoki detektatutako adibideak oso gutxi dira: % 0,62 eta % 0,33, hurrenez hurren, corpuseko sintagma kopuruarekiko (36. taula). Datu horiek erakusten digute, beraz, hizkuntza-ikasleak ez diren hiztunek idatzitako testuetan oso determinatzaile-errore gutxi egiten direla eta alarma faltsuen kopurua ere baxua dela.

	Hitz kopurua	Sintagma kopurua	Detektatutako adibideak	%
Espezialitate euskarako ikasleen testuak	19.391	5.963	37	0,62
Euskaldunon Egunkaria	53.658	16.434	55	0,33

36. taula. Euskara-ikasleak ez diren hiztunen testuetan detektatutako adibideen kopuruak, sintagma-kopuruarekiko.

Espezialitate-euskarako ikasleen testuetan 37 determinatzaile-errore detektatu dira automatikoki (% 0,62). Horietatik 2 (% 0,03) egiazko positiboak dira (erroreen gramatikak ongi detektatu dituen determinatzaile-erroreak) eta 35 (% 0,58) alarma faltsuak. Alarma faltsu horietatik, ordea, benetakoak 13 dira (% 0,21). Gainerako alarma faltsuen artean, 11 (% 0,18) dira EZEZAG etiketa duten lexikorik gabeko hitzetan aplikatutako erregelak, 9 (% 0,15) testuinguruan ez dagokien analisia duten hitzetan aplikatutakoak eta 2 (% 0,03) zalantzazko kasuak (37. taula).

Euskaldunon Egunkariako testuetan, berriz, 55 errore detektatu dira guztira (% 0,33), espezialitate-euskarako ikasleen testuetan baino are gutxiago, proportzioan. Horietatik 3 (% 0,01) egiazko positiboak dira (ongi detektatutako determinatzaile-erroreak) eta 52 (% 0,31) alarma faltsuak. Alarma faltsuen artean, 26 (% 0,15) dira benetakoak, 13 (% 0,07) EZEZAG analisia duten lexikorik gabeko hitzetan aplikatutako erregelak, 11 (% 0,06) testuinguruan ez dagokien analisia duten hitzetan aplikatutakoak eta 2 (% 0,01) zalantzazko kasuak (37. taula).

	IS kop.	Errore kop. guztira	Egiazko positiboak		Gezurrezko positiboak		Benetako gezurrezko positiboak	
			kop.	%	kop.	%	kop.	%
Espezialitate euskarako ikasleen testuak	5.963	37 (%0,62)	2	0,03	35	0,58	13	0,21
Euskaldunon Egunkaria	16.434	55 (%0,33)	3	0,01	52	0,31	26	0,15

37. taula. Euskara-ikasleak ez diren hiztunen testuetan detektatutako egiazko positiboak eta gezurrezko positiboak kopuruak, IS kopuruarekiko.

Testu zuzenetan egindako ebaluazioan ikusi dugu, beraz, determinatzaile-erroreen kopurua oso baxua dela hizkuntza-ikasle ez diren hiztunek idatzitako corpusetan, hau da, determinatzaile-erroreak euskara-ikasleek egin ohi dituzten erroreak direla batik bat. Ikusi dugu, halaber, alarma faltsuen kopurua oso baxua dela testu zuzenetan; eta horixe izan da gure helburuetako bat: nahiago izan dugu alarma faltsuen kopurua ahalik eta baxuen mantendu, errore gehiago detektatu eta alarma faltsuen kopuruak gora egitea baino. Izan ere, egokiagoa da, gure ustez, errore mota gutxiago detektatzea, alarma faltsuen kopurua areagotzea baino; batetik, alarma faltsuek erabiltzaileen jakintza zalantzan jar dezaketelako (berezihi hizkuntza-ikasleen kasuan) eta bestetik, alarma faltsu asko dituzten tresnek eraginkortasuna galtzen dutelako.

VII.7.Hitzak deklinatzean determinatzaile atzizkia ongi ez erabiltzeagatik sortutako erroreen azterketa

Eskuzko etiketatzean, hitzak deklinatzean *-a* determinatzaile atzizkia behar denean kentzeagatik edota behar ez denean gehitzeagatik egindako erroreak etiketatu ditugu, beste determinatzaile-errore motekin batera. Errore hauek gainontzeko erroreak ez bezalakoak dira, beste ezaugarri batzuk dituzte. Hitzaren eta atzizkiaren arteko lotura egitean sortzen diren hitz-mailako erroreak dira:

- Hitzak deklinatzean *-a* determinatzaile atzizkia kentzeagatik egin diren erroreak DETK_DEKL kodearekin etiketatu ditugu eskuz. Errore hauen artean hitz batzuk forma erroredunak dira beti: **etxen*, **amakin...*; beste batzuk, berriz, ortografikoki zuzenak zein erroredunak izan daitezke, esaldiaren arabera: **nire lehenengo **kotxerekin** joan ginen, hainbat **kotxerekin** moldatzen dira, *nola ospatzen den urte hau beste **hiritan**, hainbat **hiritan** gertatzen da*

*gauza bera, *nere amak **auzokideri** oparitu zion, edozein **auzokideri** ez zaio gustatuko...*

- Hitzak deklinatzean *-a* determinatzaile atzizkia gehitzeagatik egiten diren erroreak, berriz, DETG_DEKL kodearekin etiketatu ditugu. Hitz hauek ortografikoki okerrak dira kasu guztietan: **eskuminak Vincennesko parkeko **basauntzaei**, *neguak hotza eta gosea ekarri zituen **herriara**, *gu **zelaia**tik igo ginen...*

IXA taldean hitz-mailako erroreen detekzioa eta tratamendu automatikoa EDBL datu-basean (Aldezabal *et al.*, 2001) biltegitratutako informazioan eta bi mailatako morfologian (Koskeniemi, 1983) oinarrituta egiten da. EDBL hainbat aplikaziotarako oinarrizko informazio morfologikoa eta lexikoa gordetzen duen datu-basea da, besteak beste XUXEN zuzentzaile ortografikoaren (Agirre *et al.*, 1992) informazio-iturria. Datu-basearen helburua erroreen tratamendu automatikoa ez den arren, badu erroreen sailkapen bat, lexiko mailako forma ez-estandarrak barne hartzen dituenak: **auzogune (auzune), *esku hutsik (esku-hutsik) edota *lehen bait lehen (lehenbailehen) bezalako konposizio-erroreak; *altaboz (bozgorailu), *naranja (laranja) edo *silla (aulki) bezalako mailegu okerrak; *aitu (aditu), *goxoki (gozoki) edota *parre (barre) motako aldaera dialektalak...; eta deklinabide-atzizki ez-estandar batzuk ere badaude, hala nola: *-agaitik, *-agaz, *-gatio, *-taz, *-tikan...* Biltegitratutako adibide horiei esker, zuzentzaile ortografikoa gai da lexiko mailako hainbat errore detektatzeko; eta horien artean, DETK_DEKL eta DETG_DEKL motako errore batzuk. Baina aipatu azpikategoria horiei dagozkien adibide guztiak ez zirela detektatzen ikusi genuen eskuz etiketatutako adibideak aztertuta.

Zuzentzaileak errore horiek detektatu eta diagnostikatzeko modu bat EDBL datu-basean deklinabide-atzizki erroredun berriak sartzea izan zitekeela aurreikusi genuen. Adibidez, EDBLn *-ara deklinabide-atzizki okerra (*-a* determinatzaile singularraz eta *-ra* kasu-markaz osatua) balego, zuzentzaileak **etxeara* errorea detektatuko luke, *etxe* lema zuzenari **-a+ra* deklinabide-atzizki erroreduna gehitu zaiola identifikatuko luke (diagnosi linguistikoa egingo luke) eta horrela, ordezeko zuzenaren proposamen egokia eman ahalko luke. Adibide horrekin batera, beste errore mota batzuk ere aurkitu ditugu: **-a+tik* deklinabide-atzizki okerra sartuz, **bideatik (bide+*-a+tik)* motako erroreak identifikatuko lirateke, **-a+ko*

deklinabide-atzizki okerrarekin **institutuako* (*institutu+*-a+ko*) motakoak, etab. Hain zuzen ere, *-a* determinatzaile atzizki singularraz eta lekuzko kasu-markaz (*-n*, *-tik*, *-rik*, *-ra*, *-raino*, *-ko*, *-rako* eta *-rantz*) osatzen dira **-a+n*, **-a+tik*, **-a+rik*, **-a+ra*, **-a+raino*, **-a+ko*, **-a+rako* eta **-a+rantz* atzizki-konbinazio okerrak.

Euskara-ikasleen corpusetan detektatutako adibideak oinarri hartuta, hainbat errore diagnostikatzeko beharrezkoak diren deklinabide-atzizki okerren zerrenda osatu dugu (38. taula). Lehenengo zutabea DETK_DEKL edo DETG_DEKL azpikategoriako errorea den zehazten da; bigarren zutabea deklinabide-atzizki erroredunak zerrendatzen dira, errore-adibide batzuekin batera; hirugarren zutabea deklinabide-kasu erroredun bakoitzari dagokion forma zuzena ikus daiteke.

Errore mota	Deklinabide erroreduna	Baliokide zuzena
DETK_DEKL	*-kin (*soldatakin, *amakin...)	-(r)ekin / -arekin / -ekin (soldatarekin, amarekin...)
	*-n (*etxen, *mendin, *herrikon...)	-an (etxean, mendian, herrikoan...)
	*-an (*oihanan, *ertzan...)	-(e)an (oihanean...)
DETG_DEKL	*-etik (*nagusietik, *jaietik...)	-tik (nagusitik, jaitik...)
	*-atik (bideatik, mendiatik...)	-tik (bidetik, menditik...)
	*-ara (*iturriara, *parkeara, *ibaiarako ...)	-ra (iturrira, parkera, ibairako ...)
	*-araino (*tokiaraino, *lakuaraino...)	-raino (tokiraino, lakuraino...)
	*-ako (*institutuako, *etxeako...)	-ko (institutuko, etxeko...)
	*-arantz (*kontzertuarantz, *zinearantz...)	-rantz (kontzerturantz, zinerantz...)
	*-aei (*basahuntzaei, *ikastetxeaiei...)	-ei (basahuntzei, ikastetxeei...)
	*-arik (*soinuarik, *sudurrarik...)	-(r)ik (soinurik, sudurririk...)

38. taula. Errore mota batzuk detektatu eta diagnostikatzeko EDBLn sartutako deklinabide-atzizki okerren zerrenda.

EDBL datu-basea horrelako formekin aberasteko aukera interesgarria iruditzen zaigu; izan ere, deklinabide-atzizki erroredun berriak datu-basean gordeta, hainbat erroreren detekzioa eta diagnosia bideratu ahal izan dugu. Datu-basean gordetako adibide erroredunak erroreen sailkapenari lotuz gero, gainera, detektatutako erroreen (hitz-mailako erroreen) diagnosi linguistikoa egiteko aukera izango genuke. Eta hori ekarpen garrantzitsua iruditzen zaigu erroreen eta desbideratzeen tratamendu automatikoaren alorrean.

EDBL, baina, ez da baliabide nahikoa erroreen eta desbideratzeen diagnosi osoak egin ahal izateko, hau da, ez da nahikoa erroreak eta desbideratzeak hizkuntzen i(r)a kasuntzaren ikuspegitik tratatzeko: EDBL datu-basean ez da aurreikusten, esaterako, gordetako adibide bakoitzari legokiokeen hizkuntza-maila eta ondorioz, datu-base honekin ezingo dugu maila bakoitzari egokitutako ortografia-zuzentzailerik garatu. Hori egin ahal izateko, diseinua birplanteatu beharko litzateke. Helburuak desberdinak izanik, ordea, euskara-maila neurtzeari begira dagoen datu-base batean bestelako beharrak sortuko lirateke eta dagoenarekin bateraezina gertatuko litzateke azkenean. Ondorioz, egokiagoa litzateke aparteko datu-base bat egitea, eta bertan gehitzea hitz erroredun bakoitzaren kodea, dagokion hizkuntza-maila, sakoneko kodea, etab. zehazteko eremuak. Lan hori, baina, ez da berehalakoa.

VII.8. Laburbilpena

Erroreen eskuzko etiketatzea egin ondoren, determinatzaile-errore batzuen detekzio automatikoa bideratu dugu, MG formalismoan oinarritutako erregelen bidez eta euskararako analizatzaile sintaktikoaren analisia abiapuntutzat hartuta. Zehazki, determinatzaile-erroreak detektatzeko erregelak hirugarren mailan desanbiguatutako analisisetan oinarrituta idatzi ditugu, XUXENg gramatika-zuzentzailerako hartu zen irizpideari jarraituz.

Baliatu dugun formalismoari dagokionez, testuetan erroreen hasierak eta bukaerak zehazteko eta errore-etiketak esleitzeko baliagarria izateaz gain, MGk diagnosi-mezu egokiak idazteko aukera eskaintzen du, errore bakoitza zein motatakoa den zehazteko aukera. Eta erroreen detekzioarekin batera diagnosi linguistikoa egitea ekarpen garrantzitsua da aztergai ditugun bi ikerketa-alorretan: euskarazko erroreen tratamendu automatikoan eta ordenagailuz lagundutako hizkuntzen i(r)a kasuntzan. Erroreen diagnosi linguistikoa egiteko, idatzi ditugun erregelak euskarazko erroreen eta desbideratzeen sailkapenari lotuta daude, hau da, automatikoki esleitutako etiketa bakoitza sailkapeneko kategoria edota azpikategoria linguistiko bati dagokio.

Determinatzaile-erroreen detekzio automatikoan hainbat muga edo arazo aurkitu ditugu. Errore mota jakin baten azterketa eta tratamendu automatikoa egitean,

bestelako erroreak ere egon ohi dira corpus errealean: errore ortografikoak, errore lexikoak, hitz-segida ez-gramatikalak... Errore mota horiek guztiak oztopo dira tratatu nahi diren beste errore batzuen detekzio automatikoa bideratzeko, ‘zarata’ sortzen baitute. Errore ortografiko eta lexiko asko, esaterako, ezezagunak zaizkio analizatzaile sintaktikoari. *Parse*rrak lexikorik gabeko hitz gisa analizatzen ditu ezagutzen ez dituen forma horiek, eta EZEZAG etiketa esleitzen die. Ondorioz, detektatu nahi diren erroreak detektatzen dira, edo ez, edo alarma faltsuak sortzen dira, EZEZAG etiketa duten hitzen analisisen eta erregeletan zehaztutako informazioaren arabera. Beraz, gramatika-erroreen detekzioa egin aurretik errore ortografikoak, lexikoak edota hitz-segida ez-gramatikalak zuzendu behar direla uste dugu, gramatika-erroreen detekzio automatikoan emaitza hobeak lortuko bailirateke horrela. Badakigu, ordea, errore ortografiko eta lexiko batzuk automatikoki zuzentzea posible den arren, hainbat errore lexikoren eta hitz-segida ez-gramatikalen zuzenketa automatikoa ezinezkoa dela.

Bestalde, erroreen detekzio automatikorako baliatzen ditugun bitartekoak hizkuntza-ekoizpen gramatikalak aztertzeke garatu diren sistemak dira. Ondorioz, testu erroredunetan ez dizkigute egitura erroredunei dagozkien analisiak ematen: adibidez, *auzi hau* sintagman ongi jartzen dira sintagma mugatzen duten %SIH eta %SIB etiketak, baina **auzia hau* sintagma erroredunean ez, *auzia* sintagmatzat hartzen baitu analizatzaileak (%SINT etiketa esleituz). Arazo hori konpontzeko bide bat hizkuntza-ekoizpen ez-gramatikalak aztertzeke baliabideak garatzea litzateke, erroreen detekzioa helburu duen analizatzaile sintaktiko bat sortzea. Lan hori oso garestia da, ordea, eta orain artean ez dugu eskura izan horretarako bitartekorik. Ondorioz, euskara gramatikala eta estandarra aztertzeke sortu dugun analizatzaile sintaktikoak ematen digun informazioa oinarri hartuta bideratu dugu euskarazko erroreen tratamendu automatikoa.

Gainera, erregelak egiteko baliatzen ditugun analisisen aurreko urratsetan akatsak gertatu ahal dira: posible da, esaterako, desanbiguatzaileak hitzei testuinguru jakin batean ez dagokien analisisa esleitzea. Horrek ere eragina dauka, noski, erroreen detekzioan: hitzek testuinguru jakin batean dagokien analisisa ez badute, erroreak detektatu gabe gera daitezke edota alarma faltsuak sor daitezke.

Erroreen tratamendu automatikoaren alorrean oztopo da, halaber, hizkuntzaren beraren konplexutasuna edota anbiguotasuna. Badaude esaldi-mailan zuzentzat hartu arren testu-mailan erroredunak diren egiturak edota sintagma-mailan erroreduntzat hartu arren esaldi-mailan zuzenak izan daitezkeenak. Baina errore mota horiek automatikoki tratatzea zaila da oraindik, eta euskararen azterketa eta prozesamenduaren alorrean hainbat ekarpen egin beharra dago horrelako erroreen detekzioa bideratu ahal izateko (testu-mailako azterketa automatikoa, galdera-erantzunen sistemen garapena, anaforaren tratamendua... beharrezkoak izango dira, besteak beste).

Eskuz determinatzaile-errore mota guztiak etiketatu diren arren, detekzio automatikoari dagokion lehen hurbilpen honetan DETK (determinatzailea kendu), DETG (determinatzaile atzizkia gehitu), DETE (determinatzailea errepikatu), DETO (determinatzaile-sintagmako osagaien ordena okerra) eta DETMUG (mugatasun kontuak) azpikategorietako errore batzuk detektatzeko erregelak egin ditugu: agerpen gehien dituzten errore motak, hain zuzen. Bigarren urrats batean helduko diogu aztertu gabe geratu diren gainerako erroreen detekzio automatikoari.

Orain artean, 85 erregela idatzi ditugu, MG formalismoan oinarrituta. Erregelen ebaluaziorako baliatu dugun euskara-ikasleen probarako corpusean, erregelen doitasuna % 45,45koa izan da eta estaldura % 44,77koa. Doitasuna, ordea, % 80ra igotzen da, kontuan hartzen ez badira erroreak detektatzeko erregelai ez dagozkien beste arazo batzuen ondorioz sortu diren alarma faltsuak (% 79,17), hala nola, errore ortografikoak, lexikoak edota hitz-segida ez-gramatikalak, analizatzailearen aurreko urratsetan gertatu ahal izan diren akatsak... Baina horrelako batzuk ekiditea posible den arren, beste batzuk saihestea oso zaila da.

Euskara-ikasle ez diren hiztunen testuetan ere ebaluatu ditugu erregelak, egiazko positiboen (ongi detektatutako erroreen) eta gezurrezko positiboen (alarma faltsuen) kopuruak ezagutzeko. Espezialitate-euskarako ikasleen testuak eta *Euskaldunon Egunkaria* corpusaren lagin bat baliatu ditugu horretarako. Testu horietan oso adibide gutxi detektatu dira: % 0,62 eta % 0,33, hurrenez hurren (corpus-laginetako sintagma kopuruarekiko). Ebaluazio horretan ikusi dugu euskara-ikasle ez diren hiztunek oso determinatzaile-errore gutxi egiten dituztela eta alarma faltsuen kopurua baxua dela. Gure ustez, egokiagoa da errore mota gutxiago detektatzea, alarma

faltsuen kopurua areagotzea baino; izan ere, alarma faltsuek erabiltzaileen jakintza zalantzan jar dezakete (hizkuntza-ikasleen kasuan, batik bat), eta tresnek eraginkortasuna galduko lukete alarma faltsu asko erakutsiz gero.

MG bidez detektatzen ditugun determinatzaile-erroreez gain, hitzak deklinatzean determinatzaile atzizkia ongi ez erabiltzeagatik sortzen diren erroreak ere etiketatu dira eskuz, DETK_DEKL eta DETG_DEKL etiketekin. Hitz-mailako erroreak dira horiek eta horrelako adibideak automatikoki detektatzeko EDBL datu-basean deklinabide-atzizki erroredun berriak gehitu ditugu. Baina EDBL ez da forma erroredunak tratatzeko helburuarekin garatutako datu-basea eta, beraz, ez da baliabide nahikoa erroreak eta desbideratzeak hizkuntzen i(ra)kasuntzaren ikuspegitik tratatzeko. Izan ere, EDBLk ez du erroreen diagnosi oso bat egiteko aukerarik ematen eta datu-basean gordetako informazioarekin ezingo dugu, esaterako, hizkuntza-maila bakoitzari egokitutako ortografia-zuzentzailerik garatu. Horretarako, hitz erroredun bakoitza zein mailari dagokion zehazteko aukera eskaintzen duen beste datu-base lexikal bat osatu beharko litzateke. Lan hori oso garestia da, ordea.

Tesi-lan honetan aztertu ditugun determinatzaile-erroreak detektatzeko erregelak euskararako garatzen ari garen XUXENg gramatika-zuzentzailean (Otegi, 2003 eta 2006) integratzeko asmoa dago. Horrez gain, erregela bakoitzean zehaztutako informazioa (erroreari dagokion sailkapeneko kategoria edo azpikategoria, errorearen deskripzioa, errore motaren adibideak, zuzenketa posibleak...) baliagarria izango da euskararen i(ra)kaskuntzaren alorrean, ordenagailu bidez euskara ikasteko garatzen ditugun laguntza-tresnetan ohar edota informazio gisa erabiltzeko. Izan ere, erroreen azterketa eta prozesamendua bi ikerketa-alorretarako egiten dugu: erroreen detekzio automatikoa eta ordenagailuz lagundutako hizkuntzen i(ra)kaskuntza. Eta bi ikuspuntuak uztartzea izan da gure helburuetako bat.

Tesi honetan erroreen eta desbideratzeen analisirako hainbat irizpide finkatu eta baliabide batzuk aurkeztu ditugu. Izan bedi lan hau euskarazko erroreen eta desbideratzeen azterketaren eta prozesamenduaren alorrean ikertzen, aurrerapausoak ematen eta ekarpenak egiten jarraitzeko informazio- eta motibazio-iturri.

VIII. Ondorioak

VIII.1. Ekarpen eta ondorio nagusiak

Tesi-lan honen xede nagusietako bat izan da euskarazko erroreen eta desbideratzeen analisirako irizpide batzuk finkatu eta lan-ingurune oso bat garatzea, bi ikerketa-alorretan ekarpenak egiten jarraitzeko: batetik, erroreen azterketa eta prozesamenduaren alorrean, euskararako ortografia-, gramatika- eta estilo-zuzentzailea edota aldaki dialektalen markatzailea bezalako tresnak garatzea dugu helburu, eta tesi honetan egindako determinatzaile-erroreen analisisa eta tratamendu automatikoa ekarpen garrantzitsua da euskararako gramatika-zuzentzailea osatze bidean; bestetik, ordenagailuz lagundutako hizkuntzen i(ra)kaskuntzaren alorrean, euskararen ikasketa prozesua ezagutu ahal izateko baliagarriak diren bitartekoak eta hizkuntza i(ra)kasteko, hobetzeko edota lantzeko laguntza-tresnak prestatzea da gure erronketako bat, eta horretarako hainbat tresna diseinatu ditugu jada. Alor batean zein bestean helburuak desberdinak diren arren, bietan da ezinbestekoa erroreen eta desbideratzeen analisi sakonak egitea. Eta horretarako aukera eskaintzen digun lan-ingurune oso bat garatu dugu guk. Bi

ikuspuntuak uztartuz, alor batean egiten dugun lana baliagarria zaigu besterako, eta alderantziz.

Euskarazko erroreen eta desbideratzeen analisia bideratzeko garaian, garrantzitsua iruditu zaigu euskarak bizi duen egoera soziolinguistikoa kontuan hartzea: euskarak estandarizazio-prozesuan murgilduta jarraitzen du 1968an euskara idatziaren batasuna bideratzeko ahaleginak hasi zirenetik, eta hizkuntzaren normalizazioa ez da oraindik erabatekoa. Gainera, euskara batuarekin hainbat euskalki edo dialekto bizi dira gure hizkuntza-komunitatean. Kode, estilo eta erregistro guztiak dira baliagarriak eta beharrezkoak, guztiak izan daitezke zuzenak eta egokiak komunikazio-egoeraren edota eremu geografikoaren arabera: testuinguru formaletan kode estandarrari eta estilo zainduari dagozkion ezaugarriak bete behar diren arren, egoera informaletan estilo zaindugabea edota kode dialektalak hobetsi behar dira. Eta hori kontuan hartu beharra dugu Hizkuntzaren Azterketa eta Prozesamenduaren (HAP) alorrean garatuko ditugun tresnak baliagarriak izatea nahi baldin badugu. Horregatik, kode estandarra helburu duten baliabide informatikoetatik haratago, gure hizkuntza-komunitateko egoerari eta hiztunen beharrei erantzungo dieten tresnak prestatzeko beharra aldarrikatzen dugu. Testuinguru honetan, bada, erroreak eta desbideratzeak bereizi ditugu.

Errore eta desbideratze kontzeptua

Erroreen analisiaren eta tratamendu automatikoaren alorrean erroretzat zer hartu eta zer ez finkatu beharra dago eta erroreen aurrean nola jokatu den erabaki behar da. Literaturan, oro har, ez dago guztiontzat baliagarria den eredu edo definizio estandarrik; bakoitzak dituen helburuen arabera finkatu ohi ditu erroreen analisirako baliagarriak zaizkion irizpideak. Euskara estandarrari dagozkion gramatika-arauak betetzen ez dituzten egitura erroredunak hartzen ditugu guk erroretzat; eta desbideratze esaten diegu gramatikalak izan arren testuinguru jakin batean desagokiak diren egiturei. Desbideratzeen barruan sartzen ditugu estilo kontu orokorrak, aldaki dialektalen erabilera edota erregistroaren egokitasuna.

Erroreak eta desbideratzeak bereizi eta bereiz tratatuko dituzten tresnak sortu behar direla uste dugu. Erroreen artean, ortografia-zuzentzaileari dagokio arau ortografikoak urratzen dituzten hitzen detekzioa eta zuzenketa; eta gramatika-zuzentzaileak detektatuko ditu euskarari dagozkion gramatika- eta

puntuazio-arauak betetzen ez dituzten egitura erroredunak. Desbideratzeen barruan, berriz, estilo-zuzentzaileak detektatuko ditu gramatikaren ikuspuntutik zuzenak izan arren estiloari dagokionean hobetu daitezkeen egiturak: behin eta berriz errepikatzen diren hitzak, esaldi luzeegiak, hitzen ordena desegokia...; estilo-zuzentzaileak tratatu beharrekoak dira, halaber, errore semantiko-pragmatikoak; aldaki dialektalen markatzaileak komunikazio-egoera edota eremu geografiko jakinei dagozkien egiturak ezagutu eta forma estandarretatik bereiziko ditu; eta erregistroaren egokitasuna kontuan hartzen duten baliabideak ere garatu beharko lirateke aurrerantzean.

Erroreak eta desbideratzeak bereizi eta bereiz tratatuko dituzten tresnak sortzea, beraz, ekarpen garrantzitsua da gure hizkuntza-komunitateko beharrak ase nahi baditugu. Horretarako, hainbat bitartekok eratzen duten lan-ingurune oso bat garatu dugu.

Erroreen eta desbideratzeen analisirako lan-ingurunea

Euskarazko erroreen eta desbideratzeen analisirako lan-ingurunean bi ikuspuntu edota ikerlerro uztartzen ditugu: erroreen tratamendu automatikoa eta ordenagailuz lagundutako hizkuntzen i(ra)kaskuntza. Erroreen analisi bera bi ikerketa-alorretan baliatzea ez da ohikoa izaten, eta guk ekarpen interesgarria egin dugula uste dugu, adar batean egiten dugun lanaren zati bat baliagarria baitzaigu besterako, eta alderantziz.

Aipatu lan-ingurunea bitarteko nagusi hauek osatzen dute: corpusek, corpusetan dauden adibideak etiketatzeko editoreak, euskarazko erroreen eta desbideratzeen sailkapenak eta etiketatutako adibideak gordetzeko sortu ditugun *Erroreak* eta *Ikasleak* datu-baseek. Bitarteko horiek guztiak loturik daudenez, erroreen eta desbideratzeen analisi osoak egiteko aukera daukagu, etiketatutako adibideekin eta horiei dagokien informazio linguistikoarekin batera datu teknikoak eta psikolinguistikoak ere gordetzen baititugu.

Corpusa

Corpusa da aztertu beharreko adibideen iturburu nagusia, erroreen eta desbideratzeen alorrean egiten ditugun azterketen abiapuntua. Hiztun desberdinen testuak baliatzen ditugu: batetik, hizkuntza-ikasleenak, horietan egon ohi delako

errore-adibide gehien eta ikasleen ekoizpenak ezinbestekoak direlako hizkuntzaren ikasketa prozesua aztertu ahal izateko; eta bestetik, hiztun arruntenak, hiztunok ere tarteka egiten ditugulako erroreak. Erabiltzaile jakin batzuentzako tresnak garatuko baditugu, garrantzitsua da erabiltzaile horien ekoizpen errealak abiapuntutzat hartzea.

Hala, euskarazko erroreen eta desbideratzeen analisirako osatu dugun euskara-ikasleen testu-bilduma da tesi-lan honen beste ekarpenetako bat: 113.290 hitzeko corpora bildu dugu behe-, erdi- eta goi-mailako euskara-ikasleen testuekin. Corpus hori determinatzaile-erroreekin etiketatuta dago, eta lagin txikiagoetan bestelako errore batzuen etiketatzea ere egin da (komunztadura-erroreak, mendeko perpaus erroredunak...). Izan ere, informazio linguistikoz hornitutako testuak baliabide eraginkorrak dira bai erroreen tratamendu automatikoa bideratzeko eta bai hizkuntza-ikasleen i(ra)kaskuntza prozesuari dagozkion ezaugarriak edota urratsak aztertzeke. Horregatik, hain zuzen, denetariko errore eta desbideratze motekin etiketatutako corpus sendo bat eratzea da gure erronketako bat: euskarazko erroreen eta desbideratzeen azterketa eta prozesamenduaren alorrean ekarpenak egiten jarraitzeko aukerak eskainiko dizkigun corpora eratu nahiko genuke etorkizunean. Euskara-ikasleen corpusarekin batera, baina, bestelako hiztunen testuak ere aztertu behar dira. Hala, espezialitate-euskarako ikasleen lanak (Euskara Teknikoko ikasleenak) eta hiztun arruntek idatzitako hainbat testu baliatzen ditugu, erroreak ongi detektatzen diren ikusteko, alarma faltsurik sortu bada konpontzeko edota tratamendu automatikorako baliatu diren teknikak ebaluatzeke.

Hizkuntza-ikasleen corpora biltzeko aplikazioa eta kode-sistema

Euskara-ikasleen corpusari dagokionez, IRAKAZI aplikazioak testuak bildu, antolatu eta modu errazean identifikatzeko aukera eskaintzen digu. Euskaltegietako ikasleen testuak gordetzeko kode-sistema bat definitu dugu, lau eremuk osatzen dutena: testuaren jatorriak (euskaltegiak) eta urteak; hizkuntza-mailak; ikaslearen identifikazio-hizkiek eta ariketa motak. Horrela, fitxategiaren izenak informazio zehatza esplizituki adierazten digu, eta erreferentzia-kodeari esker erraz jakin dezakegu, esaterako, zein euskaltegitako testuak dauzkagun, euskaltegi eta maila bakoitzeko zenbat testu ditugun, ikasle beraren zenbat lan dauzkagun edota ariketa bera zenbat ikaslek egin duten. Ordenagailuz lagundutako hizkuntzen i(ra)kaskuntzaren alorrean, garrantzitsua da corpora irizpide jakin batzuen arabera

antolatzea, horrek bermatuko baitu, hein handi batean, ikerketaren arrakasta edo porrota. Euskara-ikasleen corpus sendo eta adierazgarri bat osatzen jarraitzeko irizpideak definitzea izan da, hain zuzen, tesi-lan honen beste ekarpen bat. Irizpideak definituta, hizkuntza-ikasleen corpus ahalik eta osatuena, handiena eta heterogeneoena (euskaltegi, ikasle, maila edota zonalde desberdinetako ikasleena) biltzen jarraitzea da gure xedea.

Corpusak hizkuntza-azterketak egiteko baliatu nahi badira, testuak informazio linguistikoarekin hornitu behar dira, hau da, testuak etiketatu egin behar dira.

Errore-editorea

Eskuzko etiketatzea lan handia eta garestia izan ohi da. Euskarazko errorearen eta desbideratzearen etiketatzea errazteko, EtikErro editorea garatu dugu. Editoreak erroreak etiketatzeko laguntza emateaz gain, aukera eskaintzen digu etiketatutako adibideak eta horiei dagokien informazio linguistikoa euskarazko erroreak eta desbideratzeak biltegiratzeko garatu dugun datu-basera esportatzeko. Informazio linguistiko hori datu teknikoekin edota psikolinguistikoekin osatzen dugu ondoren. Eta horixe da, hain zuzen, gure editoreak alor honetan garatu diren editoreekiko duen funtsezko desberdintasun bat; izan ere, EtikErro ez da testuen etiketatze-lanera mugatzen, gainontzeko editoreen moduan; aitzitik, errore-adibideekin eta informazio linguistikoarekin datu-baseak aberasteko ere balio du, eta horrek datu-bilketari dagokion lana errazten eta aurrezten digu.

Sailkapena

Corpusetan detektatutako adibideak etiketatzeko (adibideen diagnosi linguistikoa egiteko), errore-sailkapenetan zehaztutako kategoria edota azpikategoria linguistikoak hartu ohi dira oinarri. Horregatik, sailkapenak oinarrizko baliabideak dira errore-analisiaren alorrean. Horrelako taxonomiak eratzea ez da lan erraza izaten, ordea, eta bakoitzak dituen helburuen arabera errore-kategoriak definitu ohi dira. Ondorioz, subjektibotasuna handia izan ohi da, bai errorearen sailkapenak osatzen direnean, eta bai errore-adibideak sailkatzen direnean ere. Hori dela eta, garrantzitsua da sailkapeneko kategorien eta azpikategorien definizioak eta errore- edota desbideratze-adibideak etiketatzeko irizpideak argi eta garbi zehaztea. Funtsezkoa da, halaber, testuetan detektatutako errore mota errealak barne hartzen dituzten sailkapenak garatzea. Gainera, sailkapen irekiak eta dinamikoak egitea da egokiena,

corpusetan adibide berriak detektatu ahala azpikategoriak gehitzen joateko aukera ematen baitigu horrek.

Eskura ditugun bitartekoekin errore mota guztiak aldi berean aztertzea ezinezkoa zaigunez, tesi-lan honetan euskarazko errorearen eta desbideratzearen sailkapenaren egitura nagusia definitu dugu eta determinatzaile-erroreei dagokien kategoria landu dugu sakonki.

Datu-baseak

Sailkatutako adibideak eta horiei dagokien informazioa biltegitratzeko bi datu-base prestatu ditugu: *Erroreak* eta *Ikasleak*. Lehenengoaren helburua errorearen tratamendu automatikoa bideratzeko baliagarria den informazioa jasotzea da. Horretarako, etiketatutako adibide bakoitzari dagokion informazio linguistikoa eta teknikoa gordetzen da. Bigarrenaren xede nagusia hizkuntzen i(ra)kaskuntzaren prozesua aztertzeko eta horren araberrako laguntza-tresnak garatzeko behar den informazioa jasotzea da. Datu-base horretan, beraz, testuetan detektatutako adibideei eta testu-ekoizleei dagozkien datu linguistikoak eta psikolinguistikoak gordetzen ditugu.

Datu-base bakoitzaren helburuak desberdinak diren arren, konektaturik daude eta biltegitratutako adibideei dagokien informazio linguistikoa partekatzen dute. Datu-base bien integrazioa ekarpen garrantzitsua izan da euskarazko errorearen eta desbideratzearen tratamendurako, testuetan etiketatutako adibideak eta horiei dagokien informazio linguistikoa bi ikerketa-alorretan baliatzeko aukera eskaintzen baitigu. Adibide bakoitzari dagokion informazio linguistikoarekin batera, datu teknikoak edota psikolinguistikoak ere jasotzen ditugu, eta errorearen eta desbideratzearen ikuspegi zabalago bat lortzen dugu horrela. Aipatu datu-baseetan gordetako informazioari esker, bada, euskarazko errorearen eta desbideratzearen diagnosi linguistikoa, teknikoa edota psikolinguistikoa egiteko aukera daukagu.

Aztertu beharreko adibideak, bestalde, bi euskarritan jasotzen ditugu: datu-baseetan, adibide solteak; eta corpusetan, adibideak testuinguruan. Adibide solteak baliatzen ditugu errorearen tratamendu automatikorako erregelak egiteko; testu osoak behar ditugu, ordea, bai testuingurua kontuan hartuz aztertu behar diren desbideratzearen tratamendurako, eta bai hizkuntza-ikasleen ibilbidea ezagutu ahal

izateko; izan ere, kasu horietan ezinbestekoa da errore- eta desbideratze-adibideekin batera, egitura zuzenak, ekiditen direnak edota ikasleei dagokien informazio psikolinguistikoa ere kontuan izatea.

Corpusak, errore-editorea, euskarazko errorearen eta desbideratzearen sailkapena eta *Erroreak* eta *Ikasleak* datu-baseak funtsezko bitartekoak zaizkigu ikusmiran ditugun ikerketa-alorretan aurrerapausoak ematen jarraitzeko, eta guztiak lan-ingurune batean biltzea HAPen alorrean egin dugun ekarpen garrantzitsua da.

Determinatzaile-erroreen eskuzko etiketatzea

Erroreen eta desbideratzearen analisirako bitartekoak prestatuta, kasu-azterketari ekin diogu. Eskuzko etiketatzea ezinbesteko urratsa da errorearen azterketa eta prozesamenduaren alorrean, etiketatutako adibideak oinarri hartuta bideratzen baita, ondoren, tratamendu automatikoa.

Corpusetan errore eta desbideratze mota guztiak aldi berean detektatu, etiketatu, sailkatu eta aztertzea litzateke egokiena. Eskura ditugun baliabideekin, ordea, ezinezkoa izan zaigu hori egitea eta tesi-lan honetan determinatzaile-erroreen eskuzko azterketa eta eskuz etiketatutako errore batzuen tratamendu automatikoa bideratu ditugu. Hala ere, lan honetan determinatzaile-erroreen analisira mugatu garen arren, osatu dugun lan-inguruneak errore eta desbideratze mota guztiak tratatzeko aukera ematen digu.

Testuetan etiketatutako determinatzaile-erroreen kopurua ez da oso handia: % 2,07 euskara-ikasleen testuetan eta % 0,98 Euskara Teknikoko ikasleen lanetan (corpusetako IS kopuruarekiko). Baina hizkuntza-maila eta testu mota guztietan gramatika-errore larritzat har daitezkeen erroreak dira eta, hortaz, euskararako gramatika-zuzentzaileak detektatu beharrekoak. Erroreen maiztasunagatik baino, beraz, larritasunagatik hautatu dugu fenomeno linguistiko erroredun horren analisisa egitea. Eta horrela, aurrerapauso berri bat eman dugu IXA taldean garatzen ari garen euskararako gramatika-zuzentzailea osatze bidean.

Erroreen detekzio automatikorako ezinbestekoa izateaz gain, errorearen etiketatzea oinarritzko abiapuntua da hizkuntza-ikasleen ibilbidea ezagutu ahal izateko. Etiketatutako errore-adibideei esker ikusi dugu euskara-ikasleen kasuan determinatzaile-erroreen kopurua gutxituz doala hizkuntza-mailan aurrera egin ahala:

behe-mailako testuetan errorearen kopurua % 2,73 izatetik goi-mailakoetan % 1,81 izatera pasatzen da. Bestalde, determinatzailea kentzea (DETK) eta errepikatzea (DETE) dira, oro har, testu-mota eta hizkuntza-maila guztietan gehien etiketatu diren errore motak. Hizkuntzaren ikasketa-prozesua aztertu ahal izateko, baina, beharrezkoa da denetariko errore motak etiketatzea, eta erroreekin batera desbideratzeak eta egitura zuzenak ere kontuan hartzea. Azterketa mota hori lan honetan egin ez den arren, horretarako baliabideak prest daude tesian aurkeztu dugun lan-ingurunean.

Determinatzaile-erroreen tratamendu automatikoa

HAP, oro har, eta errorearen tratamendu automatikoa, zehazkiago, ikerketa-alor konplexuak dira eta euskara bezalako hizkuntza eranskarietan hizkuntzaren tratamendu automatikoak hainbat zailtasun ditu. Azken urteotan aurrerapenak egin diren arren, egiteko asko dugu oraindik esku artean.

Ez dago errore mota guztien detekzioarako baliagarria den teknirik; aitzitik, errore motaren arabera, bide bat edo beste jarraitzea komeni izaten da: esaldi mailakoak diren komuntadura-erroreak detektatzeko, adibidez, Saroi tresna garatu da IXA taldean; data-egituretan egiten diren erroreak XFST bidez detektatzen dira; eta postposizio-lokuzioetan egindako erroreak detektatzeko Murriztapen Gramatika (MG) formalismoan oinarritutako hainbat erregela idatzi dira (Oronoz, 2009). Determinatzaile-erroreak sintagma-mailako errore lokalak dira eta horien detekzio automatikoa bideratzeko erregelak ere MG baliatuz idatzi ditugu. Formalismo hori sintagma-mailako erroreak detektatzeko egokia da eta, gainera, aukera ematen digu detektatutako errorearen diagnosi linguistikoa egiteko, erregela bakoitza errorearen eta desbideratzearen sailkapeneko kategoria edota azpikategoria bati lotuta baitago.

MG formalismoa baliatuz eta euskararako garatu dugun analizatzaile sintaktikoaren analisiak oinarri hartuta, 85 erregela idatzi ditugu determinatzaile-errore batzuk automatikoki detektatzeko. Lehen hurbilpen honetan, ez dugu eskuz etiketatutako adibide erroredun guztien detekzioa bideratu, testuetan gehien etiketatutako errore-adibideena baizik: DETK (determinatzailea kendu), DETG (determinatzaile atzizkia gehitu), DETE (determinatzailea errepikatu), DETO (determinatzaile-sintagmako osagaien ordena okerra) eta DETMUG (mugatasun kontuak) kategoriako errore batzuen detekzio automatikoa egin dugu.

Idatzitako erregelen ebaluazioari dagokionez, euskara-ikasleen probarako corpusean erregelen doitasuna % 45,45koa izan da eta estaldura % 44,77koa. Baina doitasuna % 80ra igotzen dela ikusi dugu, determinatzaile-erroreak detektatzeko erregelei ez dagozkien zenbait arazo kontuan hartzen ez badira, hala nola, determinatzaile-erroreak ez diren gainerako erroreak (ortografikoak, lexikoak edota hitz-segida ez-gramatikalak), analizatzaileak aurreko urratsetan egindako akatsak, eskuzko lanean etiketatzaileak ahaztutako adibideak... Horrelako ebaluazioak baliagarriak dira, beraz, aurreko urratsetan gertatu ahal izan diren akatsak detektatu eta konpon daitezkeen aztertzeko, hau da, hizkuntzaren azterketa eta prozesamenduaren alorreko beste urrats batzuk errepasatzeko.

Determinatzaile-erroreak detektatzeko egin ditugun erregelen ebaluaziotik ondorioztatu dugu gramatika-erroreen detekzio automatikoa egin baino lehen komeni dela bestelako erroreak zuzentzea, errore horiek ‘zarata’ sortzen baitute. Errore ortografiko batzuen zuzenketa automatikoa bideragarria ikusten dugu eta horrek gramatika-erroreen detekzioa hobetuko luke. Ordea, beste errore batzuk automatikoki zuzentzea ezinezkoa da (errore lexiko batzuk edota ortografikoki zuzenak diren hitzek eratutako segida ez-gramatikalak, esaterako). Guk nahiago izan dugu errore gutxiago detektatzea, alarma faltsuen kopurua igotzea baino. Izan ere, alarma faltsuek erabiltzaileen jakintza zalantzan jar dezakete (hizkuntza-ikasleen kasuan, batik bat) eta tresnek eraginkortasuna galduko lukete alarma faltsu asko erakutsiko balituzte.

Erregelak idazteko baliatzen den desanbiguazio-mailak ere eragina du emaitzetan: errore mota batzuen kasuan desanbiguatutako gabeko analisiak baliatuz emaitza hobeak lortzen dira eta beste kasu batzuetan, berriz, desanbiguatutako analisiak erabiliz (Ornoz, 2009). Guk hirugarren mailan desanbiguatutako analisiak oinarri hartuta idatzi ditugu determinatzaile-erroreen detekzio automatikoa egiteko erregelak, XUXENG gramatika-zuzentzailerako hartu zen irizpideari jarraituz.

Bestalde, aztertu ditugun corpusak ez dira oso handiak eta oinarritzat hartu dugun errore-adibideen kopurua ere ez da oso handia. Testu gehiago aztertuz gero, beraz, determinatzaile-errore berriak lortuko genituzke seguruenik eta sailkatutako errore mota bakoitzari dagokion adibideen kopurua ere handitu egingo litzateke.

Erroreen eta desbideratzeen azterketa eta prozesamenduaren alorrean ikertzen jarraitzea ezinbestekoa zaigu euskararako hain beharrezkoak eta onuragarriak diren tresna (erdi)automatikoak garatzeko edota garatzen laguntzeko, teknologia berrien eragina gure hizkuntzan ahalik eta onena izan dadin. Garrantzitsua da, gainera, gure hizkuntza-komunitatean bizi dugun egoera soziolinguistikoa kontuan hartzea sortzen ditugun baliabideak eraginkorrak eta erabilgarriak izango badira. Eta horretarako bideak irekitzen dizkigu tesi-lan honetan aurkeztu dugun erroreen eta desbideratzeen analisirako lan-inguruneak.

VIII.2. Tesiarekin lotura duten lanak

Tesi-lan hau idatzi bitartean egin ditugun azterketek ondoko artikulua argitaratzeko bidea eman digute:

- Uria L., Arrieta B., Díaz de Ilarraza A., Maritxalar M., Oronoz M. 2009. ***Determiner errors in Basque: Analysis and Automatic Detection***. XXV. Congreso de la SEPLN (Donostia).
- Aldabe I., Arrieta B., Díaz de Ilarraza A., Maritxalar M., Niebla I., Oronoz M., Uria L. 2008. ***Hizkuntzaren Tratamendu Automatikoa Euskararen Irakaskuntzan***. BAT Soziolinguistika aldizkaria, 2008 (I), 66 zk., 61-69 or.; ISSN: 1130-8435.
- Aldabe I., Arrieta B., Díaz de Ilarraza A., Maritxalar M., Oronoz M., Uria L., Amoros L. 2007a. ***Learner and Error Corpora Based Computational Systems***. In Corpora and ICT in Language Studies: PALC 2005, J. Walinski, K. Kredens & S. Gozdz-Roszkowski (eds.), Peter Lang. Vol. 13, 2007. ISBN 978-3-631-56099-0.
- Aldabe I., Arrieta B., Díaz de Ilarraza A., Maritxalar M., Niebla I., Oronoz M., Uria L. 2007b. ***Lengoaia Natureleko Prozesamendurako Tresnen Integrazioa Hizkuntzen Ikasketarako Inguruneetan***. UPV/EHU, LSI, TR 14-2007, Donostia.
- Aldabe I., Arrieta B., Díaz de Ilarraza A., Maritxalar M., Niebla I., Oronoz M., Uria L. 2007c. ***Basque error corpora: a framework to classify and store it***. In the Proceedings of the 4th Corpus Linguistic Conference on-line archive: <http://www.corpus.bham.ac.uk/corplingproceedings07/>.
- Aldabe I., Arrieta B., Díaz de Ilarraza A., Maritxalar M., Niebla I., Oronoz M., Uria L. 2006. ***The Use of NLP tools for Basque in a multiple user CALL environment and its feedback***. TAL & ALAO workshop. TALN 2006. Leuven, Belgium. In Proceedings

of the 13th Conference Sur Le Traitement Automatique des Langues Naturelles. Volume 2. p.: 815-824; ISBN: 2-87463-024-1.

- Aldabe I., Arrieta B., Díaz de Ilarraza A., Gojenola K., Maritxalar M., Oronoz M., Otegi A., Uria L. 2005a. ***Erreus web aplikazioa***. UPV/EHU / LSI / TR 26-2005.
- Aldabe I., Arrieta B., Díaz de Ilarraza A., Maritxalar M., Oronoz M., Uria L., Leire Amoros. 2005b. ***IRAKAZI: a web-based system to assess the learning process of Basque language learners***. EuroCALL. Cracovia. Polonia.
- Aldabe I., Arrieta B., Díaz de Ilarraza A., Maritxalar M., Oronoz M., Uria L. 2005c. ***Propuesta de una clasificación general y dinámica para la definición de errores***. Revista de Psicodidáctica, EHU. Vol 10, Nº 2, p. 47-60 (ISSN: 1136-1034).
- Aldabe I., Aldezabal I., Aranzabe M., Arrieta B., Díaz de Ilarraza A., Gojenola K., Maritxalar M., Oronoz M., Otegi A., Uria L. 2005d. ***Euskarazko errorearen sailkapena ERROREAK eta DESBIDERATZEAK datu-baseetan***. UPV/EHU / LSI / TR 27-2005.

Artikulu eta barne-txosten horiek izan dira tesi-lan honen elikagai eta tesiak ere, aldi berean, euskarazko errorearen eta desbideratzeen azterketa eta prozesamendurako baliabide, helburu edota ideia berriak sustatzeko balio izan digu.

VIII.3. Etorkizuneko lanak

Euskarazko errorearen eta desbideratzeen azterketa eta prozesamenduaren alorrari dagokion lana konplexua da eta egiteko asko dugu oraindik esku artean. Tesi honetan oinarritzko irizpide batzuk finkatu eta lan-ingurune oso bat garatu dugu ikusmiran ditugun bi ikerketa-alorretan ekarpenak egiten jarraitzeko. Hemen aurkeztutako lana osatzeko, etorkizun hurbilean egin beharreko azterketa hauek aurreikusten ditugu:

- Lehen hurbilpen honetan landu ez ditugun determinatzaile-errorearen tratamendu automatikoa bideratzea, urrats honetan ikusi eta ikasitakoak kontuan hartuta.
- Erregelak XUXENG-n integratzea, euskararako gramatika-zuzentzaileak landu ditugun errore motak detekta ditzan.
- Determinatzaile-errorearen inguruan bildu dugun informazioa ordenagailuz lagundutako hizkuntzen i(r)a-kaskuntzaren alorrean garatu ditugun aplikazioetan integratzea; hau da, determinatzaile-errore bakoitzaren kategoriei

eta azpikategoriei dagozkien definizioak edota deskripzioak, adibideak, zuzenketak... idazketa-prozesurako edota idazlanen ebaluazio automatikorako prestatu ditugun aplikazioetan integratzea hizkuntza-ikasleei oharrak, aholkuak edota proposamenak eskaintzeko.

Hala ere, ikusmiran ditugun ikerketa-alorretan ate asko daude zabalik, tesi honetan aurkeztutako lanari jarraipena emateko. Interesgarria litzateke, beraz, euskarazko errorearen eta desbideratzearen analisiaren inguruan ekarpen berriak egiteko, beste azterketa hauek ere bideratzea:

- Euskaltegi gehiagoren lankidetzaren lortzea, hizkuntza-ikasleen testu berriak eskuratzeko eta ordenagailuz lagundutako hizkuntzen i(r)a(k)askuntzaren alorrean garatzen ditugun tresnak ebaluatzeko.
- Hizkuntzaren i(r)a(k)askuntzaren alorrean esperientzia duten irakasleekin lankidetzaren harremanak sortzea, euskara-ikasleen testu-azterketak egiteko eta ondorioak ateratzeko garaian beraien ikuspuntua eta esperientzia kontuan hartu ahal izateko. Izan ere, ordenagailuz lagundutako hizkuntzen i(r)a(k)askuntza diziplinarteko alorra da, hainbat adituren elkarlana eskatzen duena (hizkuntzalari, informatikari eta hizkuntza-irakasleena, gutxienez).
- Errore-etiketatzailerik gehiagorekin lan-talde bat osatzea, euskarazko errorearen eta desbideratzearen azterketa eta tratamendu automatikoari dagozkion aztergaiak talde-mailan argitzeko, zalantzazko kasuetan iritziak alderatzeko edota hartu beharreko erabakiak hartzeko.
- Corpus erroreduna osatzen jarraitzea: zenbat eta testu gehiago bildu, orduan eta oinarri sendoagoa izango dugu euskarazko errorearen eta desbideratzearen analisirako.
- Euskarazko errorearen eta desbideratzearen sailkapena osatzen duten azpikategoriak zehazten eta lantzen jarraitzea, corpusetan adibide berriak aurkitu ahala, errore mota gehiagoren diagnosi linguistikoa egin ahal izateko.
- Euskara-ikasleen corpusean errore eta desbideratze desberdinak etiketatzea, etiketatutako corpus adierazgarri bat eratze aldera.
- Detekzio eta diagnosi linguistikoaz gain, determinatzaile-errorearen zuzenketa automatikoari heltzea.

- Euskararen ikasketa-prozesua zein den ezagutzeko ikasle edota ikasle-talde baten ibilbidea aztertzea, tesi-lan honetan aurkeztu ditugun bitartekoak baliatuz.
- Forma estandarrak eta ez-estandarrak bereizten dituzten tresnak garatzea, hau da, aldaki dialektalak erroretzat hartu beharrean forma ez-estandartzat hartzen dituzten tresnak prestatzea. Horrela, erabiltzaileek jakin ahal izango dute aldaki dialektal gisa markatutako forma horiek testuinguru, eremu geografiko edota erregistro batzuetan desagokiak diren arren, beste batzuetan egokiak eta baliagarriak direla.
- Bizkaierarako XUXEN (XUXENb) egin den moduan, aldaki dialektal jakinen ezaugarriak zehaztu eta gordetzea, beste euskalki batzuetarako zuzentzaileak garatu ahal izateko.
- Hizkuntza-erregistroaren egokitasuna automatikoki tratatzeko tresnen garapena bideratzea.

Tesi-lan hau euskarazko erroreen eta desbideratzeen azterketaren eta prozesamenduaren alorrean ikertzen, aurrerapausoak ematen eta ekarpenak egiten jarraitzeko informazio- eta motibazio-iturri izatea gustatuko litzaiguke.

Eskerrak ematen

Eskerrak eman nahi dizkiegu euskarazko errorearen eta desbideratzearen azterketa eta prozesamenduaren inguruan idatzi dugun tesi-lan honetan euskara-ikasleen corpusa eratzeko, eratu dugun sailkapena ebaluatzeko, sortu zaizkigun zalantzak argitzeko, garatutako aplikazioak probatzeko... laguntza eskaini diguten hizkuntzalariei, informatikariei, hizkuntza-irakasleei eta euskaltegiei.

Erreferentziak

(Abney, 1987) Abney S. *The Noun Phrase in Its Sentential Aspect*. MIT PhD. Dissertation. Cambridge.

(Aduriz et al., 2006a) Aduriz I., Aranzabe M., Arriola J., Atutxa A., Díaz de Ilarraza A., Ezeiza N., Gojenola K., Oronoz M., Soroa A., Urizar R. *Methodology and steps towards the construction of EPEC, a corpus of written Basque tagged at morphological and syntactic levels for the automatic processing* *Corpus Linguistics Around the World*. Book series: Language and Computers. Vol 56 (pag 1- 15). ISBN 90-420-1836-4 Ed. Andrew Wilson, Paul Rayson, and Dawn Archer. Rodopi. Netherlands. 2006a.

(Aduriz et al., 2006b) Aduriz I., Aranzabe M., Arriola J., Díaz de Ilarraza A. *Sintaxi Partziala*. Fernández Beatriz & Laka Itziar (arg.), Andolin gogoan: Essays in Honour of Professor Eguzkitza, 31-49. Bilbo: UPV/EHU Argitarapen Zerbitzua. ISBN: 84-8373-890-2. 2006b.

(Aduriz et al., 2001) Aduriz, I., Arriola, J. M., Díaz de Ilarraza, A. *Desanbiguazio morfologikoa, azterketa sintaktikoaren lehen urratsak eta aplikazioak Murritzapen Gramatikaren eredu konputazionala jarraituz*. Euskaltzaindiaren XV. Biltzarra. Baiona, France. 2001.

(Aduriz, 2000) Aduriz I. *EUSMG: Morfologiatik syntaxira Murritzapen Gramatika erabiliz. Euskararen desanbiguazio morfologikoaren tratamendua eta azterketa sintaktikoaren lehen urratsak*. Doktoretza-tesia, Filologia eta Historia-Geografia Fakultatea. UPV-EHU, Gasteiz. 2000.

(Aduriz et al., 1998) Aduriz I., Agirre E., Aldezabal I., Alegria I., Ansa O., Arregi X., Arriola J.M., Artola X., Díaz de Ilarraza A., Ezeiza N., Gojenola K., Maritxalar M., Oronoz M., Sarasola K., Soroa A., Urizar R. *A framework for the automatic processing of Basque*. Proceedings of Workshop on Lexical Resources for Minority Languages, Granada, Spain. 1998.

- (Aduriz *et al.*, 1994) Aduriz I., Aldezabal I., Arriola J., Artola Zubillaga X. ***Euskararen normalizazioa eta linguistika konputazionala***. Euskaltzaindiaren XIII. Biltzarra / XIII. Congreso de Euskaltzaindia, 39. liburukia, 1579-1588. Leioa (Basque Country). 1994.
- (Aduriz, 1994) Aduriz I. ***Errore ortografikoen azterketa eta zuzenketa bi mailatako morfologiaren arabera***. UPV/EHU Euskal Filologia. Barne-txostena-doktoregoa. 1994.
- (Agirre *et al.*, 2005) Agirre E., Aldezabal I., Etxeberria J., Izagirre E., Mendizabal K., Pociello E., Quintian M. ***EUSEMCOR: euskarako corpora semantikoki etiketatzeko eskuliburua; editatze-, etiketatze- eta epaitze-lanak***. UPV/EHU, LSI, TR23-2005. 2005.
- (Agirre *et al.*, 1992) Agirre E., Alegria I., Arregi X., Artola X., Díaz de Ilarraza A., Maritxalar M., Sarasola K. ***Xuxen: A Spelling Checker/Corrector for Basque based in Two-Level Morphology***. Proceedings of ANLP'92, 119-125. Povo Trento. 1992.
- (Aït-Mokhtar eta Chanod, 1997) Aït-Mokhtar S. eta Chanod J.P. ***Incremental finite-state parsing***. Proceedings of the fifth conference on Applied Natural Language Processing, 72–79, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc. 1997.
- (Aldabe *et al.*, 2008) Aldabe I., Arrieta B., Díaz de Ilarraza A., Maritxalar M., Niebla I., Oronoz M., Uria L. ***Hizkuntzaren Tratamendu Automatikoa Euskararen Irakaskuntzan***. BAT Soziolinguistika aldizkaria, 2008 (I), 66 zk, 61-69, ISSN: 1130-8435. 2008.
- (Aldabe *et al.*, 2007a) Aldabe I., Arrieta B., Díaz de Ilarraza A., Maritxalar M., Oronoz M., Uria L., Amoros L. ***Learner and Error Corpora Based Computational Systems***. In Corpora and ICT in Language Studies: PALC 2005, J. Walinski, K. Kredens & S. Gozdz-Roszkowski (eds.), Peter Lang. Vol. 13, 2007. ISBN 978-3-631-56099-0. 2007a.
- (Aldabe *et al.*, 2007b) Aldabe I., Arrieta B., Díaz de Ilarraza A., Maritxalar M., Niebla I., Oronoz M., Uria L. ***Lengoaia Naturaleko Prozesamendurako***

- Tresnen Integrazioa Hizkuntzen Ikasketarako Inguruneetan*. Barne-txostena, UPV/EHU, LSI, TR14-2007, Donostia. 2007b.
- (Aldabe *et al.*, 2007c) Aldabe I., Arrieta B., Díaz de Ilarraza A., Maritxalar M., Niebla I., Oronoz M., Uria L. ***Basque error corpora: a framework to classify and store it***. In the Proceedings of the 4th Corpus Linguistic Conference on-line archive: <http://www.corpus.bham.ac.uk/corplingproceedings07/>. 2007c.
- (Aldabe *et al.*, 2006) Aldabe I., Arrieta B., Díaz de Ilarraza A., Maritxalar M., Niebla I., Oronoz M., Uria L. ***The Use of NLP tools for Basque in a multiple user CALL environment and its feedback***. TAL & ALAO workshop. TALN 2006. Leuven, Belgium. In Proceedings of the 13th Conference Sur Le Traitement Automatique des Langues Naturelles. Volume 2. p.: 815-824; ISBN: 2-87463-024-1. 2006.
- (Aldabe *et al.*, 2005a) Aldabe I., Arrieta B., Díaz de Ilarraza A., Gojenola K., Maritxalar M., Oronoz M., Otegi A., Uria L. ***Erreus web aplikazioa***. UPV/EHU / LSI / TR 26-2005. 2005a.
- (Aldabe *et al.*, 2005b) Aldabe I., Arrieta B., Díaz de Ilarraza A., Maritxalar M., Oronoz M., Uria L., Leire Amoros. ***IRAKAZI: a web-based system to assess the learning process of Basque language learners***. EuroCALL. Cracovia. Polonia. 2005b.
- (Aldabe *et al.*, 2005c) Aldabe I., Arrieta B., Díaz de Ilarraza A., Maritxalar M., Oronoz M., Uria L. ***Propuesta de una clasificación general y dinámica para la definición de errores***. Revista de Psicodidáctica, EHU. Vol 10, Nº 2, p. 47-60 (ISSN: 1136-1034). 2005c.
- (Aldabe *et al.*, 2005d) Aldabe I., Aldezabal I., Aranzabe M., Arrieta B., Díaz de Ilarraza A., Gojenola K., Maritxalar M., Oronoz M., Otegi A., Uria L. ***Euskarazko erroreent sailkapena ERROREAK eta DESBIDERATZEAK datu-baseetan***. Barne-txostena UPV/EHU / LSI / TR 27-2005, Euskal Herriko Unibertsitatea, Informatika Fakultatea, Donostia. 2005d.
- (Aldezabal *et al.*, 2007a) Aldezabal I., Ceberio K., Esparza I., Estarrona A., Etxeberria J., Quintian M., Izagirre E., Uria L. ***EPEC (Euskararen Prozesamendurako Erreferentzia Corpusa) segmentazio-mailan etiketatzeko eskuliburua***. UPV/EHU, LSI, TR11-2007.

- (Aldezabal *et al.*, 2007b) Aldezabal I., Aranzabe M., Arriola J.M., Díaz de Ilarraza A., Estarrona A., Fernandez K., Quintian M., Uria L. ***EPEC (Euskararen Prozesamendurako Erreferentzia Corpusa) dependentziekin etiketatzeko eskuliburua***. UPV/EHU, LSI, TR12-2007.
- (Aldezabal *et al.*, 2001) Aldezabal I., Ansa O., Arrieta B., Artola X., Ezeiza A., Hernández G., Lersundi M. ***EDBL: A General Lexical Basis for the Automatic Processing of Basque***. Proceedings of the IRCS Workshop on Linguistic Databases. Philadelphia (AEB). 2001.
- (Alegria *et al.*, 2006) Alegria I., Arrieta B., Díaz de Ilarraza A., Izagirre E., Maritxalar M. ***Using Machine Learning Techniques to Build a Comma Checker for Basque***. Coling-ACL. Sydney. Australia. ISBN: 1-932432-69-8 pp.1-8. 2006.
- (Alegria *et al.*, 2005) Alegria I., Artola X., Díaz de Ilarraza A., Ezeiza N., Sologaistoa A., Soroa A., Valverde A., Areta N., Gurrutxaga A., Leturia I., Saiz R. ***Zientzia eta teknologiaren corpusa. Euskara zientifiko-teknikoa: Normalizaziotik homologazitora***. Mendebalde Kultura Alkartea. ISBN: 84-931882-5-5. Bilbao. 2005.
- (Alegria *et al.*, 2004) Alegria I., Ansa O., Artola X., Ezeiza N., Gojenola K., Urizar R. ***Representation and Treatment of Multiword Expressions in Basque***. ACL workshop on Multiword Expressions. ISBN: 1-932432-40-X. 2004.
- (Alegria *et al.*, 2003) Alegria I., Aranzabe M., Ezeiza A., Ezeiza N., Urizar R. ***Robustez y flexibilidad de un lematizador/etiquetador***. VIII Simposio Internacional de Comunicación Social. ISBN 959-7174-01-4. 2003.
- (Alegria, 1995) Alegria I. ***Euskal morfologiaren tratamendu automatikorako tresnak***. Doktoretza-tesia, Euskal Herriko Unibertsitatea (UPV/EHU), Donostia. 1995.
- (Alexopoulou, 2005) Alexopoulou A. ***El error: un concepto clave en los estudios de adquisición de segundas lenguas***. RLA: Revista de lingüística teórica y aplicada. ISSN 0033-698X, Nº 43, 1, pags. 75-92. 2005.

- (Alonso, 2003) Alonso J.C. **Euskara eta hezkuntza**. Bat: Soziolinguistika aldizkaria. ISSN 1130-8435, N°. 49 (Ejemplar dedicado a: Hezkuntza mundua eta hizkuntza berreskurapena), pags. 147-156. 2003.
- (Aranzabe, 2008) Aranzabe M. **Dependentzia-ereduan oinarritutako baliabide sintaktikoak: zuhaitz-bankua eta gramatika konputazionala**. Doktoretza-tesia, Euskal Filologia Saila, EHU/UPV, Donostia. 2008.
- (Areta et al., 2008) Areta N., Gurrutxaga A., Leuria I. **Begiratu bat corpus-baliabideei**. BAT Soziolinguistika aldizkaria, 62. alea. 2008.
- (Armendariz et al., 2004) Armendariz A. J., López-Cuadrado J., Vadillo J.A. eta Pérez T. A. **HEUSKLEARNING: Un hiperentorno educativo para el aprendizaje del euskara**. IV Encuentro Europa-América Latina sobre Formación y Cooperación Tecnológica y Profesional, Isla de Margarita (Venezuela). 2004.
- (Arppe, 2000) Arppe A. **Developing a Grammar Checker for Swedish**. Proceedings from the 12th Nordiske datalingvistikkdager, Department of Linguistics, Norwegian University of Science and Technology (NTNU). Nordgard. December 9-10. 2000.
- (Arrieta et al., 2003) Arrieta B., Díaz de Ilaraza A., Gojenola K., Maritxalar M., Oronoz M. **A database system for storing second language learner corpora**. Learner corpora workshop. Corpus linguistics 2003. Lancaster, UK. Volume 16, Part 1. p.: 33-41; ISBN: 1 86220 131 5. 2003.
- (Arriola et al., 2005) Arriola J.M., Esparza I., Ezeiza N., Gojenola K., Sologaitoa A. **Analizatzaile morfosintaktikoa**. Barne-txostena, EHU/UPV, Donostia. 2005.
- (Artiagoitia, 2004) Artiagoitia X. **Izen Sintagmaren birziklatzea: IS-tik inguruko funtzio buruetara**. In Albizu P. eta Fernández B. (arg.) Euskal Gramatika XXI. mendearen atarian: arazko zaharrak, azterbide berriak. Arabako Foru Aldundia - EHU, 11-38. 2004.
- (Artiagoitia, 1998) Artiagoitia X. **Determinatzaile-sintagmaren hipotesia euskal gramatikan**. Uztaro 27, 33-61 orriak. 1998.
- (Artiagoitia, 1997) Artiagoitia X. **DP predicates in Basque**. In A. Taff (eds.), Working papers in Linguistics, Vol. 15, Seattle: University of Washington. 161-198. 1997.

- (Atkins *et al.*, 1992) Atkins B., Clear J., Ostler N. ***Corpus Design Criteria***. Literary and Linguistic Computing. Vol. 7, No. 1, 1-16. 1992.
- (Azkue, 1925) Azkue R.M. ***Morfología vasca***. La Gran Enciclopedia Vasca. Bilbo. 1969.
- (Badia *et al.*, 2004) Badia T., Gil A., Quixal M., Valentín O. ***NLP-enhanced error checking for Catalan unrestricted text***. Proceedings of the fourth international conference on Language Resources and Evaluation, LREC 2004, 1919–1922, Lisbon, Portugal. 2004.
- (Barrios *et al.*, 2008) Barrios M., Osa E., Muñoa I., Elorza I. eta Boan K. ***Zenbait orientabide erregistroen trataeraz***. Bilbo: Euskaltzaindia, Jagon Saila. 2008.
- (Beauvois, 1997) Beauvois M. H. ***High-Tech, High-Touch: From Discussion to Composition in the Networked Classroom***. Computer Assisted Language Learning, 10 (1), 57-69. 1997.
- (Becker *et al.*, 1999) Becker M., Bredenkamp A., Cysmann B., Klein J. ***Annotation of Error Types for German News Corpus***. In Proceedings of the ATALA workshop on Treebanks, Paris. 1999.
- (Behrens, 2008) Behrens H. ***Corpora in language acquisition research***. In Corpora in Language Acquisition Research; history, methods, perspectives. Edited by Heike Behrens, John Benjamins Publishing Company. Volume 6; p. xix. 2008.
- (Bell, 1974) Bell R.T. ***Error analysis: a recent pseudoprocedure in applied linguistics***. International Review of Applied Linguistics, Vols 25-26: 35-49. 1974.
- (Bibiloni, 1997) Bibiloni G. ***Llengua estàndard i variació lingüística***. Ed. Eliseu Climent, Valencia. 1997.
- (Birn, 2000) Birn J. ***Detecting grammar errors with Lingsoft's Swedish grammar checker***. In Proceedings of the 12th Nordic Conference in Computational Linguistics, Nodalida'99. Department of Linguistics, Norwegian University of Science and Technology (NTNU), Trondheim, pp. 28-40. 2000.
- (Bowker eta Pearson, 2002) Bowker L. eta Pearson J. ***Working with Specialized Language: A Practical Guide to Using Corpora***. London/New York: Routledge. 2002.

- (Brady *et al.*, 2005) Brady I, Navarro M., Perinán C. (Eds.). ***Nuevas tendencias en lingüística aplicada***. Murcia. 2005.
- (Burt eta Kiparsky, 1974) Burt M.K. eta Kiparsky C. ***Global and local mistakes***. In New Frontiers in Second Language Learning edited by J. H. Schumann and N. Stenson, Newbury House Publishers Inc. 1974.
- (Bustos, 1998) Bustos Gisbert J. M. ***Análisis de errores, problemas y categorización***. DICENDA Cuadernos de Filología Hispánica 16, Madrid, Servicio de publicaciones UCM, pp. 11-40. 1998.
- (Cabré, 1990) Cabré T. ***La lingüística aplicada avui***. In M^a Teresa Cabré *et al.* (eds.), *La lingüística aplicada: noves perspectives/noves professions/noves orientacions*. Fundació Caixa de Pensions, Universitat de Barcelona, 11-29. 1990.
- (Camacho eta Lonbide, 2005) Camacho A. eta Lonbide P. ***Euskararen irakaskuntza Europako Erreferentzia Markoaren baitan***. Hizpide 60, 69-82 orr. 2005.
- (Castro-Castro *et al.*, 2008) Castro-Castro D., Lannes-Losada R., Maritxalar M., Niebla I., Pérez-Marqués C., C. Alamo-Suarez N., Pons-Porrata A. ***A Multilingual Application for Automated Essay Scoring***. Lecture Notes in Advances in Artificial Intelligence - LNAI 5290 - IBERAMIA ISBN 3-540-99308-8 Springer New York pp. 243-251. 2008.
- (Cermeno, 2008) Cermeno O. ***Euskarazko errore sintaktikoen detekzioa ikasketa automatikoa erabiliz. Kasu praktikoa: determinatzaile eta komunztadura-erroreak***. Karrera Bukaerako Proiektua. Informatika Fakultatea, EHU, Donostia, 2008ko iraila.
- (Corder, 1974) Corder S.P. ***Idiosyncratic Dialects and Error Analysis***. In Richards, J. (Ed.). *Error analysis: Perspectives on Second Language Acquisition*. 158-171. Essex: Longman. 1974.
- (Corder, 1971) Corder S.P. ***Describing the language learner's language***. Interdisciplinary Approaches to Language. CILT Reports and Papers. 1971.
- (Corder, 1967) Corder S.P. ***The Significance of Learner's Errors***. *IRAL* (International Review of Applied Linguistics) 5, 161-170. 1967. Posteriormente,

- este artículo fue recogido en SP. Corder: *Error Analysis and Interlanguage* (Oxford: Oxford University Press, 1981).
- (Dagneaux *et al.*, 2005) Dagneaux E., Denness S., Granger S., Meunier F., Neff J., Thewissen J. ***Error Tagging Manual Version 1.2***. Centre for English Corpus Linguistics. Université Catholique de Louvain, Louvain-la-Neuve. 2005.
- (Dagneaux *et al.*, 1998) Dagneaux E., Denness S., Granger S. ***Computer-Aided Error Analysis***. System, Vol. 26, 163-174. 1998.
- (Dagneaux *et al.*, 1996) Dagneaux E., Denness S., Granger S., Meunier F. ***Error Tagging Manual Version 1.1***. Centre for English Corpus Linguistics. Université Catholique de Louvain, Louvain-la-Neuve. 1996.
- (Demuth, 2008) Demuth K. ***Exploiting corpora for language acquisition research***. In Corpora in Language Acquisition Research; history, methods, perspectives. Edited by Heike Behrens, John Benjamins Publishing Company. Volume 6, p. 200. 2008.
- (de Haan, 2000) de Haan P. ***Tagging non-native English with the TOSCA-ICLE tagger***. In Mair, C. and M. Hundt (eds.) *Corpus linguistics and linguistic theory*. 69-79. Amsterdam: Rodopi. 2000.
- (Díaz, 2005) Díaz A.M. ***Tipología de errores gramaticales para un corrector automático***. En Proceedings del XXI Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural, núm. 35, pp. 409-416. 2005.
- (Díaz de Ilarraza *et al.*, 2008) Díaz de Ilarraza A., Gojenola K., Oronoz M. ***Detecting Erroneous Uses of Complex Postpositions in an Agglutinative Language***. Proceedings of Coling-2008 Conference, Manchester. 2008.
- (Díaz de Ilarraza *et al.*, 2007) Díaz de Ilarraza A., Gojenola K., Oronoz M., Otaegi M., I. Alegria. ***Syntactic Error Detection and Correction in Date Expressions using Finite-State Transducers***. Workshop on Finite State Methods in Natural Language Processing, Postdam. 2007.
- (Díaz-Negrillo eta García-Cumbreras, 2007) Díaz-Negrillo A. eta García-Cumbreras M.A. ***A tagging tool for error analysis on learner corpora***. ICAME Journal, Computers in English Linguistics, Nº 31, April 2007.

- (Díaz-Negrillo eta Fernández-Domínguez, 2006) Díaz-Negrillo A. eta Fernández-Domínguez J. **Error Tagging Systems for Learner Corpora**. Revista Española de Lingüística Aplicada (RESLA), ISSN 0213-2028, Vol. 19, pags. 83-102. 2006.
- (Dulay *et al.*, 1982) Dulay H., Burt M., Krashen S. **Language two**. Oxford: Oxford University Press, New York. 1982.
- (Dulay eta Burt, 1974) Dulay H. eta Burt M. **You can't learn without goofing**. In Error analysis. ed. J. C. Richards. London: Longman. 1974.
- (D'Aquino eta Ribas, 2001) D'Aquino A. eta Ribas R. **Interferencias morfosintácticas entre italiano y español L2: un estudio empírico**. Atti del XIX Convegno [Associazione ispanisti italiani]: Roma, 16-18 settembre 1999 / coord. por Antonella Cancellier, Renata Londero, Vol. 2, 2001 (Italiano e spagnolo a contatto), ISBN 88-8098-146-3, pags. 119-130. 2001.
- (EAGLES, 1996) EAGLES. **Preliminary recommendations on corpus typology**. EAG-TCWG-CTYP/P. Pisa: Consiglio Nazionale delle Ricerche. Istituto di Linguistica Computazionale (<http://www.ilc.cnr.it/EAGLES96/corpintr/node13.html#SECTION00040000000000000000>). 1996.
- (Eguren, 2006) Eguren L. **Marcas de predicación en vasco**. In Beatriz Fernández e Itziar Laka (eds.), Andolin Gogoan. Essays in Honour of Professor Eguzkitza, Bilbao, Servicio Editorial de la UPV-EHU, pp. 233-250. 2006.
- (Elordui eta Zabala, 2009) Elordui A. eta Zabala I. **Euskara batuaren garapen lexiko-diskurtsiboa: batasunetik aniztasun funtzionalerantz**. In Etxepare R. Gomez R. Eta Lakarra J.A. (ed.). Beñat Oyharçabalen omenezko artikulu-bilduma. ASJU (argitarabidean).
- (Elordui, 2006) Elordui A. **Eskolaren hizkuntza-eredua harreman-sareen trinkotzearen ikuspegitik. Unibertsitate aurreko hizkuntza ereduaren balorazioa eta proposamenak**. BAT Soziolinguistika Aldizkaria 60, 141-160. 2006.
- (Elordui eta Maia, 2002) Elordui A. eta Maia J. **Eskola eta dialektoak: irakaskuntzan gertatzen diren arazoak eta irtenbideak**. Revista Psicodidáctica, nº 013, Euskal Herriko Unibertsitatea, Vitoria-Gasteiz. 2002.
- (Engwall, 1992) Engwall G. **Comments**. In Svartvik, J. (ed.). 164-169. 1992.

- (Euskaltzaindia, 2004) Euskaltzaindia. *Adierazpena: Euskalkien erabileraz irakaskuntzan, komunikabideetan eta administrazioan*. 137. araua: www.euskaltzaindia.net
- (Euskaltzaindia, 1993) Euskaltzaindia. *Euskal Gramatika Laburra: Perpaus bakuna*. Bilbo. 1993.
- (Euskaltzaindia, 1985) *Euskal gramatika: lehen urratsak-I*. Euskaltzaindia, Gramatika Batzordea, Iruñea, 1985 (berrargitalpena).
- (Evans) Evans D. *Corpus building and investigation for the Humanities: An on-line information pack about corpus investigation techniques for the Humanities* (<http://www.corpus.bham.ac.uk/Unit1.pdf>).
- (Ezeiza, 2002) Ezeiza N. *Corpusak ustiatzeko tresna linguistikoak. Euskararen etiketatzaile sintaktiko sendo eta malgua*. Doktoretza-tesia, University of the Basque Country, Donostia. 2002.
- (Ezeiza et al., 1998) Ezeiza N., Aduriz I., Alegria I., Arriola J.M., Urizar R. *Combining Stochastic and Rule-Based Methods for Disambiguation in Agglutinative Languages*. COLING-ACL'98. Pgs. 380 - 384. Vol 1. Montreal (Canada). August 10-14. 1998.
- (Fernández, 1997) Fernández S. *Interlengua y Análisis de Errores en el aprendizaje del español como lengua extranjera*. Ed. Edelsa, Madrid. 1997.
- (Fernández, 1991) Fernández S. *Análisis de errores e Interlengua en el aprendizaje del español como lengua extranjera*. Col. Tesis, U.Complutense, Madrid. 1991 (1997 en Edelsa).
- (Flowerdew, 1997) Flowerdew L. *Interpersonal strategies: investigating interlanguage corpora*. RELC Journal 28 (1): 72-88. 1997.
- (Flowerdew, 1996) Flowerdew J. *Concordancing in language learning*. In M. Pennington (eds.) *The Power of CALL*, pp. 97-113. Houston, TX: Athelstan. 1996.
- (Francis, 1982) Francis W.N. *Problems of assembling and computerizing large corpora*. Computer Corpora in English Language Research, ed. S. Johansson. Bergen: Norwegian Computing Centre for the Humanities. 7-24. 1982.

- (Fries, 1945) Fries C. *Teaching and Learning English as a Second Language*. Ann Arbor: University of Michigan Press. 1945.
- (Garcia, 2008) Garcia E. *Hizkuntzaren tratamendu automatikorako tresnak*. Bat: Soziolinguistika aldizkaria, ISSN 1130-8435, 66. zk ('Teknologi berriak eta euskara' gaiari eskainitako alea), 93-106 or. 2008.
- (Goenaga, 1980) Goenaga P. *Gramatika bideetan*. Erein D.L., Donostia, 1980.
- (Gojenola, 2000) Gojenola K. *Euskararen sintaxi konputazionalerantz. Oinarritzko baliabideak eta beren aplikazioa aditzen azpikategorizazio-informazioaren erauzketan eta errorearen tratamenduan*. Informatika Fakultatea, UPV-EHU, Donostia, 2000.
- (Granger, 2004) Granger S. *Computer learner corpus research: current status and future prospects*. Applied Corpus Linguistics. A Multidimensional Perspective. Eds. U. Connor and T. A. Upton. Amsterdam & Atlanta: Rodopi. 123-145. 2004a.
- (Granger, 2003) Granger S. *Error-tagged learner corpora and CALL: A promising synergy*. CALICO Journal 20 (3) (special issue on Error Analysis and Error Correction in Computer-Assisted Language Learning), pp. 465-480. 2003.
- (Granger, 2002) Granger S. *A Bird's-eye View of Computer Learner Corpus Research*. In Granger S., Hung J. and Petch-Tyson S. (eds) Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching. Amsterdam & Philadelphia: Benjamins, pp. 3-33. 2002.
- (Granger *et al.*, 2002) Granger S., Dagneaux E. and Meunier F. (eds.). *International Corpus of Learner English*. CDROM-Version 1.1. Université Catholique de Louvain: Centre for English Corpus Linguistics. 2002.
- (Granger, 1998a) Granger S. *The computer learner corpus: a versatile new source of data for SLA research*. Learner English on Computer. Ed. S. Granger. London: Longman. 3-18. 1998a.
- (Granger, 1998b) Granger S. (editor). *Learner English on Computer*. London & New York: Addison Wesley Longman (228 pp.). 1998b.
- (Granger, 1996) Granger S. *From CA to CIA and back: an integrated approach to computerized bilingual and learner corpora*. In Aijmer, K., B. Altenberg and M.

- Johansson (eds.) *Languages in Contrast: Papers from a Symposium on Text-based Cross-linguistic Studies*, Lund 4-5 March 1994, Lund: Lund University Press, pp. 37-51. 1996.
- (Granger, 1994) Granger S. *The learner corpus: a revolution in applied linguistics*. English Today 39 (10/3): 25-9. 1994.
- (Granger, 1993) Granger S. *The International Corpus of Learner English*. In Aarts, J., P. de Haan and N. Oostdijk (eds.) *English Language Corpora: Design, Analysis and Exploitation*. Amsterdam: Rodopi, pp. 57-69. 1993.
- (HABE, 1999) HABE. *Helduen Euskalduntzearen Oinarrizko Kurrikulua* (HEOK), 120-131. or., Donostia. 1999.
- (Hammarberg, 1974) Hammarberg B. *On the insufficiency of error analysis*. IRAL (International Review of Applied Linguistics), Vol. 12: 185-92. 1974.
- (Haugen, 1983) Haugen E. *The Implementation of Corpus Planning: Theory and Practice*. In Cobarrubias and Fishman (eds.). *Progress in Language Planning International Perspectives*. Berlin/New York/Amsterdam: Mouton. 1983.
- (Hendrickson, 1979) Hendrickson J.M. *Evaluating Spontaneous Communication Through Systematic Error Analysis*. Foreign Language Annals, Vol. 12, No. 5, 357-364. 1979.
- (Hoffman, 1998) Hoffman L. *El lloc que ocupa la recerca sobre els llenguatges d'especialitat en la lingüística aplicada. In Llenguatges d'especialitat*. Selecció de textos. Barcelona. Institut Universitari de Lingüística Aplicada: 91-102. 1998.
- (Hualde eta Ortiz de Urbina, 2003) Hualde J.I. eta Ortiz de Urbina J. (eds.). *A Grammar of Basque*. Berlin/New York: Mouton de Gruyter, ISBN: 3 11 017683 1. 2003.
- (Hutchinson, 1996) Hutchinson J. *UCL Error Editor*. Louvain-la-Neuve: Centre for English Corpus Linguistics, Université Catholique de Louvain. 1996.
- (IXA taldea eta Elhuyar Fundazioa, 2007) Ixa taldea, Elhuyar Fundazioa. *Testu-corpusak: ezaugarriak, eraketa eta tresnak*. Hizkuntza, komunikazioaren eta teknologiaren garaian. IVAP aldizkaria. ISBN 978-84-7777-316-0. 2007.

- (Izumi *et al.*, 2003) Izumi E., Toyomi S., Thepchai S., Kiyotaka U., Isahara H. ***The development of the spoken corpus of Japanese learner English and the applications in collaboration with NLP techniques***. In Archer *et al.* (eds.), 359-366. 2003.
- (Jain, 1984) Jain M.P. ***Error Analysis: Source, Cause and Significance***. En J.C. Richards (Ed.): *Error Analysis: Perspectives of Second Language Acquisition* (London: Longman), pp 189-215. 1984.
- (James, 1998) James C. ***Errors in Language Learning and Use: Exploring Error Analysis***. Longman, London and New York. 1998.
- (Johannessen *et al.*, 2002) Johannessen J.B., Hagen K., Lane P. ***The performance of a grammar checker with deviant language input***. Proceedings of the 19th International Conference on Computational Linguistics, 1–8, COLING, Taipei, Taiwan, 2002.
- (Juozulynas, 1994) Juozulynas V. ***Errors in the compositions of 2nd year german students: an empirical study for parser-based ICALL***. Indiana University, CALICO Journal, Vol. 12, No. 1, 5-15. 1994.
- (Karlsson *et al.*, 1995) Karlsson F., Voutilainen A., Heikkilä J., Anttila A. (eds.). ***Constraint Grammar. A Language-Independent System for Parsing Unrestricted Text***. Berlin and New York: Mouton de Gruyter, 1995.
- (Karttunen *et al.*, 1997) Karttunen L., Gaál T., Kempe A. ***Xerox Finite State Tool***. Technical report, Xerox Research Centre Europe. 1997.
- (King, 1983) King A. ***Batua bai, euskalkiak ere bai!*** Argia, 1983-II-13.
- (Koskenniemi, 1983) Koskenniemi K. ***Two-level Morphology: a general computational model for word-form recognition and production***. University of Helsinki, Helsinki, 1983.
- (Koutsis *et al.*, 2007) Koutsis I., Markopoulos G., Episimiotis G.M. ***A multilingual tool for hierarchical annotation of texts***. Corpus Linguistics Conference, Birmingham, July 27-30, 2007.
- (Krishna, 1975) Krishna V. ***The Syntax of Error***. Journal of basic Writing Spring, 43-49. 1975.

- (Lado, 1957) Lado R. *Linguistics Across Cultures. Applied Linguistics for Language Teachers*. Ann Arbor, Michigan: Michigan University Press. (Traducción española: (1973), Lingüística contrastiva: Lenguas y culturas, Madrid: Alcalá). 1957.
- (Lafitte, 1944) Lafitte P. *Grammaire Basque (Navarro-Labourdin Littéraire)*. Berrargitaratua, Elkar, Donostia, 1979.
- (Laka, 1996) Laka I. *A Brief Grammar of Euskara, the Basque Language*. Euskararako Errektoreordetza, EHU, [http://www.ehu.es/grammar web orrian eskuragarri](http://www.ehu.es/grammar_web_orrian_eskuragarri). 1996.
- (Laka, 1993) Laka I. *Unergatives that assign ergative, unaccusatives that assign accusative*. In Papers on Case and Agreement, Vol I., J. Bobaljik & C. Phillips (eds), MITWPL 18, Cambridge, Massachusetts, USA, pp.149-172.
- (Larsen-Freeman eta Long, 1994) Larsen-Freeman D. eta Long M. *Introducción al estudio de la adquisición de segundas lenguas*. Madrid. Gredos. 1994.
- (Larringan, 1993) Larringan L.M. *Euskara batua eta euskalkiak. Aintzat hartu beharreko praktikak*. Jakin 76. 11-39. 1993 maiatza-ekaina.
- (Lee, 1990) Lee N. *Notions of 'error' and appropriate corrective treatment*. Hong Kong Papers in Linguistic and Language Teaching, 13. 1990.
- (Leech, 1993) Leech G. *Corpus annotation schemes*. Literary and Linguistic Computing. 8(4). 275-281. 1993.
- (Leech eta Fliegelstone, 1992) Leech G. eta Fliegelstone S. *Computers and corpus analysis*. In Butler, C.S. (ed.). Computers and written texts. Oxford: Basil Blackwell. 115-140. 1992.
- (Liu Fang, 2004) Liu Fang. *A Computer-assisted Error Analysis of Prepositional Uses in Non-English Majors' Writing*. A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of Master of Arts, Foreign Linguistics and Applied Linguistics, Supervisor: Professor Zhou Jianglin, Huazhong University of Science & Technology, Wuhan 430074, P.R.China. 2004.
- (Lozano, 2008) Lozano C. *CEDEL2: Corpus Escrito del Español L2*. In Actas del XXVI Congreso de AESLA. Universidad de Almería. 2008.

- (Maia *et al.*, 2005) Maia J., Rojo J., Serna M. ***Euskarazko hutsak zuzentzen, teknologia berriak baliatuz***. Revista Psicodidáctica, ISSN 1136-1034, Vol. 10, Nº. 1, pags. 55-63. 2005.
- (Maia, 2001a) Maia J. ***Estandarra eta dialektoak nola ulertu eta tratatu irakaskuntzan***. Revista Psicodidáctica, nº 012, Euskal Herriko Unibertsitatea, Vitoria-Gasteiz. 2001a.
- (Maia, 2001b) Maia J. ***Eskola eta tokian tokiko hizkuntz aldaera***. Euskalkia eta hezkuntza, Mendebalde arg. 2001b.
- (Maritxalar, 1999) Maritxalar M. ***MUGARRI: Bigarren Hizkuntzako ikasleen hizkuntza ezagutza eskuratzeko sistema anitzeko ingurunea***. Doktoretza-tesia, Euskal Herriko Unibertsitatea (UPV/EHU), Donostia, 1999ko azaroaren 30a.
- (Martín del Rey, 2004) Martín del Rey M.A. ***Análisis de Errores de la Interlengua de español en estudiantes italianos***. Publicado por Elenet.org, ISBN: 2-9524532-0-9. 2004.
- (Mason eta Uzar, 2000) Mason O. eta Uzar R. ***NLP meets TEFL: Tracing the zero article***. In Lewandowska-Tomaszczyk, B. and J.P. Melia (eds.) 2000, pp. 105-116. 2000.
- (McEnery *et al.*, 2005) McEnery T., Xiao R. eta Tono Y. ***Corpus-based Language Studies: An advanced resource book***. Routledge Applied Linguistics Series. Taylor & Francis Group plc, London, UK. ISBN 0415286220. 2005.
- (McEnery eta Wilson, 1996) McEnery T. eta Wilson A. ***Corpus Linguistics***. Edinburgh University Press, ISBN 0-7486-0808-7 (hardback); ISBN 0-7486-0482-0 (paperback). 1996.
- (Meunier, 1998) Meunier F. ***Computer tools for learner corpora***. Learner English on Computer. Ed. S. Granger. London: Longman. 19-37. 1998.
- (Meyer, 2002) Meyer C.F. ***English Corpus Linguistics: An introduction***. Cambridge University Press. United Kingdom. 2002.
- (Milroy eta Milroy, 1991) Milroy eta Milroy. ***Authority in Language: Investigating Language Prescription and Standardisation***. Routledge (2nd edition). London, New York. September 20, 1991.

- (Milton, 1998) Milton J. **WORDPILOT: enabling learners to navigate lexical universes**. In S. Granger and J. Hung (eds), 97-98. 1998.
- (Milton eta Chowdhury, 1994) Milton John eta Chowdhury N. **Tagging the interlanguage of Chinese learners of English**. Proceedings joint seminar on corpus linguistics and lexicology, Guangzhou and Hong Kong, 19-22 June, 1993, Language Centre, HKUST, Hong Kong, 1994, p. 127-143. 1994.
- (Milton eta Tsang, 1993) Milton J. eta Tsang E. **A corpus-based study of logical connectors in EFL students' writing**. In R. Pemberton & E. Tsang (eds.) Studies in Lexis. Language Centre, The Hong Kong University of Science and Technology, pp. 215-246. 1993.
- (de Mönnink, 2000) de Mönnink I. **Parsing a learner corpus**. Corpus Linguistics and Linguistics Theory Eds C. Mair and M. Hundt. Amsterdam: Rodopi. 81-70. 2000.
- (Mujika, 1978) Mujika J.A. **Perpaus bakunaren sintasia. Adibideak, ariketak**. Gero argitaletxea. 1978.
- (Myles, 2005) Myles F. **Interlanguage corpora and SLA research**. Second Language Research 21 (4): 373-391. 2005.
- (Nemser, 1971) Nemser W. **Approximative systems of foreign language learners**. International Review of Applied Linguistics (IRAL), 9, págs. 115-123. 1971.
- (Nesselhauf, 2004) Nesselhauf N. **Learner corpora and their potential for language teaching**. How to Use Corpora in Language Teaching. Ed. J.M. Sinclair. Amsterdam: John Benjamins. 125-152. 2004.
- (Norrish, 1981) Norrish J. **Language learners and their errors**. Macmillan, 1981.
- (Oyharçabal, 2002) Oyharçabal B. **Euskaltzaindiaren corpusez**. In Hizkuntza-corpusak. Oraina eta geroa. Donostia: UZEI [on-line] [kontsulta: 05-01-22]. http://www.uzei.org/corpusajardunaldia/03_murkia.pdf . 2002.
- (Oyharçabal, 1996) Oyharçabal B. **Euskara batua eta euskalkiak**. Aitzolen mendeurrena eta II. Jagon Jardunaldiak. 41. zenb. Tolosa, 1996-X-17/19.
- (Oyharçabal, 1993) Oyharçabal B. **Euskara batuari iparraldetik behatuz**. Jakin 76: 79-89. 1993.

- (Oronoz, 2009) Oronoz M. *Euskararen errore sintaktikoak detektatzeko eta zuzentzeko baliabideen garapena: datak, postposizio-lokuzioak eta komunztadura*. Doktoretza-tesia, Euskal Herriko Unibertsitatea (UPV/EHU). 2009.
- (Otegi, 2006) Otegi A. *Zuzentzaile sintaktikoa word-en integratzeko liburutegi baten sorkuntza*. Karrera Bukaerako Proiektua. Informatika Fakultatea, EHU, Donostia, 2006ko maiatza.
- (Otegi, 2003) Otegi A. *Estilo kontuak eta puntuazio-ikurren erroreak detektatzeko sistema*. Karrera Bukaerako Proiektua. Informatika Fakultatea, EHU, Donostia, 2003ko iraila.
- (Pano, 2008) Pano A. *Estrategias comunicativas y lingüísticas en listas y foros de debate en Internet*. Alma Mater Studiorum, Università de Bologna. 2008.
- (Porquier, 1975) Porquier R. *Analyse d'erreurs en français langue étrangère: études sur les erreurs grammaticales dans la production orale libre chez les adultes anglophones*. Tesis, Univ. De París VIII. 1975.
- (Pravec, 2002) Pravec N. *Survey of learner corpora*. ICAME Journal 26:81-114. [Internet document available at <http://nora.hd.uib.no/icame/ij26/pravec.pdf>]. 2002.
- (Rojo, 2002) Rojo G. *Sobre la lingüística basada en el análisis del corpus. In Hizkuntza-corpusak*. Oraina eta geroa. Donostia: UZEI. [on line] [kontsulta: 09-02-24 <<http://www.uzei.com/Modulos/UsuariosFtp/Conexion/archivos54A.pdf>>]
- (Santos, 1993) Santos I. *Análisis contrastivo, análisis de errores e interlengua en el marco de la Lingüística contrastiva*. Madrid. Síntesis. 1993.
- (San Martin, 1998) San Martin J. *Euskararen ostarteak*. Donostia, Erein, 1998.
- (Saurí, 2004) Saurí R. *Un corpus para el asturiano: Las tecnologías lingüísticas en la consolidación de las lenguas minorizadas*. In Revista de Filología Asturiana. Volume 3/4. Años 2003/2004: 135-174. 2004.
- (Schachter eta Celcé-Murcia, 1977) Schachter J. eta Celcé-Murcia M. *Some Reservations Concerning Error Analysis*. TESOL Quarterly 11, 441-51, 1977.

- (Schumann eta Stenson, 1974) Schumann J.H. eta Stenson N. (editors). ***New Frontiers in Second Language Learning***. Ed. John H. Schumann & Nancy Stenson, Newbury House Publishers, Inc., Massachusetts. 1974.
- (Scott, 1996) Scott M. ***WordSmith Tools***. Oxford: Oxford University Press. 1996.
- (Selinker, 1974) Selinker L. ***Interlanguage***. In Richards, J. (Ed.). *Error analysis: Perspectives on Second Language Acquisition*. 31-54. Essex: Longman. 1974.
- (Selinker, 1969) Selinker L. ***Language transfer***. *General Linguistics* 9, 67-92. 1969.
- (Sinclair, 1996) Sinclair J. ***EAGLES. Preliminary recommendations on Corpus Typology***. <http://www.ilc.pi.it/EAGLES96/corpus/corpus/corpus.html>. 1996.
- (Sinclair, 1991) Sinclair J.M. ***Corpus, Concordance, Collocation***. Oxford: Oxford University Press. 1991.
- (Skinner, 1985) Skinner B. F. ***Aprendizaje y comportamiento***. Ed. Martínez Roca. 1985.
- (Sperling, 1995): <http://www.eslcafe.com/>
- (Sridhar, 1981) Sridhar N. ***Contrastive Analysis, Error Analysis and Interlanguage: Three Phases of One Coal***. En J. Fisiak (Ed.): *Contrastive Linguistics and the Language Teacher* (Oxford: Pergamon), pp. 207-243. 1981.
- (Stevens, 1969) Stevens P. ***Two ways of looking at error analysis***. ERIC: 037 714, 1969.
- (Suri eta McCoy, 1993) Suri L. eta McCoy K. ***A Methodology for Developing an Error Taxonomy for a Computer Assisted Language Learning Tool for Second Language Learners***. Technical Report, TR-93-16. Dept. Of CIS, University of Delaware. 1993.
- (Teubert, 2006) Teubert W. (ed). ***Corpus Linguistics-Critical Concepts in Linguistics***. London: Routledge. 2006.
- (Ties, 2006) Ties I. ***Corpus linguistics: a general introduction***: http://www.eurac.edu/NR/rdonlyres/33276E8D-C0EB-40C8-B0A3-9B6037F46152/11459/CL_introduction1.pdf. 2006.

- (Tono, 2003) Tono Y. *Learner corpora: design, development and applications*. In Archer *et al.* (eds.), Proceedings of Corpus Linguistics 2003. Lancaster University, 800-809. 2003.
- (Tono, 2002) Tono Y. *The Role of Learner Corpora in SLA Research and Foreign Language Learning: The Multiple Comparison Approach*. Unpublished PhD thesis. Lancaster University. 2002.
- (Tono, 2000) Tono Y. *A computer learner corpus-based analysis of the acquisition order of English grammatical morphemes*. In Burnard, L. and T. McEnery (eds.) Rethinking Language Pedagogy from a Corpus Perspective. Frankfurt: Peter Lang, pp. 123-132. 2000.
- (Tono eta Aoki, 1998) Tono Y. eta Aoki M. *Developing the optimal learning list of irregular verbs based on the native and learner corpora*. In S. Granger and J. Hung (eds) 1998, pp. 113-118. 1998.
- (Torijano, 2004) Torijano J.A. *Errores de aprendizaje, aprendizaje de los errores*. Arco/Libros S.L., Madrid. 2004.
- (Trask, 2003) Trask L. *The noun phrase: nouns, determiners, and modifiers; pronouns and names*. J.I. Hualde & J. Ortiz de Urbina (arg), A Grammar of Basque. Dordrecht: Kluwer, 92-134. 2003.
- (Trudgill, 1983) Trudgill P. *Accent, Dialect and School*. London, Edward Arnold (publishers) (7. argitaraldia, lehendabizikoa 1975ean). 1983.
- (Txillardegui, 1978) Txillardegui. *Euskal Gramatika*. Ediciones Vascas, Bilbo. 1978.
- (Uria *et al.*, 2009) Uria L., Arrieta B., Díaz de Ilarraza A., Maritxalar M., Oronoz M. *Determiner errors in Basque: Analysis and Automatic Detection*. XXV. Congreso de la SEPLN (Donostia). 2009.
- (Uzar, 1997) Uzar R. *Was PELE a linguist?* In Lewandowska-Tomaszczyk, B. & P. J. Melia (eds.) PALC '97 (Practical Applications in Language Corpora, Łódź, Poland 10-14 April 1997). 1997.
- (Villasante, 1998) Villasante A.L. *Euskararen auziaz*. Estella. 1998.
- (Villasante, 1983) Villasante, L. *Estudios de Sintaxis Vasca*. Editorial Franciscana Aranzazu, Donostia. 1983.

- (Villasante, 1972) Villasante, L. ***La Declinación del Vasco Literario Común***. Editorial Franciscana de Aranzazu, Donostia. 1972.
- (Yang eta Akahori, 1998) Yang J.C eta Akahori K. ***Error Analysis in Japanese Writing and Its Implementation in a Computer Assisted Language Learning System on the World Wide Web***. CALICO J., vol.15, no. 1-3, pp. 47-66. 1998.
- (Yang Huizhong eta Gui Shichun, 2005) Yang Huizhong eta Gui Shichun. ***CLEC-based Chinese EFL Learners English Analysis***. Shanghai: Shanghai Foreign Language Education Press. 2005.
- (Zabala, 2004) Zabala I. ***Los predicados complejos en vasco***. E. Pérez, Ll. Gràcia e I. Zabala (eds.), Estructuras morfológicas y estructuras sintácticas: la frontera de la composición en lenguas románicas y en vasco: 445-534. Universidad de Deusto. Donostia. 2004.
- (Zabala, 1993) Zabala I. ***Predikazioaren teoriak gramatika sortzailean (euskararen kasua)***. Doktoretza-tesia, Euskal Herriko Unibertsitatea (UPV/EHU). 1993.
- (Zuazo, 2005) Zuazo K. ***Euskara batua: ezina ekinez egina***. Elkar. 2005.
- (Zuazo, 2000) Zuazo K. ***Euskararen sendabelarrak***. Alberdania, Irun. 2001.
- (Zubiri eta Zubiri, 1995) Zubiri I. eta Zubiri E. ***Euskal Gramatika Osoa***. Didaktiker SA, Bilbo. 1995.
- (Zubiri, 1990) Zubiri I. ***Ikas-Txip, euskara ordenadore bidez***. Zutabe, 24, 9-55. Donostia: H.A.B.E. 1990.

Bibliografia

Atal honetan tesia osatzeko kontsultatu ditugun baina testuan erreferentziatu ez ditugun lanak bildu ditugu:

Aduriz I. eta Díaz de Ilarraza A. ***Morphosyntactic disambiguation and shallow parsing in Computational Processing of Basque***. In Oyharzabal B., editor, *Inquiries into the lexicon-syntax relations in Basque*. ASJUren gehigarria. Euskal Herriko Unibertsitatea, Bilbo. 2003.

Aduriz I., Aldezabal I., Aranzabe M., Arrieta B., Arriola J., Atutxa A., Díaz de Ilarraza A., Gojenola K., Oronoz M., Sarasola K., Urizar R. ***The design of a digital resource to store the knowledge of linguistic errors***. DRH2002 (Digital Resources for the Humanities). Edinburgo. 2002.

Alcaraz E. eta Martínez M. ***Diccionario de lingüística moderna***. Barcelona: Ariel. 1997.

Aldezabal I., Arriola J.M., Díaz de Ilarraza A., Sarasola K. ***Hizkuntzalaritza konputazionala***. UEU. 2005.

Aldezabal I., Alegria I., Ansa O., Arriola J.M., Ezeiza N. ***Designing spelling correctors for inflected languages using lexical transducers***. Proceedings of EACL'99, 265-266. Bergen, Norway. 1999.

Alegria I., Ceberio K., Ezeiza N., Hernández G., Soroa A. ***Euspell: corrección ortográfica del euskera en software libre***. FLOSSIC-2007. 213-220. Jerez. Arg. digitala. ISBN 978-84-9828-124-8. 2007.

Ansa O., Arregi X., Arrieta B., Díaz de Ilarraza A., Ezeiza N., Fernandez I., Garmendia A., Gojenola K., Laskurain B., Martínez E., Oronoz M., Otegi A., Sarasola K., Uria L. ***Integrating NLP Tools for Basque in Text Editors***.

- Workshop on International Proofing Tools and Language Technologies. University of Patras. Greece. 2004.
- Bredenkamp A., Klein J., Crysmann B. *Annotation of error types for a german news corpus*. In ATALA sur les Corpus Annotés pour la Syntaxe Treebanks, June 18-19, Paris, France. 1999.
- Bueno Velazco C. eta Martínez Herrera J.M. *Aprender y enseñar inglés: cinco siglos de historia*. Language teaching methods: a 500 years ' journey. Rev. Hum. Med. vol.2 no.1 Ciudad de Camaguey Jan.-Apr. 2002.
- Burt M.K. eta Kiparsky C. *Global and local mistakes*. In New Frontiers in Second Language Learning edited by J. H. Schumann and N. Stenson, Newbury House Publishers Inc. 1974.
- Bartholomae D. *The study of error*: <http://astro.temple.edu/~sparkss/studyoferror.htm>
- Cassany D., Luna M., Sanz G. *Enseñar lengua*. Bartzelona, Graó. 1994.
- Chujo K., Utiyama M., Nishigaki C. *Towards building a usable corpus collection for the ELT classroom*. In Corpora in the Foreign Language Classroom. Edited by Encarnación Hildalgo, Luis Quereda and Juan Santana, pp. 47-69. Rodopi. 2004.
- Corder S.P. *Error Analysis and Interlanguage*. Oxford Univeristy Press. 1981.
- Díaz de Ilarraza A., Gojenola K., Oronoz M. *Reusability of a corpus and a treebank to enrich verb subcategorisation in a dictionary*. Conference on Recent Advances in Natural Language Processing (RANLP07). Pages 280-284. ISBN: 978-954-91743-7-3, 27-29 September 2007, Borovets, Bulgaria. 2007.
- Díaz de Ilarraza A., Maritxalar M., Oronoz M., Maritxalar A. *Integration of NLP Tools in an Intelligent Computer Assisted Language Learning Environment for Basque: IDAZKIDE*. Proceedings of Natural Language Processing and Industrial Applications Moncton, Canada. 1998.
- Ensunza M., Etxebarria J.M., Iturbe J. *ZIENTZIA ETA TEKNIKARAKO EUSKARA: Zenbait baliabide*. Udako Euskal Unibertsitatea. Bilbo. 2002.
- Esnal P. *Erroreak helduen euskalduntzearen azken urratsetan*. ZUTABE, 17 zk., 9-66 orr. 1988.

- Foster J. eta Vogel C. *Parsing ill-formed text using an error grammar*. Artificial Intelligence Review, 21(3-4):269–291. 2004.
- Granger S. *The International Corpus of Learner English: a new source for foreign language learning and teaching and second language acquisition research*. TESOL Quarterly 37 (3): 538-546. 2003.
- Gutiérrez Quintana E. *Análisis de la interlengua de italianos aprendices de español*. Estudios de lingüística, ISSN 0212-7636, Nº 19, pags. 223-242. 2005.
- HABE. *Helduen euskalduntzerako programazioa*. Donostia: HABE. 1981.
- Hashemi S. *Detecting grammar errors in children's writing: A finite state approach*. In Proc. 13th Nordic Conf. in Computational Linguistics (Nodalida-01). 2001.
- Hasselgård H. 2001: <http://folk.uio.no/hhasselg/UV-corpus.htm>
- Howard H. Kleinmann. *Avoidance Behavior in Adult Second Language Acquisition*. In Second Language Learning: Contrastive Analysis, Error Analysis and Related Aspects, edited by Betty Wallace Robinett & Jacqueslyn Schachter, Ann Arbor, the University of Michigan Press, 1983.
- Hundt M. 2007:
<http://www.rzuser.uni-heidelberg.de/~hu1/Korpuslinguistik/Corpus%20Linguistics%20and%20Corpora.htm>
- Kaplan R. (editor). *On the scope of applied linguistics*. 1st printing. Rowley, MA: Newbury House. 1980.
- Lado R. *Patterns of Difficulty in Vocabulary*. Language Learning 6, 23-41. 1956.
- Madrid D. *Errores gramaticales en la producción escrita de los angloamericanos*. En A. Romero *et al.*: Educación lingüística y literaria en el ámbito escolar. Granada: Grupo Editorial Universitario, pp. 609-623. 1999.
- Matte Bon F. *De nuevo la gramática*. Actas de las Segundas Jornadas Internacionales de Didáctica del Español como LE. pp. 109-124. 1988.
- Naber D. *A Rule-Based Style and Grammar Checker*. PhD thesis, Technische Fakultät. Universität Bielefeld.

- Nada AbiSamra. *An Analysis of Errors in Arabic Speakers' English Writings*. American University of Beirut; Instructor: Dr. Kassim Shaaban; Second Language Acquisition; English 345; January 2003.
- Rambell. *Error typology for automatic proof-reading purposes*. En A. Sagvall Hein, editor, Reports from the SCARRIE project, Uppsala. 1999.
- Scott M. *WordSmith Tools version 4*. Oxford: Oxford University Press. ISBN: 0-19-459400-9. 2004.
- Scott M. *Comparing corpora and identifying key words, collocations, and frequency distributions through the WordSmith Tools suite of computer programs*. In M. Ghadessy, A. Henry & R.L. Roseberry (eds.) Small corpus studies and ELT: theory and practice. Amsterdam: Benjamins. pp. 47- 67. 2001.
- Urkia M. 2002. *XX. mendeko euskara-corpusa*. In Hizkuntza-corpusak. Oraina eta geroa. Donostia: UZEI [on line] [kontsulta: 05-01-22] http://www.uzei.org/corpusajardunaldia/03_murkia.pdf
- Vandeventer Faltin A. *Syntactic error diagnosis in the context of CALL*. These, 2003.
- Vázquez G. *Análisis de errores y aprendizaje de español / lengua extranjera*. Peter Lang, Frankfurt am Main, Bern, New York, Paris. 1991.
- Wallace Robinett B. eta Schachter J. (editors). *Second Language Learning: Contrastive Analysis, Error Analysis and Related Aspects*. Ann Arbor, the University of Michigan Press, 1983.

ERANSKINAK

A eranskina:

Beste hizkuntza batzuetan garatu diren errore-sailkapenak

Hendrickson (1979)

- Lexical errors
- Morphological errors
- Syntactic errors
- Spelling errors

Suri & McCoy (1993)

Gramatika-kategoriak:

- NP and VP Conjunctions: 18
 - Omitted conjunction: 10
 - Inappropriate conjunction: 7
 - Extra conjunction: 1
- Inappropriate sentential conjunction: 2
- Prepositions: 79
 - Omitted preposition: 27
 - Inappropriate preposition: 35
 - Extra Preposition: 17
- Determiners: 92
 - Omitted determiner: 49
 - Inappropriate determiner or determiner formation: 17
 - Extra Determiner: 26
- Incorrect Subject-Verb Agreement: 14
- Tense and Aspect: 95
 - Dropped Tense: 6
 - Extra Auxiliary: 7
 - Missing Auxiliary: 2
 - Incorrect Modal: 3
 - Missing Modal: 2
 - Extra Modal: 1
 - Other tense/aspect problems: 74
- BE, HAVE (non-Auxiliary): 28
 - Omitted BE: 18
 - Lack of BE/HAVE distinction: 10
- Other Omitted Main Verbs: 15
- Incorrect Main Verbs: 13 (Poor lexical choice)
- Relative Clauses: 26
 - Relative pronoun deletion: 5
 - Resumptive pronoun: 1
 - Incorrect WH-relative pronoun: 4

- Other: 16
- Mixing up English words or phrases which share a single ASL sign: 24
- Adjective and Adverb Problems: 27
 - Incorrect Adjective Choice: 3
 - Incorrect Adjective Formation: 13
 - Mixing up Adjectives and Adverbs: 2
 - Incorrect Adjective Order: 1
 - Missing Adjective: 2
 - Other Adverb Problems: 6
- Incorrect Number on Noun: 36
- Problems with Noun Formation : 6
- Problems with Referent Formation : 5
- Pronouns: 25
 - Incorrect pronoun choice (including pleonastic): 12
 - Inappropriate pronoun use (where full definite descriptions are required): 11
 - Lack of pronoun use (overuse of definite descriptions): 2
- Pleonastic Pronoun Problems: 20
 - Object Deletion: 4
 - Subject Deletion: 8
 - Incorrect Pleonastic Pronoun: 5
 - Other Pleonastic Pronoun Problems: 3
- Redundancy Problems: 7
- Other Problems that may be related to Focus/Discourse Structuring: 76
 - Noun Phrase Omission (subject: 16; object:15): 31
 - Problems carrying over general/specific description strategies: 6
 - Structuring Problems with “because”: 9
 - Other (may be related to topic-comment structures, or verbs of ASL): 25
 - Other: 5
- Other Illegal Ellipses: 10
- Run-on Sentences: 8
- Idioms, word choice, hard to define problems: 36

Diskurtso mailako errore-kategoriak:

- Inappropriate determiner or determiner formation: 17
- Dropped Verb Tense: 6
- Relative Clauses: 27
- Redundancy Problems: 7
- Pleonastic Pronoun Problems: 20
 - Object Deletion: 4
 - Subject Deletion: 8
 - Incorrect Pleonastic Pronoun: 5
 - Other Pleonastic Pronoun Problems: 3
- Run-on Sentences: 8
- Problems with Referent Formation: 5
- Pronouns: 25
 - Incorrect pronoun choice (including pleonastic): 12
 - Inappropriate pronoun use (where full definite descriptions are required): 11
 - Lack of pronoun use (overuse of definite descriptions): 2
- Other Problems that may be related to Focus/Discourse Structuring: 76
 - Noun Phrase Omission (subject: 16; object: 15): 31
 - Problems carrying over general/specific description strategies: 6
 - Structuring Problems with “because”: 9
 - Other (may be related to topic-comment structures, or verbs of ASL): 25
 - Other: 5
- Other Illegal Ellipses: 10

Juozulynas (1994)

- Semantic errors
 - o Wrong word choice
 - o “made-up” words
 - o Errors in pronoun reference
- Local errors
 - o Morphological errors
 - Noun gender
 - Case endings of nouns
 - Number form of noun
 - Ending of adjective / ordinal number in group adjective / ordinal number + noun
 - Definite article / *viele* or Indefinite article / *viel* with noun
 - Ending of adjective / ordinal number in group *ein*-word + adjective/ordinal number + noun
 - Ending of adjective / ordinal number in group *der*-word + adjective/ordinal number + noun
 - Verb form
 - Indeclinable adjective or number form
 - Adjective / adverb form
 - Auxiliary verb in tense formation (*haben/sein*)
 - o Syntactic errors
 - Preposition + Required case
 - Verb + (Required preposition) + Required case
 - Word order in main clause
 - Subject – verb agreement
 - Word order in dependent clause
 - Missing parts of sentence/clause (subject, verb, part of verb, object, etc.)
 - Relative pronoun
 - Reflexive verb
 - Word order in dependent clause with additional infinitive
 - Infinitive particle *zu*
 - Adverb or adjective with required case and preposition
 - o Punctuation errors
 - Missing comma(s) in structure main clause + main clause
 - Missing comma(s) in structure main clause + dep. clause or dep. clause + dep. clause
 - Presence of erroneous comma
 - Missing comma(s) in infinitive group
 - Missing question mark
 - Missing exclamation mark
 - Missing quotation mark
 - o Spelling errors
 - The missing umlaut
 - Minisculization needed
 - Capitalization needed
 - Indiscrimination between ss/ s / ß
 - Erroneous umlaut
- Global errors
 - o Morphology
 - o Syntax
 - o Spelling
 - Typos
 - Letter transposition
 - Misspelling
 - Separation of compound words into parts
 - Etc.

Sónsoles Fernández (1997)

1. LÉXICOS:

1.1. Forma

- 1.1.1 Uso de un significante español próximo (*campana/campaña)
- 1.1.2 Formaciones no atestiguadas en español (*examinación)
- 1.1.3 Barbarismos –préstamos– (*various)
- 1.1.4 Género (como rasgo del nombre) (*una viaje, *la calor)
- 1.1.5 Número (*sólo tenía una gana)

1.2. Significado

- 1.2.1 Lexemas con semas comunes pero no intercambiables en el contexto (aprender-estudiar)
- 1.2.2 Cambios entre derivados de la misma raíz (*un visitado/una visita)
- 1.2.3 Registro no apropiado a la situación
- 1.2.4 Ser-estar (*ellos son en mi país)
- 1.2.5 Perífrasis (*limpiar el cuerpo/lavarse)
- 1.2.6 Otros (*un visitado/una visita)

2. GRAMATICALES:

2.1. Paradigmas

- 2.1.1 Género (formación) (*el artista/artista)
- 2.1.2 Número (formación) (*lunas de mieles)
- 2.1.3 Verbos (formación) (*cojó /cojió)
- 2.1.4 Otros (persona, determinante) (*ello/él)

2.2. Concordancias

- 2.2.1 En género (*la ambiente frío)
- 2.2.2 En número (*problemas socials)
- 2.2.3 En persona (*yo fue a Granada)

2.3. Valores y usos de las categorías

2.3.1 Artículo

- 2.3.1.1 Uso / omisión (*hay _ cosa que...; *estudió la Biología)
- 2.3.1.2 Elección (*era un día mejor de las vacaciones)

2.3.2 Otros determinantes (elección) (*este por ese y por aquel*)

2.3.3 Pronombres

- 2.3.3.1 Con función pronominal plena (*_ pasamos mal)
- 2.3.3.2 “se” gramaticalizado o lexicalizado (*te vas a ser feliz)

2.3.4 Verbos

- 2.3.4.1 Pasados (*se cayó pero no le pasaba nada)
- 2.3.4.2 Otras formas (*si vendrá a Madrid)

2.3.5 Preposiciones

- 2.3.5.1 Valores propios (*es la costumbre de japones; *miro _ mis hijos)
- 2.3.5.2 Valores idiomáticos (*aprenden en memoria)

2.4. Estructura de la oración

- 2.4.1 Orden (*coreanos libros leo)
- 2.4.2 Omisión de elementos (no incluidos en otros apartados) (*voy a _ mi amiga)
- 2.4.3 Elementos sobrantes (no incluidos en otros apartados) (*y donde cerca de la costa)
- 2.4.4 Cambios de función (*era un día ni frío ni calor)
- 2.4.5 Oraciones negativas (*yo también no fui)

2.5. Relación entre oraciones

2.5.1 Coordinación

- 2.5.1.1 Omisión del enlace (*queremos salir _ no podemos)
- 2.5.1.2 Polisíndeton (*repetición de “y”)
- 2.5.1.3 Elección errónea del enlace (*hemos comprado siete sino ocho libros)

2.5.2 Subordinación

2.5.2.1 Adjetiva

- 2.5.2.1.1 Omisión del enlace o enlace sobrante (*era la primera vez _ salía fuera)
- 2.5.2.1.2 Elección de enlace (*era mi mejor amigo quien me cae muy bien)
- 2.5.2.1.3 Concordancia verbal (*el día cuando vivían en paz, cuando no habré guerras)

2.5.2.2 Sustantiva

- 2.5.2.2.1 Omisión del enlace o enlace sobrante (*no saben _ yo quiero)
- 2.5.2.2.2 Elección de enlace (*me preguntaron que te vas)

- 2.5.2.2.3 Concordancia verbal (*a ellos dije que me fuera de vacaciones)
- 2.5.2.2.4 Coherencia de otros elementos en el estilo indirecto (*dijimos que ahora mismo)
- 2.5.2.3 Circunstanciales
 - 2.5.2.3.1 Omisión del enlace o enlace sobrante (*por_en aquel día me fui)
 - 2.5.2.3.2 Elección de enlace (*pues echaba de menos el tiempo pasó muy rápido)
 - 2.5.2.3.3 Concordancia verbal (*si tengamos tiempo)
- 3. DISCURSIVOS:
 - 3.1 Coherencia global
 - 3.2 Correferencia: deixis y anáfora
 - 3.3 Tiempo y aspecto
 - 3.4 Conectores
 - 3.5 Puntuación
- 4. GRÁFICOS:
 - 4.1 Puntuación y otros signos
 - 4.2 Tildes (*pelicula)
 - 4.3 Separación y unión de palabras (*pobre_cita, sillama)
 - 4.4 Alteración del orden de las letras (*madurgada)
 - 4.5 Confusión de fonemas (*ocopado, intelesante)
 - 4.6 Omisión de letras y letras sobrantes (*fa-cinado, cononcer)
 - 4.7 Confusión de grafemas para el mismo fonema (*estube)
 - 4.8 Mayúsculas (*Lope de vega)

James (1998)

Types of errors:

- Substance errors
 - o Misspellings
 - Punctuation errors
 - Typographic errors
 - Dyslexic errors
 - Confusibles
 - o Misspellings proper
 - Mispronunciation errors
 - Written misencodings
 - o Mispronunciations
 - Segmental
 - Combinatorial
 - Suprasegmental
- Text errors
 - o Lexical errors
 - Formal misselection
 - Misformations
 - Distorsions
 - o Semantic errors in lexis
 - Confusion of sense relations
 - Collocational errors
 - o Grammar errors
 - Morphology errors
 - Syntax errors
 - Phrase structure errors
 - Clause errors
 - Sentence errors
 - Intersentence errors (cohesion)
 - o Discourse errors
 - Coherence
 - o Pragmatic errors

- Taboos
- Size of the imposition
- Values power and social distance
- o Receptive errors

Bustos (1998)

A) ERRORES INTERLINGÜÍSTICOS

1. Interferencias que tienen su origen en la lengua materna del usuario

- 1.1. Interferencias simples
 - 1.1.1 Gráficas y Ortográficas
 - 1.1.2 Léxicas
 - 1.1.3 Morfológicas
 - 1.1.4 Morfosintácticas
 - 1.1.5 Sintácticas
 - 1.1.6 Semánticas
 - 1.1.7 Pragmático discursivas
- 1.2. Interferencias compuestas por bifurcación
 - 1.2.1 Bifurcación léxico-semántica
 - 1.2.2 Bifurcación morfológica
 - 1.2.3 Bifurcación morfosintáctica
 - 1.2.4 Bifurcación sintáctica
 - 1.2.5 Bifurcación pragmático discursiva
- 1.3. Interferencias compuestas por inexistencia

2. Interferencias que tienen su origen en otra segunda lengua

- 2.1. Interferencias simples
 - 2.1.1 Gráficas y Ortográficas
 - 2.1.2 Léxicas
 - 2.1.3 Morfológicas
 - 2.1.4 Morfosintácticas
 - 2.1.5 Sintácticas
 - 2.1.6 Semánticas
 - 2.1.7 Pragmático discursivas
- 2.2. Interferencias compuestas por bifurcación
 - 2.2.1 Bifurcación léxico-semántica
 - 2.2.2 Bifurcación morfológica
 - 2.2.3 Bifurcación morfosintáctica
 - 2.2.4 Bifurcación sintáctica
 - 2.2.5 Bifurcación pragmático discursiva
- 2.3. Interferencias compuestas por inexistencia

B) ERRORES INTRALINGÜÍSTICOS

1. Errores intralingüísticos formales

- 1.1. Errores intralingüísticos formales por acumulación
- 1.2. Errores intralingüísticos formales por confusión
- 1.3. Errores intralingüísticos formales por hipótesis falsa
- 1.4. Errores intralingüísticos formales por regularización
- 1.5. Errores intralingüísticos formales por simplificación
- 1.6. Errores intralingüísticos formales por sobregeneralización

2. Errores intralingüísticos pragmático discursivos

- 2.1. Errores intralingüísticos pragmático discursivos por acumulación
- 2.2. Errores intralingüísticos pragmático discursivos por confusión
- 2.3. Errores intralingüísticos pragmático discursivos por hipótesis falsa
- 2.4. Errores intralingüísticos pragmático discursivos por regularización
- 2.5. Errores intralingüísticos pragmático discursivos por simplificación

2.6. Errores intralingüísticos pragmático discursivos por sobregeneralización

C) ERRORES SIN CATEGORIZAR

Becker et al. (1999)

Error type	Label	Token
Syntax (general)	S	3
Subject-verb agreement	SASV	63
Antecedent-anaphor agreement	SAAA	1
Concord (NP-internal agreement)	SC	180
Word order	SO	79
Valency (general)	SG	0
Subcategorisation	SGCat	854
Case assignment	SGCas	102
Semantic selection	SGS	265
Σ Syntax		1547
Morphology	M	91
Othography (general)	O	2893
Punctuation	OI	1701
Capital vs. small letter	OC	2776
One word vs. separate words	OS	1100
Σ Orthography		7561
All		9108

Granger (2003)

Errore mota eta errore-kategoriei dagokien sailkapena:

Error domains		Error categories	
<F>	Form	<AGL> <MAJ> <DIA> <HOM> <GRA>	Agglutination Upper/lower case Diacritics Homonymy Other spelling errors
<M>	Morphology	<MDP> <MDS> <MFL> <MFC> <MCO>	Derivation-prefixation Derivation-suffixation Inflection Inflection-confusion Compounding
<G>	Grammar	<CLA> <AUX> <GEN> <MOD> <NBR> <PER> <TPS> <VOI> <EUF>	Class Auxiliary Gender Mode Number Person Tense Voice Euphony
<L>	Lexis	<SIG> <CPA> <CPD> <CPV> <CPN> <FIG>	Meaning Adjective complementation Adverb complementation Verb complementation Noun complementation Prefab
<X>	Syntax	<ORD> <MAN> <RED> <COH>	Word order Word missing Word redundant Cohesion
<R>	Register	<RLE> <RSY>	Lexis Syntax
<Y>	Style	<CLR> <LOU>	Unclear Heavy
<Q>	Punctuation	<CON> <TRO> <OUB>	Punctuation confusion Punctuation redundant Punctuation missing
<Z>	Typo		

Gramatika-kategoriari dagokion sailkapena:

Gramatical categories		Tag
Adjective	Simple	ADJ
	Comparative	AJC
	Superlative	AJX
	Complex	AJL
Adverb	Simple	ADV
	Complex	AVL
Article	Definite	ADE
	Indefinite	AIN
	Partitive	APA
	Contracted	ACO
Conjunction	Coordinator	COC
	Simple subordinator	COS
	Complex subordinator	COL
Determiner	Demonstrative	DED
	Possessive	DEP
	Indefinite	DEI
	Exclamative-Interrogative	DEX
	Relative	DER
	Numeral	DEN
Noun	Common simple	NOM
	Common compound	NOC
	Common complex	NOL
	Proper	NOP
Preposition	Simple	PES
	Complex	PEL
Pronoun	Demonstrative	POD
	Possessive	POP
	Personal	POO
	Indefinite	POI
	Exclamative-Interrogative	POX
	Numeral	PON
	Adverbial	POA
	Relative	POR
	Impersonal	POS
Verb	Finite simple	VSC
	Participle simple	VSP
	Gerund simple	VSG
	Infinitive simple	VSI
	Finite complex	VCC
	Participle complex	VCP
	Gerund complex	VCG
	Infinitive complex	VCI
Punctuation	Period	PUP
	Question mark	PUI
	Exclamation mark	PUE
	Comma	PUV
	Semi-colon	PUG
	Colon	PUD
	Suspension periods	PUS
	Parentheses	PUA
	Square bracktes	PUC
	Quotation marks	PUL
	Dash	PUT
	Slash	PUO
Sequence		SEQ

Díaz (2005)

1. Errores de puntuación
 - a. Problemas relacionados con la colocación de espacios entre signos de puntuación
 - b. Errores en el uso de la caja tipográfica
 - c. Casos en que no se produce un emparejamiento correcto entre signos de interrogación, paréntesis, comillas, etc. de apertura y cierre
 - d. Problemas de colocación de signos de puntuación en general, por omisión, adición innecesaria o incorrección en el orden seguido
2. Errores léxicos
 - a. Espaciado (omisión o adición de un espacio, dando lugar a una palabra o secuencia incorrecta)
 - b. Acentuación (omisión o adición innecesaria de una tilde, dando lugar a una palabra incorrecta o diferente de la esperada en un contexto determinado)
 - c. Incorrecciones en la escritura de nombres propios, acrónimos y abreviaturas
 - d. Uso de una variante gráfica no correcta
 - e. Errores en alguna forma del paradigma flexivo de un sustantivo, un adjetivo o un verbo
 - f. Errores producidos en construcciones complejas que superan el nivel de la palabra
 - g. Problemas de paronimia o de los llamados “pares de confusión”
 - i. Paronimia sintáctica
 - ii. Paronimia semántica
 - h. Mala selección léxica en la coaparición de palabras, en cuanto a las propiedades morfológicas de las mismas
 - i. Problemas en la escritura de extranjerismos
 - j. Americanismos léxicos; no se trata de errores, sino más bien de formas condicionadas por un uso marcado en español peninsular
3. Errores sintácticos
 - a. Mala categorización de algún elemento dentro de la oración
 - b. Construcciones sintácticas incorrectas
 - c. Problemas de rección sintáctica; aquí se incluyen numerosos casos
 - i. Errores de rección preposicional
 - ii. Uso incorrecto de la preposición en el complemento directo
 - iii. Mala rección de tiempo verbal
 - iv. Mala rección de modo verbal
 - v. Loísmo, leísmo y laísmo
 - vi. Queísmo y dequeísmo
 - d. Expresión de un complemento agente en pasivas reflejas
 - e. Incorrección en el orden de palabras
 - f. Americanismos sintácticos; tampoco aquí se trata de un error propiamente dicho, sino más bien de un uso marcado (por ser poco frecuente en el español peninsular) de ciertas estructuras
4. Errores de concordancia
 - a. Concordancia intrasintagmática
 - b. Concordancia intersintagmática
 - c. Concordancia con femeninos débiles
 - d. Concordancia *ad sensum*
5. Errores semánticos
 - a. Errores relacionados con una confusión en cuanto al significado de ciertas piezas léxicas, sin que se trate de parónimos semánticos
 - b. Mala selección léxica en la coaparición de palabras, en cuanto a las propiedades semánticas de las mismas
 - c. Problemas de duplicación de la negación
 - d. Redundancia semántica

Julian Maia et al. (2005)

1. Deklinabide arazoak
 - 1.1. Deklinabide-kasuak oker
 - 1.2. Zenbait esapidetan
2. Aditz arazoak
 - 2.1. Adizki jokatu gabeak
 - 2.2. Adizki jokatuak
 - 2.3. Aditzaren erregimena aldatzea
3. Perpauseko elementuen ordena
4. Perpaus bakunaren barneko beste puntu batzuk
 - 4.1. Esapide kalkatu batzuk
5. Lokailuak (“ere” partikularen erabilera)
 - 5.1. “ere”-ren erabileraz
 - 5.2. “baita ere” / “ezta ere”-ren erabileraz
6. Perpaus elkartuaren sintaxia
 - 6.1. Nominalizazioak (menpeko perpaus ez-jokatuak – konpletiboak)
 - 6.2. Zehar-galderak
 - 6.3. Menpeko perpausaren hasieran galdetzailea jartzea
 - 6.4. “zergatik” galdetzailea kausazko perpausaren hasieran
 - 6.5. Galdetzailea erlatiboazko perpaus baten hasieran (eta atzizkirik ez)
 - 6.6. Erdarazko QUE menperatzaile batzuen eragina
 - 6.7. Perpaus osagarri izenlagunak osatzeko atzizki desegokiak erabiltzea
7. Arlo lexiko-semantikoa
8. Ondo dauden esaldiak, baina beharbada zuzenduko direnak, tradizioa ezkutuan gelditzen delakoz

B eranskina:

Euskarazko errorearen sailkapena

1. Errore ortografikoak: Arau ortografikoak ez errespetatzeagatik sortzen diren diren erroreak, lema/hitz mailakoak.

1.1. **LEtra(k) Kentzea (ORT_LEK):** *euki, *muitu, *biotza, *hoiek...

1.1.1. **LEtra Kentzea Hasieran, H (LEKHH):** *erri, *emen, *artu...

1.1.2. **LEtra Kentzea Hasieran, E (LEKHE):** *ta...

1.1.3. **LEtra Kentzea Tartean, H (LEKTH):** *biotza...

1.1.4. **LEtra Kentzea Tartean, D (LEKTD):** *euki...

1.1.5. **LEtra Kentzea Tartean, G (LEKTG):** *muitu...

1.1.6. **LEtra Kentzea Tartean, R (LEKTR):** *hoiek...

1.1.7. **LEtra Kentzea Bukaeran, R (LEKBR):** *ze...

1.2. **LEtra(k) Gehitzea (ORT_LEG):** *haundi, *bainan, *zehatzki...

1.2.1. **LEtra Gehitzea Hasieran, H (LEGHH):** *harrisku, *heman...

1.2.2. **LEtra Gehitzea Tartean, I (LEGTI):** *laister...

1.2.3. **LEtra Gehitzea Tartean, U (LEGTU):** *haundi...

1.2.4. **LEtra Gehitzea Bukaeran, N (LEGBN):** *bainan...

1.3. **LEtra(k) Ordezkatzea (ORT_LEO):** *eritzi, *inbiria, *ospatsuak, *piskanaka, *araso...

1.3.1. **LEtra Ordezkatzea Edozein tokitan, E beharrean A (LEOEAA):** *ospatsuak...

1.3.2. **LEtra Ordezkatzea Edozein tokitan, A beharrean E (LEOEAE):** *erreztasun...

1.3.3. **LEtra Ordezkatzea Edozein tokitan, O beharrean U (LEOEOU):** *burruka...

1.3.4. **LEtra Ordezkatzea Edozein tokitan, U beharrean O (LEOEUO):** *partido...

1.3.5. **LEtra Ordezkatzea Edozein tokitan, D beharrean R (LEOEDR):** *inbiria...

1.3.6. **LEtra Ordezkatzea Edozein tokitan, H beharrean J (LEOEJH):** *zijoala...

1.3.7. **LEtra Ordezkatzea Edozein tokitan, J beharrean G (LEOEJG):** *imaginazioan...

1.3.8. **LEtra Ordezkatzea Edozein tokitan, N beharrean M (LEOENM):** *trampa...

1.3.9. **LEtra Ordezkatzea Edozein tokitan, X beharrean J (LEOEXJ):** *keja...

1.3.10. **LEtra Ordezkatzea Edozein tokitan, S beharrean Z (LEOESZ):** *garrazia...

1.3.11. **LEtra Ordezkatzea Edozein tokitan, S beharrean X (LEOESX):** *experientzia...

1.3.12. **LEtra Ordezkatzea Edozein tokitan, Z beharrean S (LEOEZS):** *arasoak...

1.3.13. **LEtra Ordezkatzea Edozein tokitan, Z beharrean X (LEOEZX):** *baxter...

1.3.14. **LEtra Ordezkatzea Edozein tokitan, X beharrean S (LEOEXS):** *piskanaka...

1.3.15. **LEtra Ordezkatzea Edozein tokitan, X beharrean Z (LEOEXZ):** *gaizotasunez...

1.3.16. **LEtra Ordezkatzea Edozein tokitan, Z beharrean TX (LEOEZTX):** *txorakeriak...

1.4. **LEtra(k) Ordenaz Aldatzea (ORT_LEOA):** *lehioa, *leihaketa...

1.4.1. **LEtra Ordenaz Aldatu, H (ORT_LEOAH):** *lehioa, *ohiana, *leihaketa...

1.5. **A ORGANIKOA kentzea (ORT_AORG):** *Donosti, *panpin, *bidai...

1.6. **Hitzak ELKartzea edo BANatzea:**

1.6.1. **Hitzak ELKartzea (ORT_ELK):** *hitzegin...

1.6.2. **Hitzak BANatzea (ORT_BAN):** *bait du...

1.7. **Errore orto(tipo)grafikoak (ORTOTIPO):**

1.7.1. **Marratxoaren erabilera okerra, hitz elkartuetan:**

1.7.1.1. **Hitz elkartuetan MARratxoa Kentzea (ORTOTIPO_MARK):**
*jaun andre, *anai arreba...

1.7.1.2. Hitz elkartuetan **MAR**ratxoa Gehitzea (ORTOTIPO_MARG):

**hitz-elkartu...*

1.7.2. **MAI**uskula eta **min**uskulen erabilera okerra (ORTOTIPO_MAI-min): **joxe garin, *. etxera...*

1.8. Bestelakoak

2. Errore lexikalak: Elementu lexiko erroredunak (asmatutako edota gaizki sortutako hitzak, esapide erroredunak, egun jada erabiltzen ez direnak...).

2.1. Errore **LEX**ikoa (**LEX**) (asmatutako lexiko erroreduna): **afaltzaile, *haurtoki...*

2.2. Lexiko **EZ-ERAB**ilia (**LEX_EZ-ERAB**): **urrutizkina, *orlegi, *azalburu...*

2.3. Maileguen egokitzapen okerrak (**LEX_MAIL**): **afamatu, *heredentzia...*

2.4. Aldaki dialektatak (**LEX_DIAL**): **nere, *haundi, *barri, *ipui, *inbidiria...*

2.5. Bestelakoak

3. Gramatika-erroreak: Gramatika-arauren bat bortxatzen duten hitzak edo segidak (egitura ez-gramatikalak).

3.1. Errore morfologikoak

3.1.1. Izen oinarriari dagozkionak: **erlien, etxian...*

3.1.2. Aditzari dagozkionak: **leike, *mirestu, *gordetu, *jasatzen, *dabilzate...*

3.2.- Deklinabide-erroreak

3.2.1. **DEKL**inabide-atzizkia **NAH**astea (**DEKL_NAH**): **lagunaz dago, Jonen autoarekin etorri naiz...*

3.2.2. **DATA**-egitura erroredunak (**DEKL_DATA**): **Donostia, 1995eko urtarrilak 15ean...*

3.3. Determinatzaile-erroreak

3.3.1. **DE**terminatzailea **Kent**zea (**DETK**): **txokolate nahi dut; *opari erosi dizkiot...*

3.3.2. **DE**terminatzaile atzizkia **Gehit**zea (**DETG**): **nahi adina dirua dauka...*

3.3.3. **DE**terminatzailea **Errepikat**zea (**DETE**): **mendia bat ikusten dut; *gizona zaharra bat...*

3.3.4. **DS**ko osagaien **Ordena** okerra (**DETO**): **gutxi pertsona ginen; *asko lan egiten du...*

3.3.5. **MUG**atasun kontuak (**DETMUG**): **edozein aldizkarietan irtetzen da; *edozein unean...*

3.3.6. **DE**T/Zenbatz. **Nah**astea (**DETN**): **asko beroa egin zuen; *jende osoak daki...*

3.3.7. **DETG** eta **DETK** sintagma berean (**DETGK**): **mozkorra handi hartu genuen...*

3.3.8. Anbiguoak diren kasuak (**DET?**): **hamar garagardoak erosi nituen...*

3.3.9. Zalantzaoko kasuak (**DET**)

3.4. Izenordainen inguruan egindako erroreak: **bere buruari ikusi da...*

3.5. Adjektibo eta Adberbioen inguruko erroreak: **hobe egin du...*

3.6. Aposizio-erroreak: **Zure lagunari, Dublinen bizi dena, sari bat eman diote...*

3.7. Postposizio-erroreak: **Izaskun buruz hizketan ibili da...*

3.8. Komuntadura-erroreak

3.8.1. **KOM**untadura eza **SIN**tagma barruan (**KOMSIN**): **gurasoak eta lagunez mintzatu zara; *guk geu...*

3.8.2. **KOM**untadura eza **APO**sizioan (**KOMAPOS**): **zure laguna, Dublinen bizi denari, sari bat eman diote...*

3.8.3. **KOM**untadura eza **Per**pausean (**KOMP**)

3.8.3.1. Aditza – Subjektua (**KOMPAS**)

3.8.3.1.1. **KOM**untadura eza **Per**pausean, **Aditza** eta **Subjektua**ren artean, **NUM**eroari dagokionean (**KOMPAS-NUM**): **aurrerapen haundia daude; *gizonek egin du...*

3.8.3.1.2. **KOM**untadura eza **Per**pausean, **Aditza** eta **Subjektua**ren artean, **KAS**uari dagokionean (**KOMPAS-KAS**): **zuk etorri zara; *nik esnatu naiz...*

- 3.8.3.2. Aditza – Objektua:
 3.8.3.2.1. **KOMunztadura** eza **Perpausean**, **Aditza** eta **Objektuaren** artean, **NUMeroari** dagokionean (**KOMPAO-NUM**): **eman dizut liburua; *nik etxeak ikusi dut...*
 3.8.3.2.2. **KOMunztadura** eza **Perpausean**, **Aditza** eta **Objektuaren** artean, **KASuari** dagokionean (**KOMPAO-KAS**): **nik etxeak ikusi ditut...*
- 3.8.3.3. Aditza – Zehar-objektua
 3.8.3.3.1. **KOMunztadura** eza **Perpausean**, **Aditza** eta **Zehar Objektuaren** artean, **NUMeroari** dagokionean (**KOMPAZO-NUM**): **ziberespazioan dabiltzan pertsonen dagokion izena da; *emaitzei dagokiona...*
 3.8.3.3.2. **KOMunztadura** eza **Perpausean**, **Aditza** eta **Zehar Objektuaren** artean, **KASuari** dagokionean (**KOMPAZO-KAS**): **nik haiek eman diet...*
- 3.8.3.4. Aditza – Predikatua
 3.8.3.4.1. **KOMunztadura** eza **Perpausean**, **Aditza** eta **Predikatuaren** artean (**KOMPAP**): **gure erleak oso soziablea dira...*
- 3.8.4. **KOMunztadura** eza **Mendekoetan** (**KOMM**): **goxoki asko jaten duen umeek dute...*
- 3.8.5. **KOMunztadura** eza **Testuinguruan** (**KOMT**):
- 3.9. Aditzaren inguruan egindako erroreak
 3.9.1. Aditzaren Denbora, Aspektua edota Modua nahastea: **goaz mendira?...*
 3.9.2. Aditz-paradigmen nahasketa
 3.9.2.1. **Nor** eta **Nor-NorK** aditz-paradigmak nahastea (**PARADIG_N_N-NK**): **ez da funtzionatzen...*
 3.9.2.2. **Nor** eta **Nor-NorI** aditz-paradigmak nahastea (**PARADIG_N_N-NI**): **nagusia zuzendu da...*
 3.9.2.3. **Nor-NorK** eta **Nor-NorI** aditz-paradigmak nahastea (**PARADIG_N-NI_N-NK**): **niri hori ez zait molestatzen...*
 3.9.2.4. **Nor-NorK** eta **Nor-NorI-NorK** aditz-paradigmak nahastea (**PARADIG_N-NK_N-NI-NK**): **Joni ikusi diot...*
 3.9.2.5. **Nor-NorI** eta **Nor-NorI-NorK** aditz-paradigmak nahastea (**PARADIG_N-NI_N-NI-NK**): **gustatzen dit...*
- 3.10. Mendeko perpaus erroredunak
 3.10.1. Konpletiboak (**MEN-KON**): **ez dut uste etorriko dela...*
 3.10.2. Zehar-galderak (**MEN-ZG**): **ez dakit nor da; *galdetu ea joango bada...*
 3.10.3. Helburuzkoak (**MEN-HEL**): **etxera noa afaltzeko; *paseatzeko joan dira...*
 3.10.4. Kausazkoak (**MEN-KAU**): **zergatik ez zinen joan? zergatik ez neukan gogorik...*
 3.10.5. Baldintzazkoak (**MEN-BAL**): **edukiz gero dirua; *kontuz ez ibiliz gero...*
 3.10.6. Erlatibozkoak (**MEN-ERL**): **mutil bat etorri da zu ezagutzen zaituela; *nik eman dizut liburua polita da...*
 3.10.7. Denborazkoak (**MEN-DEN**): **helduko denean, abisatu; *ikusiko dudanean esango diot...*
 3.10.8. Kontzesiboak (**MEN-KONT**): **izan arren berandu, joan egingo naiz; *ez bada ere oso handia, guztiok sartuko gara...*
 3.10.9. Moduzkoak (**MEN-MOD**): **Jon zu bezala da; *ez dira uste nuen bezala...*
 3.10.10. Konparaziozkoak (**MEN-KONP**): **zuk baino dirua gehiago daukat...*
 3.10.11. Nominalizazioa (**MEN-NOM**): **ezin nuen liburua irakurtzen...*
- 3.11. Juntagailuen eta lokailuen inguruan egindako erroreak
- 3.12. Perpausen egitura okerra (hitzen ordena okerra, sintaxiari dagokionean): **jakin dudanez auzokide baten bitartez Udalak dirua eskaintzen du; arrantza motak erabiltzen zirenak; oso erle fina ez baitzen...*
- 3.13. Bestelakoak

4. Puntuazio-ikurrak: Puntuazio-ikurren erabilerari dagozkion erroreak.

- 4.1. Puntuazio-Ikurra behar ez denean Erabiltzea (PI_E)
- 4.2. Puntuazio-Ikurra behar denean Ez Erabiltzea (PI_EE)
- 4.3. Puntuazio-Ikurrak Ordezkatzea edo nahastea (PI_O)
- 4.4. Bestelakoak

5. Estilo kontuak: Estilo aldetik hobetu daitezkeen egiturak.

- 5.1. ESALdi LUZeegiak (ESALUZ)
- 5.2. HITZen ORDena, Estiloari dagokionean (HITZOR-E)
- 5.3. MAIZegi erabiltzen/errepikatzen diren hitzak/egiturak (MAIZ)
- 5.4. EKIDIIn diren egiturak (EKIDI)
- 5.5. Bestelakoak

6. Errore semantiko - pragmatikoak: Testuinguruan edota esanahiari dagokionean zentzurik ez duten elementuak.

- 6.1. Hitz bat beste batekin nahastea: *froga – proba, haziera – hazkuntza...*
- 6.2. Sasi-adiskideak (euskarazko hitz bat beste hizkuntza bateko antzeko hitz/egitura batekin nahastea; forma bera edo berdintsua izan arren, esanahia oso ezberdina duena): *xelebre* (célebre), *azienda* (hacienda), **mundiala* (mundu mailakoa esateko)...
- 6.3. Pare dikotomikoak (fonetikoki berdinak ala antzekoak diren hitzak, baina ezberdin idatzi eta esanahi ezberdina dutenak erabiltzea, testuinguru desegokian): *hura – ura; ari – hari; hasi – hazi, atso – atzo...*
- 6.4. Kalko okerrak (hizkuntza batek beste baten eredutik berera egokitzen duen egitura): **siesta bota, *ilea hartu...*
- 6.5. Esamolde okerrak (esamolde edo egitura zuzenak ez dagokien egoeratan/mementotan erabiltzea): **lur eta zur...*
- 6.6. Errore pragmatikoak: *jaten ari den bati ‘bejondeizula’; arratsaldean ‘egun on’, tresna – traste – gailu...*
- 6.7. Bestelakoak

C eranskina:

Determinatzaile-erroreen sailkapena

3.3. Determinatzaile-errorea

3.3.1. DETK (DETerminatzailea Kendu)

- 3.3.1.1. IZE + DET kendu
- 3.3.1.2. IZE + ADJ + DET kendu
- 3.3.1.3. IZE + Zenbatzaile Orokorra + DET kendu
- 3.3.1.4. DETK aditz-esapideetan
- 3.3.1.5. DETK postposizio-egituretan
- 3.3.1.6. DETK izenki-predikatuei
- 3.3.1.7. DETK hitza deklinatzean

3.3.2. DETG (DETerminatzaile atzizkia Gehitu)

- 3.3.2.1. Determinatzaile atzizkia erabiltzea behar ez denean
- 3.3.2.2. DETG aditz-esapideetan
- 3.3.2.3. DETG postposizio-egituretan
- 3.3.2.4. DETG izenki-predikatuei
- 3.3.2.5. DETG hitza deklinatzean

3.3.3. DETE (DETerminatzailea Errepikatu)

- 3.3.3.1. IZE + DETatzizkia + ADJ + DETatzizkia
- 3.3.3.2. IZE + DETatzizkia + Artikulu partitiboa
- 3.3.3.3. IZE+DETatzizkia + Artikulu zehaztugabea
- 3.3.3.4. IZE + DETatzizkia + ADJ + Artikulu zehaztugabea
- 3.3.3.5. IZE + ADJ + DETatzizkia + Artikulu zehaztugabea
- 3.3.3.6. IZE + DETatzizkia + ADJ + DETatzizkia + Artikulu zehaztugabea
- 3.3.3.7. IZE + DETatzizkia + Erakuslea
- 3.3.3.8. IZE + DETatzizkia + ADJ + Erakuslea
- 3.3.3.9. IZE + ADJ + DETatzizkia + Erakuslea
- 3.3.3.10. IZE + DETatzizkia + ADJ + DETatzizkia + Erakuslea
- 3.3.3.11. IZE + DETatzizkia + Zenbatzaile Orokorra + DETatzizkia
- 3.3.3.12. IZE + DETatzizkia + Zenbatzaile Zehaztugabea
- 3.3.3.13. Zenbatzaile zehaztugabea + IZE + DETatzizkia
- 3.3.3.14. Zenbatzaile zehaztugabea + IZE + DETatzizkia + ADJ
- 3.3.3.15. Zenbatzaile zehaztugabea + IZE + ADJ + DETatzizkia
- 3.3.3.16. Zenbatzaile zehaztugabea + IZE + DETatzizkia + ADJ + DETatzizkia
- 3.3.3.17. Determinatzaile nolakotzaile/galdetzaile/zehaztugabea + IZE + DETatzizkia
- 3.3.3.18. Determinatzaile nolakotzaile/galdetzaile/zehaztugabea + IZE + DETatzizkia + ADJ + DETatzizkia
- 3.3.3.19. Zenbakia + IZE + DETatzizkia
- 3.3.3.20. Zenbakia + IZE + DETatzizkia + ADJ
- 3.3.3.21. Zenbakia + IZE + ADJ + DETatzizkia
- 3.3.3.22. Zenbakia + IZE + DETatzizkia + ADJ + DETatzizkia

3.3.4. DETO (DSko osagaien ordena okerra)

- 3.3.4.1. Zenbatzaile Zehaztugabea + IZE
- 3.3.4.2. Artikulu zehaztugabea + IZE

3.3.5. DETMUG (MUGatasun kontuak)

- 3.3.5.1. Zenbatzaile Zehaztugabea + IZE + DETatzizkia deklinatuta
- 3.3.5.2. Determinatzaile Zehaztugabea + IZE + DETatzizkia deklinatuta
- 3.3.5.3. Zenbakia + IZE + DETatzizkia deklinatuta

3.3.6. DETN (DETerminatzailea edo zenbatzailea Nahastu)

3.3.7. DETGK (DETG eta DETK sintagma berean)

3.3.8. DET? (era isolatuan zuzentzat eman daitezkeen arren testuinguruan erroredunak direnak)

3.3.9. DET (zalantzazko kasuak)

D eranskina:

Euskara-ikasleen testuetan etiketatutako determinatzaile-erroreak⁴³

3.3.1. DETK (DETerminatzailea Kendu)

3.3.1.1. IZE + DET kendu

Erregelak: DETK1, DETK1A, DETK1B, DETK1C, DETK4A, DETK5A1, DETK5A2, DETK5A3, DETK5A4, DETK5B1, DETK5B2, DETK5C1, DETK5C2, DETK5D1, DETK5D2, DETK6A, DETK6B

<DETK>Txupinazo<DETK> eta gero peñara joaten gara (ae98_2_ac_a)

<DETK>Pote<DETK> hartu genuen (ae98_2_ao_a)

Ni egun batean <DETK>banko<DETK> lapurtzeko du (ae98_2_bx_a)

komunera joan nengoen eta <DETK>aurpegi<DETK> garbitu dut (ae98_2_l_a)

Bapatean nire lagunak bota zuen <DETK>autzontzi<DETK> (ae98_2_r_a)

eta <DETK>perretxiko<DETK> bota nituen (ae98_2_bs_a)

Guk <DETK>kamping<DETK> irekitzen genuen. (ae98_2_ao_a)

notak hartu nuen eta <DETK>sobre<DETK> ez nuen ireki (ae98_2_k_a)

nire ama oso larrituta zegoela eta <DETK>sobre<DETK> ireki nuen (ae98_2_k_a)

jatetxean <DETK>marisko<DETK> jan genuen (ae98_2_bh_a)

ondoren biok <DETK>perretxiko<DETK> jan ditugu (ae98_2_bs_a)

<DETK>boleto<DETK> hartu nuen (ae98_2_af_a)

eta ez zegoen <DETK>iturri<DETK> bestela tabernan joan ginen (ae98_2_be_a)

gero <DETK>eguzki<DETK> hartu nuen (ae98_2_bm_a)

Arratsaldean hondartzara joan nintzen eta <DETK>eguzki<DETK> hartu nuen (ae98_2_af_a)

betidanik <DETK>piano<DETK> jotzen du (ae98_2_bb_a)

Gerostik <DETK>piano<DETK> jotzea erabaki nuen (ae98_2_bb_a)

<DETK>txis<DETK> egin nuen (ae98_2_bs_a)

garbitu platera dut, <DETK>ohe<DETK> egin, lehioak garbitu dut... (ae98_2_bo_a)

Orduan hasi da nire <DETK>egun<DETK> (ae98_2_n_a)

ni eta nire <DETK>lagun<DETK> tabernara joan ginen (ae98_2_ah_a)

Baina nire <DETK>irakasle<DETK> ezin zuen ulertu hori (ae98_2_ay_a)

Nire <DETK>aurpegi<DETK> garbitu dut (ae98_2_bo_a)

Itziarrekin nire <DETK>soineko<DETK> jantzi nuen (ae98_2_g_a)

Nire laguna eta nik <DETK>txokolate<DETK> txurroekin jan ginen (ae98_3_y_a)

Niri <DETK>itxaso<DETK> asko gustatzen zait (ae98_3_b_a)

egun osoan <DETK>euri<DETK> ari zela (ae98_3_g_a)

<DETK>Batxillergo<DETK> bukatu eta gero nire ametsak betetzea hasi ziren (ae98_4_z_a)

<DETK>Partido<DETK> bukatu eta gero joan ginen afaltzerako (ae98_4_al_a)

bokatak jan genituen eta gero <DETK>euri<DETK> hasi zen (ae98_4_y_a)

Bazkaria bukatu zuenean ipuin <DETK>kontalari<DETK> etorri zen (ae98_4_ab_a)

ilargi gabe ikusteko obeto da, eta <DETK>ilargi<DETK> ez zegoen (ae98_4_am_a)

<DETK>Partido<DETK> hasi zenean ni jada oso urduri nengoen (ae98_4_al_a)

Seiak ziren eta <DETK>partido<DETK> ez zen hasi zaspi ta erditaraino (ae98_4_al_a)

Hotelean batzuk gela batera joan ginen <DETK>txokolate<DETK> erretzera (ae98_4_af_a)

nire lehen <DETK>porro<DETK> erre nuenean (ae98_4_o_a)

poliki poliki dendak ta hori rekojitu genuen eta <DETK>trabesi<DETK> egin berriro (ae98_4_am_a)

<DETK>Jende<DETK> han lan egiten duena hizkuntza batzuk jakin dute (ae98_4_q_a)

futbolinak eta dardok jolastu dugu bitartean <DETK>garagardo<DETK> edaten dugu (ae98_4_ñ_a)

<DETK>Diborzio<DETK> eginez geroztik beste gizona harrapatzera hasi nintzen (ae98_4_ak_a)

Ni beti beti haserre nenbilen <DETK>"uniforme"<DETK> eramatera bultzatu baitzidaten (ae98_4_b_a)

zaia zen halako <DETK>espermatozoide<DETK> lortzea (ae98_4_g_a)

⁴³ Adibideak errore motaren eta euskaltegiaren arabera sailkatuta daude: koloreek euskaltegia adierazten dute, hau da, euskaltegi bakoitzari kolore bat dagokio.

Horregatik ez da beharreskoa zuk <DETK>frantses<DETK> esagutzen duzula (ae98_4_q_a)
ondartza joan zen eta jolastu genuen, <DETK>gatzelu<DETK> egin genuen (ae98_4_t_a)
eta <DETK>hamaiketako<DETK> hartzera nire lagunak konbidatu dut (ae98_4_i_a)
ondoren <DETK>afari<DETK> prestatu genuen eta edaria erosi genuen (ae98_4_c_a)
eta sagardotegia batean, kriston <DETK>giro<DETK> dugu (ae98_4_i_a)
Bruselen ordu bat esperatu nuen eta beste <DETK>hegazkin<DETK> hartu nuen (ae98_4_r_a)
nire <DETK>helburu<DETK> lortu izan baitnuen (ae98_4_ap_a)
Nire <DETK>urtebetetxe<DETK> ailegatu zenean, lagun batek deitu zidan (ae98_4_ag_a)
Nire <DETK>amets<DETK> errealitatea egin zuelako (ae98_4_r_a)
Bidaia oso txarra zen, gure <DETK>kotxe<DETK> oso zaharra zelako (ae98_4_m_a)
Nire <DETK>amets<DETK> da eskaraz bizitzen orduan oso ongi sentituko naiz (ae98_4_ad_a)
orain lan egiten dut baina ez da nire <DETK>lanbide<DETK> (ae98_4_ao_a)
Ondartzara allegatu ginen eta gure <DETK>"kanpamento"<DETK> jarri genuen (ae98_4_t_a)
Ayamonte herri zoragarria zen, bere <DETK>jende<DETK> oso jatorra zen (ae98_4_ax_a)
Gure lehenengo <DETK>hiri<DETK> Tanger izan zen (ae98_4_af_a)
<DETK>Kontzertu<DETK> ta gero senide batzuk txarla bat bota zuen (ae98_5_k_a)
"el de Trizio" aurpegi desenkatatua zuen <DETK>partido<DETK> bukatu ta gero (ae98_5_i_a)
Azkar, azkar afaltzeko <DETK>ordu<DETK> iritsi zen (ae98_5_k_a)
<DETK>Eguzki<DETK> sartu zenean joan ginen (ae98_5_as_a)
Gogoia geneukan juerga botatzeko eta hasi ginen <DETK>sagardo<DETK> edaten (ae98_5_z_a)
<DETK>Sagardo<DETK> edatera, kontzertu entzutera eta dantzatzera elkarrekin joan ginen (ae98_5_ay_a)
Sagardo edatera, <DETK>kontzertu<DETK> entzutera eta dantzatzera elkarrekin joan ginen (ae98_5_ay_a)
<DETK>Kontzertu<DETK> bukatutakoan kalera tabernaz beteta joan ginen (ae98_5_ax_a)
hamarretan zezen platzara joan ginen <DETK>kontzertu<DETK> ikustera (ae98_5_ax_a)
Adibidez <DETK>partidu<DETK> bukatu zuen uste dut 7 laurden gitxietan (ae98_5_s_a)
atea ireki nuen eta <DETK>tipo<DETK> sartu zen (ae98_5_af_a)
Orain dela hiru hilabete <DETK>kotxe<DETK> erosi nuen (ae98_5_m_a)
joan nintzen nire aitarekin <DETK>kotxe<DETK> erostera (ae98_5_b_a)
Gero guztiak joan guinen <DETK>kotxe<DETK> frogatzera (ae98_5_b_a)
hotelez hotel joan ginen <DETK>leku<DETK> bilatzeko (ae98_5_ax_a)
eta poz pozik eskegi nuen <DETK>telefono<DETK> (ae98_5_ai_a)
Orain dela bi urte <DETK>afari<DETK> prestatu nuen (ae98_5_m_a)
pertsonek batzuk <DETK>bainu<DETK> artzera joan ginen (ae98_5_n_a)
Gure <DETK>irakasle<DETK> animoak eman zigun (ae98_5_be_a)
Momentu batean nire <DETK>paradisu<DETK> desagertu izan zen (ae98_5_l_a)
<DETK>Klase<DETK> eta gero lehen bait lehen zuzendariaren gela joan nintzen (ae98_6_g_a)
Denbora <DETK>libre<DETK>, denbora <DETK>libre<DETK> eta denbora <DETK>libre<DETK>! (ae98_6_s_a)
poz pozik itsasoratu ginen, <DETK>arraun<DETK> artuta eta gogoekin (ae98_6_q_a)
eta gero jendea etortzen da <DETK>kontzertu<DETK> suspendituta zegoelarik (ae98_6_a_a)
<DETK>Leku<DETK> hartuz gero kampa atera ginen (ae98_6_h_a)
ondartzara <DETK>eguzki<DETK> hartzera joan ginen (ae98_6_n_a)
<DETK>Hamaiketako<DETK> jan eta gero, zezen plazara joan ginen (ae98_6_n_a)
nik hirugarren <DETK>bainu<DETK> itsasoan hartu erabaki nuen (ae98_6_i_a)
gero beste <DETK>lagun<DETK> igo zen eta azkenean nire txanda zen (ae98_6_m_a)
nire <DETK>bizi<DETK> aldatu nuen (ae98_6_p_a)
gaua heldu zen eta neska bere <DETK>amodio<DETK> aitortu zidan (ae98_6_o_a)
<DETK>Txkeo<DETK> bukatutakoan esan zioten ingresatu behar zutela (ae98_7_v_a)
<DETK>Landetxe<DETK>, oso ahundia zen eta oso polita gainera. (ae98_7_n_a)
<DETK>Une<DETK> iritsi zenean baietz esan oso lasaia geratu zen (ae98_7_n_a_bis)
<DETK>Sexu<DETK> frogatu gabe geratu nintzela uste nuen (ae98_7_c_a)
nire etxearen ondoan <DETK>sex-shop<DETK> ireki zuten egunean izan zela (ae98_7_aa_a)
<DETK>txokolat<DETK> txurroekin gozaltzera joan ginen (ae98_7_x_a)
Bazkaldu eta gero <DETK>zurrapapote<DETK> prestatu genuen (ae98_7_j_a)
poteatzen eman genuen gau osoa, <DETK>egun<DETK> argitu arte (ae98_7_x_a)
Hamaiketan "peña" batera joan ginen <DETK>hamaiketako<DETK> egitera (ae98_7_x_a)
baina bere <DETK>soineko<DETK> aurkitzeko ez dakit zen bat buelta ematen behar zuen (ae98_7_n_a_bis)
Costa Rikan pasatzen ginenean gure <DETK>kotxe<DETK> apurtu zitzaion (ae98_7_o_a)
baina Elena, gure <DETK>irakasle<DETK>, ez zuen nahi (ae98_7_a_a)
Nire <DETK>auzokide<DETK> oso polita eta gaztea zen (ae98_7_c_a)
<DETK>Kontzertu<DETK> bukatu eta gero Iruñeko Alde Zaharrera joan ginen (ae98_8_g_a)
Amak edo aitak, seme-alabek hartzaindegira eramatean <DETK>lanbide<DETK> praktikatu ahal du (ir90_6_e_d)
Eta egunero <DETK>ordutegi<DETK> gutxitu ahal da (ir90_6_e_d)
Lantegi batzuk <DETK>hartzaindegia<DETK> dauka (ir90_6_e_d)
bere herritik berak <DETK>pultso<DETK> hartzen duelako (ir90_6_p_k)
lapurtutako <DETK>material<DETK> nabarmentzen dizut (ir90_6_n_a)
<DETK>txakur<DETK> eduki badauzka bi, gainera (ir90_6_a_a)

Herri honetatik begirada partikular batez <DETK>mundu<DETK> ikusten da (ir90_6_p_k)
Marraskiak <DETK>Jon Zabaletaren<DETK> dira (ir90_6_i_g)
Nahiz eta haurrentzako <DETK>idatzitako<DETK> izan (ir90_6_d_g)
Poloniako irakasle baten <DETK>zuzenbide<DETK> lortu zuen (ir90_6_d_e)
Telebista, video eta abar beharrezkoa da Euzkera <DETK>klase<DETK> emateko (ir90_6_f_a)
Berak proposatzen dio plan bat bere <DETK>arazo<DETK> konpontzeko (ir90_6_h_g)
Edonorren heriotzak gure <DETK>errespetu<DETK> merezi du (ir90_7_e_a)
beste "play-boy" batek, <DETK>italiar eta gaztea<DETK> orain, "ligatu" zuela (ir90_7_d_a)
denon artean "Atxaga" <DETK>aldizkari<DETK> aurrera eramatea lortuko dugu (il92_7_i_a)
laguntza emango bazenigu guk <DETK>aldizkari<DETK> edukiko genuke (il92_7_a_a)
Aldizkariaren bidez geure <DETK>lan<DETK> errezagoa da (il92_7_a_a)
Aldizkari honen bidez Drogak eta bere <DETK>arisku<DETK> jakitera emango diegu (il92_7_c_a)
AMESGAIZTOA : <DETK>Gau<DETK> azkar iragan zen, azkarregi (il92_8_g_d)
Iha ezin zuten <DETK>janari<DETK> lortu (il92_8_o_a)
Gure gizonak ordez, <DETK>aurpegi<DETK> emateko ohitura du (il92_8_b_a)
Gizakiak, joera natural bezala, <DETK>osasun<DETK> hobetzea izan du (il92_8_c_d)
Egoaldekoak ezin dira bizi bere herrietan, <DETK>janari<DETK> ez daukate eta (il92_8_y_e)
<DETK>Hamaiketako<DETK> egitea lagunarekin joan zinateke (il92_8_j_a)
egurraren bila joaten den <DETK>baso<DETK> gero eta urrutiago daudela iruditzen zaio (il92_8_t_a)
<DETK>gizon<DETK> andreza amorratua zela esaten da (il92_8_q_e)
Ez haserre izeba, hau da gaurko <DETK>bizimodu<DETK> (il92_8_q_d)
Bere <DETK>haurtzaro eta nerebezaro<DETK> jaioterrian igaro (izan) zituen (il92_8_a_d)
nere <DETK>talde<DETK> hirugarren pisuko zazpigarren gelan zegoela esan zidaten (il92_8_i_c)
Joan den astean erizaina bat kirofanotik <DETK>egotzi (expulsada) <DETK> izan zen (il92_8_x_e)
<DETK>Txiste<DETK> kontatzeak badu filosofiaren bat (il92_9_n_f)
Kontatzaile batek <DETK>txiste<DETK> kontatzen joan ahala (il92_9_n_f)
<DETK>ilargi<DETK> mugitzeari ekiten diote (il92_9_m_i)
izugarritzko <DETK>oinaze<DETK> ba zuen ere aurpegian alaitasuna ikusten zen (il92_9_c_d)
<DETK>esne<DETK> erosterako bost kilometro ibili behar zuen (il92_9_f_b)
Gainera biek <DETK>uniforme<DETK> zeramaten (il92_9_e_d)
Jateko <DETK>keinu<DETK> badago, dirurik eskatzeko edo emateko baita (il92_9_f_g)
Esnea berotzeko <DETK>kanpin-gas<DETK> erabiltzen genuen (il92_9_h_c)
egunak pasa ziren neurrian <DETK>pientso<DETK> gutxitu egin zion (il92_9_c_b)
<DETK>nahiko<DETK> jandakoan itzuli estera (il92_9_m_a)
<DETK>txiste<DETK> bukatzean edo bukatu baino lehen parrez hasten den kontalaria (il92_9_b_f)
Iluntzean <DETK>artalde<DETK> itzuli zen (il92_9_d_a)
Ni bezain <DETK>luze<DETK> ikusten dut (il92_9_l_d)
<DETK>jende<DETK> bere etxean atera gabe zegoen (il92_9_c_b)
Lehen egunian <DETK>pientso besterik<DETK> ez zion eman (il92_9_c_b)
Badirudi 1000 pertsona bere <DETK>lanpostu<DETK> galdu egin duela (il92_9_l_e)
Badakite bere <DETK>antzerki<DETK> egiten (il92_9_e_c)
artzain batek bere <DETK>artalde<DETK> atera zuen lurrarena (il92_9_d_a)
Horrela egin zuen eta bere <DETK>artalde<DETK> larrera eraman zuen (il92_9_c_a)
datorren astean <DETK>klase<DETK> daukagu (il95_9_g_d)
bukatu genuenean <DETK>jende<DETK> txaloka hasi zen (il95_9_g_b)
<DETK>Erbeztzeratu<DETK> izaten zen ere bai (il95_9_j_a)
nahi baina ez, gure <DETK>andereino<DETK> ez da horrelakoa (il95_9_g_d)
<DETK>laburpen<DETK> egingo bagenu pobreagoak izatea dagokigula esango genuke (il95_10_s_c)
Britainia Handian aurre egin baitzakete <DETK>indar<DETK> hartuz. (il95_10_a_e)
Gai honek <DETK>hitz jario<DETK> dakarkigu (il95_10_a_k)
ekonomia goitik behera jarri behar bait genuke, bai gure <DETK>burumolde<DETK> ere (il95_10_a_k)
ate ondoan <DETK>haize<DETK> ematen ari ziren erleek zer egiten ote zuten? (il95_10_i_d)
medikuntzak ia-ia <DETK>mirari<DETK> egin lezake arlo honetan, (il95_10_j_g)
lurrikarak gertatzen jarraituko dituztela eta <DETK>erremedio<DETK> geldiarazteko ez ditugula (il95_10_j_f)
bi buelta dago <DETK>lehendakari<DETK> aukeratzeko (il95_10_s_g)
<DETK>Erabaki<DETK> hartzean alde honak edo txarrak miatzen hasi ziren (il95_10_j_i)
Espainako ekonomia aurreratuko zuela, arazo <DETK>multzo<DETK> agertu arren (il95_10_r_h)
gero <DETK>akordeoi<DETK> tekladoarekin iritsi zen (il95_10_r_g)
lurrikarek beste <DETK>arazo<DETK> dute eta zera da (il95_10_s_f)
Erleek haizearekin beste <DETK>ondorio<DETK> bilatu dute (il95_10_s_d)
baina euskal <DETK>unibertsitate<DETK> oso berria da (il95_10_c_g)
Kazetariak eztiaren egiteko modua <DETK>kontu<DETK> balitz bezala kontatu digu (il95_10_s_d)
<DETK>Tratamendu<DETK> jarraitu arren, %3ek haurdunaldia lortzen dute (il95_10_c_f)
hizkuntzen elkarbizitzeak <DETK>garrantzi<DETK> du oso (il95_10_g_e)
otsailaren azken egunean Roldan <DETK>atxilotu<DETK> izan da (il95_10_s_a)
eserlekuaren tapizeria erretzen <DETK>hasi<DETK> zegoen (il95_10_s_j)

zergatik gure <DETK>Vaticano<DETK> kontra ote dago? (il95_10_j_g)
 zenbaiten ustearen arabera gehiegizko <DETK>potere<DETK> lortu baitute (il95_10_g_k)
 Garraio publikoetan joatea kotxea joatea baina <DETK>merkeago<DETK> ateratzen da (il95_12_c_a)
 <DETK>erreproduzio<DETK> egitea (il95_12_e_d)
 Zer nolako <DETK>lan<DETK> egingo lezake horrelako umetxok? (il95_12_m_c)
 <DETK>Sorte<DETK> izanda, istripueen osorioak arinak izan daitezke (il95_12_c_a)
 gasolina edo <DETK>gasolio<DETK> bota behar denean (il95_12_c_a)
 Iparraldeko hondartzek, ur kalitatean, <DETK>ospe<DETK> dute (il95_12_c_e)
 antzinako gizonen margotzeko edo <DETK>arte<DETK> egiteko ohiturei buruz (il95_12_e_d)
 Hasteko, jokalarien <DETK>kopuru<DETK> haundia da. (il97_10_k_l)
 Goian <DETK>aipatutako<DETK> ukatu gabe, dirudienez futbola zerbait gehiago da (il97_10_k_l)
 Horrek esan nahi du, <DETK>sensibilitate<DETK> badaukadala oinetan (il97_10_b_n)
 boletoa garestiagoak izango dira eta guk <DETK>diru<DETK> barra-barra irabaziko dugu (il97_10_b_o)
 <DETK>klase<DETK> emateko behar adina (il97_10_h_m)
 Neska polita izan ezezik, <DETK>inteligente<DETK>, libre eta idealista hutsa ere izan zen (il97_10_e_o)
 Neska polita izan ezezik, inteligente, <DETK>libre<DETK> eta idealista hutsa ere izan zen (il97_10_e_o)
 Gure <DETK>poteo<DETK> jarraitu genuen (il97_10_f_o)
 uste dut gizarte honek bere <DETK>"tribu"<DETK> ondo sartuta daukala bere odoletan (il97_10_f_l)
 Ehuneko hirurogei eta seik gehiago ikasteko <DETK>ikastaro<DETK> egingo luke (tr05_2_or5_e)
 nahiz eta oso <DETK> aurreratu<DETK> egon, gaur egungo jendea nahiko tradizionala da (tr05_12_or11_h)

Harridura esapideak

<DETK>Ze zuerte<DETK> nirea! (ae98_2_af_a)
 Ba da ordua lanera itzultzeko, eta <DETK>ze gose<DETK>!! (ae98_6_c_a)
 nirekin elkartzeko gogoak adierazten zidan, hau poza, <DETK>hau ustekabe<DETK>. (ae98_7_l_a)
 <DETK>Zer nolako suerte<DETK> ez da? (ir90_7_g_a)

3.3.1.2. IZE + ADJ + DET kendu

Erregelak: DETK2A, DETK2B, DETK4B

Urrengo goizan <DETK>erresaka haundi<DETK> nuen (ae98_2_ao_a)
 Iruñeara joaten gara afaltzeko eta <DETK>juerga haundi<DETK> egiten dugu (ae98_2_w_a)
 Beceste nahiz eta <DETK>herri txiki<DETK> izan kriston martxa dauka (ae98_4_p_a)
 Ez zen <DETK>denbora luze<DETK> igaro beste pote bat hartzera joan ginenean (ae98_5_bf_a)
 <DETK>Buelta erdi<DETK> eman nuen (ae98_6_i_a)
 Beste <DETK>egun zoriontsu<DETK> izateko lagunen artean izan beharko da (ae98_7_ac_a)
 <DETK>ile kizkur<DETK> daukala, begiak oso ilunak (il92_8_q_b)
 gero eta <DETK>garrantzi haundiago<DETK> ematen zaio bizia osasuntsua edukitzeari (il92_8_c_d)
 hirurehun pertsona hil dira <DETK>meskita musulman<DETK> erre denean (il92_8_h_e)
 <DETK>Arrakasta eskas<DETK> izan dut gaur mahai honetan, (il92_8_f_a)
 txakurrak <DETK>leku libre<DETK>, korrika egiteko behar du (il92_9_f_c)
 <DETK>sistema edukatibo<DETK> dela kausa (il92_9_f_h)
 zuzena honi buruzko <DETK>ikerketa serio<DETK> egitea izango litzateke (il95_10_i_h)
 Badakit <DETK>denbora luze<DETK> daramagula ezer jakin gabe bata besteari (il95_10_c_k)
 Ikuspuntu bakoitzak bere irakurketa, hausnartze eta <DETK>erantzun propio<DETK> behar du (il95_10_a_f)
 gure <DETK>gaitza okerren<DETK> oraindik irauntzen du: Langabezia! (il95_10_r_k)
 konponketak hastekotan, <DETK>mailegu handi samar<DETK> eskatu behar izango dugu (il95_10_i_j)
 <DETK>Baserri propio<DETK> zutenek, laguntza moduan eta (il95_12_m_c)
 eta <DETK>laguntalde txiki<DETK> lekuko bezala (haien artean Schubert zegoen) hil zen (il97_10_b_u)
 <DETK>Fundamentalismo islandar<DETK> XX. mendean sortu zen Egipton, (il97_10_n_q)
 ez dakit beste <DETK>modu hobeago<DETK> igandeko arratsaldeak igarotzeko (il97_10_b_1)
 idazteko teknikak eta <DETK>mekanika industrial<DETK> ehuneko hamaikari gustatuko litzaioke (tr05_2_or5_e)

3.3.1.3. IZE + Zenbatzaile Orokorra + DET kendu

<DETK>gau oso<DETK> pasatu ginen farrez (ae98_4_an_a)
 <DETK>gela oso<DETK> odolez beteta geratu zen (ae98_4_bb_a)

3.3.1.4. DETK aditz-esapideetan

bere etxera joan ginen eta <DETK_HAUL>larru jo<DETK_HAUL> egin genuen (ae98_6_o_a)
 Orduan beren nagusiak haietaz <DETK_HAUL>erruki zuen<DETK_HAUL> (il92_9_e_a)

3.3.1.5. DETK postposizio-egituretan

Beste <DETK_POS>kontzertu bezala<DETK_POS> zen esan nahi dut (ae98_8_g_a)
 Nahiz eta Daniel haundiena izan <DETK_POS>ume bezala<DETK_POS> da (ir90_6_q_k)
 atso baten bezalakoa, ispiluan <DETK_POS>ikusitako bezalakoa<DETK_POS> (il92_8_g_d)
 Amak esan dit <DETK_POS>"pinguino" bezala<DETK_POS> nenbiela (il97_10_f_n)

3.3.1.6. DETK izenki-predikatuei

Egun bat <DETK_IP>eguzkitsu zen<DETK_IP> (ae98_2_bp_a)
 Bera <DETK_IP>putakume haundi zen<DETK_IP> (ae98_2_ay_a)
 <DETK_IP>zoriontsu izango<DETK_IP> dela gure euskera ongi hitz egiten eta ulertzen dudanean (ae98_4_ar_a)
 nik geratu nintzen mendiaren taldetan, <DETK_IP>mendizale naizelako<DETK_IP> (ae98_4_ay_a)
 Oso <DETK_IP>mendizale naiz<DETK_IP> asko gustatzeagatik (ae98_4_ak_a)
 Berak niri esaten zidan oso <DETK_IP>ametslari nintzela<DETK_IP> (ae98_4_ak_a)
 bazkaria txarra denez, <DETK_IP>barazkizale zen<DETK_IP> (ae98_4_ab_a)
 <DETK_IP>Egun oso triste zen<DETK_IP>, guk gure bizitza kutxan gordetzeko (ae98_4_ah_a)
 horresagatik gure bizitza oso <DETK_IP>importante da<DETK_IP> (ae98_4_s_a)
 Bere izena Maite zen, <DETK_IP>donostiako zen<DETK_IP> (ae98_4_t_a)
 <DETK_IP>GAU ONA ETA GARBI ZEN<DETK_IP> (ae98_5_ad_a)
 Bera oso isila zen, oso <DETK_IP>misterioso<DETK_IP> (ae98_5_av_a)
 Uste genuen harremana hori <DETK_IP>serio bihurtu<DETK_IP> zela (ae98_6_h_a)
 zezenaren urbil korri egin nuen (oso <DETK_IP>emozionante da<DETK_IP>) (ae98_6_f_a)
 oso <DETK_IP>emozionante izan<DETK_IP> zen (ae98_6_v_a)
 <DETK_IP>pertsona on naizenez<DETK_IP> nire bizitza pozik izango zela (ae98_7_r_a)
 gauza txikiak, baina <DETK_IP>importante eta zorionsuak<DETK_IP> niretzat (ae98_7_g_a)
 bere bizitza nirea baina <DETK_IP>okerrago zen<DETK_IP> (ae98_7_r_a)
 Teorian haurren hazkuntza <DETK_IP>nolako den<DETK_IP> nahiko finkatuta dago (ir90_6_e_i)
 Gauza <DETK_IP>nazioarteko da<DETK_IP> (ir90_7_g_a)
 <DETK_IP>familia berako denean<DETK_IP>, orduan, oso oso desberdina ikusten da (ir90_7_h_a)
 zenbat eta diru gehiago lortu orduan eta <DETK_IP>iraunkorrigo izan<DETK_IP> arazi (ir90_7_h_a)
 piskanaka konturatu gabe du gara edo <DETK_IP>gogorrigo bihurtzen<DETK_IP> gara (ir90_7_h_a)
 aldizkariaren bidez <DETK_IP>inportante izango<DETK_IP> lirakeke (il92_7_a_a)
 gutxi gora bera <DETK_IP>Bitelsen garaiako izango<DETK_IP> da, bere kintako (il92_8_e_b)
 gutxi gora bera Bitelsen garaiako izango da, bere <DETK_IP>kintako<DETK_IP> (il92_8_e_b)
 Gainera, <DETK_IP>puntual da<DETK_IP> eta egunero etortzen da (il92_8_t_b)
 pentsatzeagatik bere arraza <DETK_IP>goi-mailako zenik<DETK_IP> (il92_8_c_e)
 ezagutzen dut eta oso <DETK_IP>lagun gara<DETK_IP> (il92_8_d_e)
 Adolfo <DETK_IP>nekazari zen<DETK_IP> (il92_9_h_b)
 badirudi oso <DETK_IP>kristau dela<DETK_IP> (il92_9_l_d)
 Dokorea, oso <DETK_IP>pentsakor izan<DETK_IP> zenez ez zen konturatu (il92_9_f_i)
 ez dugu faltan botako oso <DETK_IP>gogaikarri dela<DETK_IP> eta (il92_9_l_d)
 Bere bizitza oso <DETK_IP>latz da<DETK_IP> eta ez du lagunik (il92_9_l_d)
 Famulus oso <DETK_IP>despistatu izan<DETK_IP> zenez, gaizki jarrita zeuzkan koordenadak (il92_9_f_i)
 Kontutan hartu ere hurrengo eskutitza <DETK_IP>zure izan<DETK_IP> behar dela (il92_9_c_e)
 Euskalherriko <DETK_IP>pertsona xeble izango<DETK_IP> zara (il92_9_c_e)
 mutil zaharra zen, <DETK_IP>arraro samar zelako<DETK_IP> (il92_9_f_b)
 "Ni ere <DETK_IP>langile naizenez<DETK_IP>, erle erreginari ezitia eman behar diot eta!" (il95_10_r_d)
 nahiz eta jubilatut izan, <DETK_IP>errepresentante denez<DETK_IP>... (il95_10_r_e)
 Sinn Feineren aldeko <DETK_IP>zinegotzi da<DETK_IP> (il95_10_a_e)
 <DETK_IP>esportatzaile diren<DETK_IP> empresentzat (il95_10_i_c)
 arrantzari buruz hitzegiten da, gure herrialdetan ohizko <DETK_IP>lanbide izan<DETK_IP> delako (il95_10_i_h)
 Nahiz eta Irlandako <DETK_IP>arazo izan<DETK_IP>, Euskal Herrian, era bereziz bizi izan ditugu (il95_10_g_f)
 prozesu hura naskagarri zirudien, erleentzat <DETK_IP>natural izan<DETK_IP> arren (il95_10_i_d)
 <DETK_IP>"Fabrika militar" bailitzan<DETK_IP>, erlauntzaren atarian erle bat dago (il95_10_s_d)
 nahiz eta ekonomia gora aldera jo oso <DETK_IP>mantsu da<DETK_IP> bere abiadura (il95_10_a_k)
 ikasten ari naiz baina nere giroa <DETK_IP>ez da euskaldun<DETK_IP> (il95_10_c_i)
 nire giroa <DETK_IP>ez baita euskaldun<DETK_IP> eta totela ematen dut (il95_10_c_k)
 Horrelakoa da gure "txantxangorria", <DETK_IP>euskaldun izango<DETK_IP> balitz bezala (il95_10_j_b)
 Elkarriaketa hastea <DETK_IP>ez zen oso errez<DETK_IP> (il95_10_j_e)
 gure mutilarentzat prozesu hura <DETK_IP>naskagarri zirudien<DETK_IP> (il95_10_i_d)
 erle hori ez da gure <DETK_IP>"Maia-erlea" bezalako izango<DETK_IP> (il95_10_j_d)
 <DETK_IP>Ez da gomendagarri<DETK_IP> bainatzea (il95_12_c_e)
 <DETK_IP>Ez da gomendagarri<DETK_IP> bainatzea (il95_12_e_e)
 aintzinatik dator sendo eta <DETK_IP>gihartsu izateko<DETK_IP> gogoia (il95_12_e_b)
 telebistatik oso <DETK_IP>hunkigarri diren<DETK_IP> ikuskizunak ematen dituztela (il95_12_e_a)
 hilgo gara gure aberria bat, <DETK_IP>haundi eta libre izan<DETK_IP> dadin (il97_10_b_s)
 Gurasoentzako <DETK_IP>gogorrigo da<DETK_IP>, beraiek ere hasi behar ba dute lanean (tr05_12_or13_i)

3.3.1.7. DETK hitza deklinatzean

<DETK_DEKL>Estazioen<DETK_DEKL> Iruñatik beste pertsona aurkitu nuen (ae98_2_as_a)
 Nire lagunak <DETK_DEKL>perretxikok<DETK_DEKL> hartu ditu (ae98_2_bs_a)
 Baina lehen Juratu <DETK_DEKL>Polizik<DETK_DEKL> hill ditut (ae98_2_bx_a)

nik bakarrik nire <DETK_DEKL>flotadorekin<DETK_DEKL> (ae98_4_t_a)
pote bat (kas laranja) hartu genuen, <DETK_DEKL>herriko<DETK_DEKL> (ae98_4_am_a)
Ni imaginatu dut egun bat <DETK_DEKL>eguzkirekin<DETK_DEKL> eta eguraldi ona (ae98_4_i_a)
anaiarekin urrun eta urrunagu <DETK_DEKL>itxason<DETK_DEKL> sartu ginen (ae98_4_t_a)
Nire flotadorekin <DETK_DEKL>itxason<DETK_DEKL> sartu nuen (ae98_4_t_a)
Nire <DETK_DEKL>flotadorekin<DETK_DEKL> itxason sartu nuen (ae98_4_t_a)
nire lehenengo <DETK_DEKL>kotxerekin<DETK_DEKL> (ae98_5_b_a)
Nik azkar esan nion baietz <DETK_DEKL>ospitaleri<DETK_DEKL> (ae98_5_ai_a)
kantatu peñarekin <DETK_DEKL>kaletan<DETK_DEKL> zehar (ae98_5_ac_a)
Donostian egun berezi hau egon nintzen nire <DETK_DEKL>neskatxorekin<DETK_DEKL>. (ae98_6_ñ_a)
eta nire herriko <DETK_DEKL>kaletatik<DETK_DEKL> izerditaz balin nenbilen (ae98_7_ab_a)
Beste gauzarik ezin genuen egin, <DETK_DEKL>eguraldigatik<DETK_DEKL> (ae98_7_ñ_a)
Elkarrizketak pikardia, fantasia eta <DETK_DEKL>humorekin<DETK_DEKL> idatzita daude (ir90_6_i_g)
Batzuk <DETK_DEKL>bestek<DETK_DEKL> baino azkarrago ikasten dute (ir90_6_e_i)
<DETK_DEKL>Konserjek<DETK_DEKL> berogailua pixturakoan, (ir90_6_a_a)
eta Fármaco <DETK_DEKL>Kimikokoin<DETK_DEKL> errezagoa (ir90_6_d_f)
<DETK_DEKL>Gakoren<DETK_DEKL> bitartez atea ireki zuen (ir90_6_m_a)
Behar bada, alargun <DETK_DEKL>tristeren<DETK_DEKL> irudia<DETK> (ir90_7_a_a)
adibidez irakasleak <DETK_DEKL>galderen<DETK_DEKL> bat egiten dionean (il92_8_a_b)
zaldi bat eukitzea ez zela interaldi bat bere <DETK_DEKL>arazori<DETK_DEKL> (il92_8_o_a)
alde batetik bestera beren herrietatik beste <DETK_DEKL>herritara<DETK_DEKL> (il92_8_g_e)
berak ez zeukan <DETK_DEKL>ohiturik<DETK_DEKL> etxean egoteko (il92_9_l_c)
<DETK_DEKL>gizakirengandik<DETK_DEKL> ikasitako balorea da (il92_9_b_c)
horregatik <DETK_DEKL>emakumerentzat<DETK_DEKL> os zaila lan munduan sartzea (il92_9_l_h)
Donostira etorri ginenean nere amak <DETK_DEKL>ausokideri<DETK_DEKL> oparitu zion (il92_9_f_c)
gaur arte <DETK_DEKL>kotxeren<DETK_DEKL> historioak asko aldatu du (il95_9_i_f)
Gure gizarteak <DETK_DEKL>kotxeren<DETK_DEKL> cultura oso sartuta daukagu (il95_9_i_f)
egunero goizeko <DETK_DEKL>bederatzitan<DETK_DEKL> hondartzara joaten nintzen (il95_9_t_b)
jeiki nintzen eta <DETK_DEKL>lehiotatik<DETK_DEKL> baieztatu nuen egia zelako (il95_9_t_b)
Ekologistek <DETK_DEKL>itsasorentzat<DETK_DEKL> txarra dela esaten omen dute (il95_10_r_h)
<DETK_DEKL>erleri<DETK_DEKL> eta erlauntzari buruzko kontu kontari batzuekin (il95_10_s_d)
telebista ikusten dut, orain ez dago txuri <DETK_DEKL>beltzen<DETK_DEKL> (il95_10_r_f)
eztia ez dela <DETK_DEKL>erleren<DETK_DEKL> kaka (il95_10_s_d)
galdetzen dio ea ezia <DETK_DEKL>erleren<DETK_DEKL> kaka den ala ez (il95_10_s_d)
ikusiko dugu nola ospatzen den urte hau beste <DETK_DEKL>hiritan<DETK_DEKL> (tr05_12_or9_f)

3.3.2. DETG (DETerminatzaile atzizkia Gehitu)

Erregelak: DETG1A, DETG1B, DETG1C, DETG1D

3.3.2.1. Determinatzaile atzizkia erabiltzea behar ez denean

Herri bat jaietan, <DETG>mozkorra haundiarekin<DETG> (ae98_3_i_a)
Autobusa igotzerakoan oso <DETG>lasaia<DETG> sartzen nintzen (ae98_4_ak_a)
hor egon ginen, etxanda eta <DETG>lasaia<DETG> (ae98_6_n_a)
hor zeuden siurtasun neurriak <DETG>ikaragarria<DETG> atsegin zitzaigun (ir90_6_n_a)
saila egiten zaizkio <DETG>urte gutxikoa mutil bat<DETG> delako (il92_9_h_d)
<DETG>alaia<DETG> ta korrika sartzen da nere etxean zaunka eginez (il92_9_g_c)

3.3.2.2. DETG aditz-esapideetan

batzutan egun horrekin <DETG_HAUL>ametsa egiten<DETG_HAUL> dut (ae98_6_c_a)
Nik <DETG_HAUL>argia ez daukadana<DETG_HAUL> zera da (il92_9_n_h)
bihar eta etzi <DETG_HAUL>jaia daukat<DETG_HAUL> eta (il95_9_i_c)
behatz haundian <DETG_HAUL>mina egiten<DETG_HAUL> zidan (il97_10_l_n)

3.3.2.3. DETG postposizio-egituretan

Gero gu <DETG_POS>hiria erdian<DETG_POS> bazkaldu gara (ae98_2_aw_a)
<DETG_POS>uhina malapartuaren bidez<DETG_POS> dana aldatu da (ir90_7_a_a)
<DETG_POS>mahaia inguruan<DETG_POS> ikasle pila bat biltzen ohi zen (il95_10_a_i)
Europako Diru Sisteman pezeta jarraitzeko <DETG_POS>neurria bezala<DETG_POS> (il95_10_s_c)

3.3.2.4. DETG izenki-predikatuei

nekatuta negoen baina oso <DETG_IP>alaia negoen<DETG_IP> (ae98_2_a_a)
galdetu diot nola <DETG_IP>posiblea den<DETG_IP>? (ae98_2_am_a)
Orain dela hogeit urte <DETG_IP>gaisoa zegoen<DETG_IP> lagun bat, kanzer edo hori (ae98_3_m_a)
Denok nekatuta baina <DETG_IP>alaia geunden<DETG_IP> (ae98_4_h_a)

Batzutan menditik igo hortik aurrera oso <DETG_IP>lasaia gelditzen<DETG_IP> nintzen (ae98_4_ak_a)
 oso <DETG_IP>indartsua nago<DETG_IP> (ae98_4_k_a)
 Nik eta nire ematea oso <DETG_IP>alaiak geunden<DETG_IP> (ae98_4_ai_a)
 ni aukeratuta nengoela esan zidatenean oso <DETG_IP>alala nengoen<DETG_IP> (ae98_4_j_a)
 nire ama eta nire izeba porru bat erre zuen eta oso <DETG_IP>alaiak jarri<DETG_IP> zuten (ae98_5_ak_a)
 Egunak pasa ondoren oso <DETG_IP>tristeia nintzen<DETG_IP> (ae98_5_bb_a)
 baina oso <DETG_IP>beldurra nintzen<DETG_IP> ezer ez egin nuen (ae98_6_r_a)
 Etxe barruan <DETG_IP>iluna zegoen<DETG_IP> (ae98_6_g_a)
 bezpera iritsi zen, oso <DETG_IP>urduria zegoen<DETG_IP> (ae98_7_ñ_a_bis)
 Meza bukatu ondoren jende guztia <DETG_IP>alala zegoen<DETG_IP> (ae98_7_ñ_a_bis)
 Une iritsi zenean baietz esan oso <DETG_IP>lasaia geratu<DETG_IP> zen (ae98_7_ñ_a_bis)
 berak oso <DETG_IP>bakartia sentitu<DETG_IP> zela azaldu zidan (ae98_7_c_a)
 Handik aurrera etxaldearen <DETG_IP>jabea<DETG_IP> abereak izan ziren (ir90_6_h_h)
 beste herrietara joateko bideak <DETG_IP>libreak zeuden<DETG_IP> (il92_8_t_a)
 "Ez <DETG_IP>da posiblea<DETG_IP>!" - esaten dio bere buruari (il92_8_e_f)
 Dirudenez guztiek <DETG_IP>mutuak geratu<DETG_IP> ziren (il92_9_e_a)
 Oso <DETG_IP>hurbila zeudenean<DETG_IP> apuntatu nituen (il97_10_f_o)
 Mikelek ez zeuzkan arrazoirik <DETG_IP>goibela egoteko<DETG_IP> (il97_10_d_o)
 irakurri dut: Pezeta, <DETG_IP>gaixoa omen dagoela<DETG_IP> (il95_10_r_c)

3.3.2.5. DETG hitza deklinatzean

<DETG_DEKL>herriatik<DETG_DEKL> kaminra joan ginen (ae98_2_bw_a)
 gu <DETG_DEKL>zelaatik<DETG_DEKL> igo ginen (ae98_2_bw_a)
 "uniformea" <DETG_DEKL>zoborrontziara<DETG_DEKL> bota nuen (ae98_4_b_a)
 zelaia <DETG_DEKL>jendeaz<DETG_DEKL> beteta zegoen (ae98_4_ag_a)
 Egun osoa <DETG_DEKL>jaiatik<DETG_DEKL> jaiara (ae98_4_s_a)
 Egun osoa jaiatik <DETG_DEKL>jaiara<DETG_DEKL> (ae98_4_s_a)
 piska bat ibili genuen (10 minutu) <DETG_DEKL>harriaraino<DETG_DEKL> (ae98_6_m_a)
 Eskuminak Vincennesko parkeko <DETG_DEKL>basauntzaei<DETG_DEKL> (ir90_6_b_c)
 Nire <DETG_DEKL>institutoako<DETG_DEKL> ikusentzungailuen arduraduna naiz (ir90_6_o_a)
 ez da beste pertsona arraut bat bezala: <DETG_DEKL>jatorriaz<DETG_DEKL> eta edertasunaz (ir90_7_d_a)
 Neguak hotza eta gosea ekarri zituen <DETG_DEKL>herriara<DETG_DEKL> (il92_8_r_a)
 karga herrestan eramaten alaitu zen <DETG_DEKL>herriara<DETG_DEKL> Adolfo (il92_8_r_a)
 lanik gabe, janarik gabe, <DETG_DEKL>goseaz<DETG_DEKL> alegia (il92_9_l_b)
 <DETG_DEKL>Augustoak<DETG_DEKL> ezin zuen bere farra berreduki, (il95_10_r_d)
 "Eman ezia!" zion <DETG_DEKL>Augustoak<DETG_DEKL> (il95_10_r_d)
 3'3 miloi langabe daude, hori da % 12'3 biztanle <DETG_DEKL>aktiboatik<DETG_DEKL> (il95_10_s_g)
 lurrikara gertatu baino lehenago jendea <DETG_DEKL>hiriatik<DETG_DEKL> ateratzea (il95_10_j_f)
 baina <DETG_DEKL>Flaco-aren<DETG_DEKL> arabera, musika hau Europan hasi zen jotzen (il95_10_r_g)
 F. Mitterrand, ez da aurkeztu bi <DETG_DEKL>motiboagatik<DETG_DEKL> (il95_10_s_g)
 Etxe pobre <DETG_DEKL>askoetatik<DETG_DEKL>, jateko eza eta (il95_12_m_c)
 Ni <DETG_DEKL>txikietatik<DETG_DEKL> joan naiz futbolera (il97_10_b_l)

3.3.3. DETE (DETerminatzailea Errepikatu)

3.3.3.1. IZE + DETatzizkia + ADJ + DETatzizkia

Erregelak: DETE1A, DETE1B, DETE1C, DETE1D1, DETE1D2

<DETE>Eguna zoriontsuena<DETE> zen, nire semea jaio zenean (ae98_2_cd_a)
 <DETE>Eguna zoriontsuena<DETE> zen, nire semea jaio zen eguna (ae98_2_ce_a)
 <DETE>Eguna zoriontsuena<DETE> zen, nire semea jaio zenean (ae98_2_cf_a)
 <DETE>Gaua osoa<DETE> pasatu ginen jolastu ari ginen (ae98_2_bz_a)
 <DETE>Eguna oberena<DETE> parranda egiteko delako (ae98_2_ac_a)
 eta nik <DETE>moskorra haundia<DETE> hartu nuen (ae98_2_bh_a)
 Ze <DETE>plana polita<DETE>, bai eh? (ae98_2_bx_a)
 Nire <DETE>eguna oberena<DETE> da noiz nire alaba jaio zen (ae98_2_ba_a)
 Oraindela 4 urte nire <DETE>laguna hoberena<DETE> kartzelatik atera zen (ae98_2_an_a)
 Nire <DETE>eguna zoriontsuena<DETE> hori da (ae98_3_y_a)
 <DETE>mozkorra haundia<DETE> harrapatu genuen (ae98_4_al_a)
 <DETE>mina haundia<DETE> sentitzen nuen orkatilan (ae98_4_v_a)
 denak itsasontzi batean <DETE>gaua osoa<DETE> igon dugu (ae98_4_i_a)
 kantatu peñarekin kaletan zehar, eta <DETE>mozkorra handia<DETE> arrapatu genuen (ae98_5_ac_a)
 oso ongi pasatu genuen, gero <DETE>gaua osoa<DETE> dantza eta dantza (ae98_5_aw_a)
 Orain <DETE>Euskal Herria osoa<DETE> ezagutzen dut eta neguan barnetegi batean egon naiz (ae98_5_ao_a)
 bere begiak, bere ezpainak goxo-goxoak, bere <DETE>lepoa luzea<DETE>, bere gorputza (ae98_7_q_a)
 egutegiaren aldaketa dela medio bi klaseen artean <DETE>astia librea<DETE> genuen (ae98_8_b_a)
 bere etxea, bere <DETE>ingurua hurbilena<DETE> ezagutzen ditu (ir90_6_c_d)

<DETE>Irtenbidea bakar<DETE> neukan (il92_8_g_d)
txakur batek gastu eta <DETE> lana handia<DETE> ematen digula (il92_9_d_c)
bildurtua eta <DETE> ekaitza haundia<DETE> jasotzen (il92_9_k_c)
balore hau <DETE> zerikusia haundia<DETE> dauka txakurrari nik eransten diodanarekin (il92_9_b_c)
Badago, aintzinatik, txakurrarekiko gizakiak duen lotura eta <DETE> harremana sakona<DETE> (il92_9_b_c)
<DETE> zinea mutua<DETE> ikusten dugunean inoiz ez zaigu bururatu komentarioak falta direla (il92_9_n_g)
Oso <DETE> txakurra atsegina<DETE> da (il92_9_g_c)
gustatuko litzaidake hain <DETE> txakurra alaia<DETE> ez dadila izan (il92_9_g_c)
Dena den, ni <DETE> gidaria ona<DETE> naiz (il95_9_i_e)
teknologiaren aurrerapenak <DETE> indarra handiagoa<DETE> emango diete komunikabideei (il95_10_m_k)
hiru puntu definitzen du: <DETE> unibertsitatea nazionala, herritarra eta euskalduna<DETE> (il95_10_c_g)
zuk pentsa ez dezazun ni <DETE> gidaria edo langilea txarra<DETE> naizenik (il95_10_s_j)
aintzinako arrantza <DETE> teknikak berriarekin<DETE> aurrez-aurre daudelako (il95_10_r_h)
Badakizu ni <DETE> marinela trebea<DETE> naizela (il97_10_f_o)

nire lagunekin <DETE> moskorrak haundiak<DETE> hartzen ditut (ae98_2_m_a)
Nik beste jende eta <DETE> gauzak berriak<DETE> ezagutu nahi nituen (ae98_4_j_a)
Nik <DETE> "aletak" berriak<DETE> nituen (ae98_4_t_a)
oso dotore zeunden, bere <DETE> arropak politenak<DETE> jantzitaz (ae98_5_ai_a)
<DETE> Ikurriñak gorriak<DETE> jarri zuten (ae98_6_i_a)
buelta bat eman genuen <DETE> gauzak ikusgarriak<DETE> ikusteko (ae98_7_x_a)
<DETE> Personaia nagusiak<DETE> Martin du izena (ir90_6_d_g)
hilabetero aterako da eta barruan <DETE> artikuloak interesgarriak<DETE> izango dira (il92_7_j_a)
Udaletxeko kultur komisio <DETE> lehendakaria jauna<DETE> -Donostia (il92_7_g_a)
<DETE> Jaun Andreak agurgarriak<DETE> (il92_7_d_a)
badira herri bakoitzeko <DETE> esakerak propioak<DETE> (il92_9_b_g)
Denborarekin txisteez aldatzen dute, <DETE> egoerak desberdinak<DETE> pasatzeagatik, (il92_9_f_f)
Militarrek <DETE> gauzak arraroak<DETE> egiten dituzte (il97_10_e_s)
<DETE> prakak urdinak<DETE>, Rusiako gerrako marinela eramaten zuen alkondara (il97_10_f_o)

Ere nire <DETE> lagunak hoberena<DETE> etorri zen (ae98_2_g_a)
Eskalatu nuenean <DETE> zentsazioak polita<DETE> nuen (ae98_6_m_a)

3.3.3.2. IZE + DETatzizkia + Artikulu partitiboa

"ez duzu <DETE_DEKL> soinuarik<DETE_DEKL> egin!" (ae98_4_aa_a)
Eguzki oso handia zen, <DETE_DEKL> hodeiarik<DETE_DEKL> gabe (ae98_5_o_a)
nire bizitzeko <DETE_DEKL> ilusioarik<DETE_DEKL> handiena izan zen (ae98_6_r_a)
munduko <DETE_DEKL> gizonarik<DETE_DEKL> alaitsuena izanen nintzateke (ae98_7_n_a)
lanik gabe, dirurik gabe, <DETE_DEKL> etorkizunarik<DETE_DEKL> gabe (il92_8_t_a)
Jaio zenean amonak <DETE_DEKL> sudurrarik<DETE_DEKL> ez zeukala esan zuen (il92_9_k_d)
urte batzuk <DETE_DEKL> txakurrarik<DETE_DEKL> gabe pasa genuen (il92_9_h_c)
<DETE_DEKL> barazkiarik<DETE_DEKL> ez egotean, inork ez zeukan zer o zer saltzeko (il92_9_g_b)
Peio Bixente agertu zen han. Herriko <DETE_DEKL> marinelarik<DETE_DEKL> haundiena (il97_10_f_o)

3.3.3.3. IZE + DETatzizkia + Artikulu zehaztugabea

Erregelak: DETE2A, DETE2B, DETE2C, DETE2D1, DETE2D2, DETE2D3

Euskal Herria <DETE> nazioa bat<DETE> izan dela (ae98_2_av_a)
Lekeitio <DETE> herria bat<DETE> erosten dut (ae98_2_bl_a)
<DETE> Cochea bat<DETE> erosten dut (ae98_2_bl_a)
lehenago nik <DETE> papera bat<DETE> sartu nuen (ae98_2_bt_a)
guk <DETE> afaria bat<DETE> egin ginenuen. (ae98_2_f_a)
denak <DETE> lana bat<DETE> behar dugu (ae98_2_av_a)
Baldin badut <DETE> lana bat<DETE> (ae98_2_n_a)
nik etxera joan nintzen eta <DETE> pseo bat<DETE> eman nuen (ae98_2_n_a)
Polizia bat etorri zen eta hasi zuen bronka ematen (<DETE> Forala bat<DETE> zen) (ae98_3_i_a)
Oso ongi gogoratzen dut egun hori, <DETE> larunbata bat<DETE> zen (ae98_6_b_a)
<DETE> motorra bat<DETE> pasatu zuen eta gure pedal hasi zen erabat mugitzen (ae98_6_l_a)
Niri <DETE> ametsa bat<DETE> irudi zitzaidan (ae98_6_z_a)
<DETE> deseoa bat<DETE> eman diezazudan etorri naiz (ae98_7_b_a)
Ipameriketako polizia batzuek <DETE> beltza bat<DETE> gaizki tratatu zuten (il92_8_c_e)
Joan den astean <DETE> erizaina bat<DETE> kirolfanotik egotzi (expulsada) izan zen (il92_8_x_e)
bere gorputza <DETE> makila bat<DETE> dela esan dezaket (il92_9_l_d)
gero eta jende gehiagok <DETE> automobila bat<DETE> edukitzea lortu zuela (il95_9_t_f)
Bazen behin, <DETE> Tolosarra bat<DETE> inauterietan erlez moztortzen zena (il95_10_i_d)
bidaiariak niren kontra jarri ziren, <DETE> zaharra baten aurka<DETE> joateagatik (il95_10_r_j)

<DETE>Barea bat<DETE> zen. N~o! Ze nazka! (il97_10_f_n)
hilgo gara gure <DETE>aberria bat<DETE>, haundi eta libre izan dadin (il97_10_b_s)

Eguraldi ona izan zen eta <DETE>ekaitza batzuk<DETE> ezik, gehiagorik ez zegoen (ae98_6_i_a)
jabea kanpora bota zuten eta berehala <DETE>araua batzuk<DETE> idatzi zituzten (ir90_6_l_h)

Eta <DETE>egunak batzuk<DETE> eta gero bazirudien zaldia ohitu egin zela (il92_9_c_b)

bukatu genuen 8etan hotelean, afaltzen, kriston <DETE>mozkorra batekin<DETE>, erdi lurrean (ae98_2_z_a)
guk <DETE>erizaina bateri<DETE> eskatu genion tabakoak erosi zizkigun (ae98_3_d_a)
eta <DETE>sagardotegia batean<DETE>, kriston giro dugu (ae98_4_i_a)
NOSKI <DETE>OHIANA BATEAN<DETE> AGERTU GINEN (ae98_5_ad_a)
<DETE>txartela batekin<DETE> zinzilikatuta nire alkondaratik (ae98_5_ae_a)
jantzi bereziekin: kasko, botak, eskularruak, txamarra eta <DETE>extintorea batekin<DETE> (ae98_5_c_a)
Atzo <DETE>xaguxarra batek<DETE> berri txar bat eman zidan (il92_8_e_a)
Nire aita beste <DETE>lana batekin<DETE> hari zen (il92_9_l_c)
<DETE>Atezaina batekin<DETE> topo egin zuen (il95_10_c_d)
hau zabaltzen badugu gure umeen <DETE>laguna bakarra<DETE>, telebista izango da (il95_10_g_k)

3.3.3.4. IZE + DETatzizkia + ADJ + Artikulu zehaztugabea

Erregelak: DETE3A1, DETE3A2, DETE3D1, DETE3D2

nire lagunei gonbidatu nuen eta <DETE>mozkorra handi bat<DETE> harrapatu genuen (ae98_4_k_a)
Gauza horiek ospatzeko <DETE>jaia handi bat<DETE> egin nuen (ae98_4_k_a)
Egoera honi <DETE>irtenbidea on bat<DETE> aurkitzea oso zaila dirudit (il95_10_a_k)

<DETE>Amantea gazte batetaz<DETE> maitemindu zen (il92_8_o_e)

3.3.3.5. IZE + ADJ + DETatzizkia + Artikulu zehaztugabea

Erregelak: DETE3B1, DETE3B2, DETE3E1, DETE3E2

Hau da <DETE>egun normala bat<DETE> baina oso ona (ae98_2_n_a)
Nire familia <DETE>liburu polita bat<DETE> eman didate "Guatemalako Mayak" (ae98_2_ax_a)
nik nire lagunari <DETE>ostika txikia bat<DETE> eman diot (ae98_2_ay_a)
Nirentzat hori zen <DETE>egun bat zoriontsuena<DETE> (ae98_2_ah_a)
<DETE>Soinetako txuria bat<DETE> jantzi nuen (ae98_4_a_a)
nire lagun bat deituko nion, <DETE>berri ona bat<DETE> kontatzeko (ae98_5_k_a)
Horregatik, <DETE>egun polita bat<DETE> kontatuko dut (ae98_5_r_a)
Linternarekin <DETE>katu txikiak batzuk<DETE> ikusi genituen (ae98_6_g_a)
ez da ikusi behar gaztelanian edo frantsesan <DETE>etsai potentziala bat<DETE> (il95_10_m_e)
kontutan hartu <DETE>hazkunde arrunta bat<DETE> % 3 batz bestea izaten dela (il95_10_r_k)

<DETE>mundu ezberdina batean<DETE> bizi nintzela amestu nuen (ae98_5_l_a)
Informatiboak <DETE>irudi gogorra baten<DETE> bila dabiltza (il92_9_j_g)
beti <DETE>multzo handia batean<DETE> bizi ohi dira (il95_10_j_d)
eta <DETE>lore handia batean<DETE> gelditzen da (il95_10_j_d)
zergatik ez diren eskontzen <DETE>eliza arrunta batean<DETE> (il95_10_r_c)

3.3.3.6. IZE + DETatzizkia + ADJ + DETatzizkia + Artikulu zehaztugabea

Erregelak: DETE3C1, DETE3C2, DETE3C3, DETE3F1, DETE3F2

<DETE>gauzak politikak batzuk<DETE> esan nizkion (ae98_6_o_a)

3.3.3.7. IZE + DETatzizkia + Erakuslea

Erregelak: DETE4A, DETE4B

<DETE>Eguna hori<DETE> aldatu nire visitas (ae98_2_s_a)
<DETE>Eguna hori<DETE> oso zoriontsuena zen nere bizitzaentzako (ae98_4_ac_a)
azkenean <DETE>idazlana hau<DETE> idaztea gustatu zait (ae98_5_bb_a)
<DETE>Idazlana hau<DETE> kostatu zait hoge eta hamar minutu (ae98_5_bb_a)
Orain nik uzte dut <DETE>dirua hori<DETE> Roldan eta Urralburukoa da eta (ae98_5_b_a)
Uste genuen <DETE>harremana hori<DETE> serio bihurtu zela (ae98_6_h_a)
momentu batean <DETE>poza hori<DETE> zanpatua izan zen (ae98_6_i_a)
Gidaria piska bat lotan zegoen (<DETE>gaua hori<DETE> jaranda egon zen) (ae98_6_m_a)
minik ez egiteko, <DETE>mina hori<DETE> bere inguruan egunero ikus dezakenean? (il92_8_x_e)
<DETE>Udala horrek<DETE> etxebizitzak berrizatzeko eta konpontzeko eskaintzen dituen... (il95_10_m_j)
Australiara heldu ondoren, <DETE>eraztuna hau<DETE> oparitu zidaten (il97_10_f_o)

<DETE>Neskek hauek<DETE> kampainako dendetan lo egin behar zuten (ae98_6_h_a)
 Iruñara itzuli ginen eta <DETE>gaua horretan<DETE> gende guztioei ezagutu nuela esan zidan (ae98_6_m_a)
 <DETE>Neskek horrekin<DETE> mendiari buruz hitz egitea hasi genuen (ae98_6_h_a)
 Orain dela hiru urte asi nintzen <DETE>lana hauekin<DETE> (ae98_7_u_a)
 klasekoak hasi baino lehen, zuzentzen naiz <DETE>saila hortara<DETE> (ir90_6_a_a)
 <DETE>Haurra hauek<DETE> eskolara aliegatu direnean (ir90_6_b_d)
 <DETE>Kontraesana honek<DETE> desoreka batean dauka jatorria. (il97_10_b_r)
 <DETE>Gaua hartan<DETE> ezin nuen lorik egin hainbeste mina oinetan neukan eta (il97_10_f_n)
 <DETE>Iurra honetan<DETE> edozein tokitan bezala euria egiten zuen (il97_10_f_o)
 <DETE>Irtenbidea hauek<DETE> hartzen zituen jendea gaizki ikusita zegoen (tr05_12_or11_h)

3.3.3.8. IZE + DETatzizkia + ADJ + Erakuslea

Erregelak: DETE4C1, DETE4C2, DETE4C3, DETE4C4, DETE4F

tokatu zaigun <DETE>Euskal Herria maite honetan<DETE>, arazo dexente dugunez (ae98_7_n_a)

3.3.3.9. IZE + ADJ + DETatzizkia + Erakuslea

Erregelak: DETE4D1, DETE4D2, DETE4D3, DETE4D4, DETE4G

<DETE>irudi gogorra hori<DETE> guri bota diezaguketela (il92_9_j_g)

3.3.3.10. IZE + DETatzizkia + ADJ + DETatzizkia + Erakuslea

Erregelak: DETE4E1, DETE4E2, DETE4E3, DETE4E4, DETE4H

3.3.3.11. IZE + DETatzizkia + Zenbatzaile Orokorra + DETatzizkia

Erregelak: DETE6A1, DETE6A2

Ni egun batean jaiki naiz eta <DETE>gerrak guztiak<DETE> bukatuta dira (ae98_2_ar_a)
 zuri <DETE>idazlana guztia<DETE> korregitu pxka bat gehiago kostatuko zaizu (ae98_5_bb_a)
 Hori da <DETE>mundua osoak<DETE> sentitu duena (ir90_7_g_a)
 Garai hartan <DETE>gauzak guztiak<DETE> bazekiten hitz egitea (il92_9_d_a)
 beste <DETE>gauzak guztiak<DETE> (gizakia izan ezik) erabaki zuten (il92_9_d_a)
 <DETE>Lurra osoa<DETE>, gu bizi garen lurrean esan nahi dut (il95_10_j_f)
 sagarraren azalak nazka ematen didanez, utzi nuen <DETE>azala guztia<DETE> mahaian (il97_10_f_m)
 azala kendu, eta <DETE>azala guztiak<DETE> niretzat (il97_10_f_m)
 Papera eta pipa <DETE>azala guztiak<DETE> erori ziren irakasleen gainera (il97_10_f_m)

3.3.3.12. IZE + DETatzizkia + Zenbatzaile Zehaztugabea

Erregelak: DETE7A, DETE7B

berak <DETE>argazkia asko<DETE> egin zuen. (ae98_2_g_a)
 mendira joan ginen eta <DETE>animaliak asko<DETE> ikusi genuen (ae98_2_ag_a)
 Tzuria eta <DETE>puntillak asko<DETE> (...) (ae98_2_b_a)
 Hori eguna lagunak, bere gurasoak eta <DETE>auzokidek asko<DETE> kartzelara joan ginen (ae98_2_an_a)
 Ondoren osabak-izebak eta <DETE>lengusinak gehiago<DETE> etorri ziren nire etxera (ae98_2_g_a)
 Orain dela <DETE>urtea asko<DETE> Irlandara joatea nahi dut (ae98_4_r_a)
 ez zuten <DETE>seme-alabak gehiagorik<DETE> (ae98_4_g_a)
 Bidaia oso txarra zen, gure kotxe oso zaharra zelako eta <DETE>mendiak asko<DETE> zegoen (ae98_4_m_a)
 Nahiz eta berandu altxatu naiz <DETE>gausak asko<DETE> egin dezaket (ae98_4_ñ_a)
 Espero dut <DETE>egun zoriontsua gehiago<DETE> edukitzea (ae98_5_m_a)
 Baina <DETE>gauzak guztien<DETE> gainera nahi nuela egin beharko nuen (ae98_5_g_a)
 Eskolan auzo txikian zegoen eta <DETE>zelaia asko<DETE> zegoen (ae98_6_g_a)
 <DETE>erreztasuna gehiago<DETE> edukiko dute hizkuntzarekin (ir90_6_b_d)
 Bigarrena-nahiz eta handia izan- <DETE>atseratua xamarra<DETE> da (ir90_6_p_k)
 eta bi egun honetan <DETE>frogak asko<DETE> egin dizkizugu (il92_8_g_d)
 Langabezia aurkitzen direnak gero eta <DETE>pertsonak gehiago<DETE> dira (il92_8_c_e)
 gero eta <DETE>belarra gutxiago<DETE> ematen dio Adolfok (il92_8_g_c)
 oheratu nintzen, <DETE>lana askorekin<DETE> egun batean pasa eta gero (il95_10_r_e)
 oso <DETE>ura gutxi<DETE> dagoela urtegietan entzunda bait daukat (il95_12_e_e)

3.3.3.13. Zenbatzaile zehaztugabea + IZE + DETatzizkia

Erregelak: DETE9A, DETE9B

ilusioa, inozentzia eta <DETE>hainbeste gauzak<DETE> (ae98_6_x_a)
 ez dakit <DETE>zenbat gauzak<DETE> egin genuen (ae98_6_y_a)
 dagoeneko <DETE>zenbait gauza interesgarria<DETE> egin du (il92_7_i_a)
 <DETE>Zenbat gauzak<DETE> egin dezakezu gauek? (il92_8_j_a)
 ez dira <DETE>hainbeste gauzak<DETE> egiten (il92_9_l_h)

Ez bagenu jasoko <DETE>hainbeste berriak<DETE> entzuten edo ikusten ditugunak... (il95_10_s_k)
 <DETE>Zenbat gonbidatuak<DETE> ospakizunetara joango dira? (il95_10_r_c)
 <DETE>hainbeste dirua<DETE> mugitzen da, non (il95_12_e_b)
 Gaua hartan ezin nuen lorik egin <DETE>hainbeste mina<DETE> oinetan neukan eta (il97_10_f_n)
 nire buruari galdetu nion zergaitik <DETE>hainbeste algarak<DETE> (il97_10_f_n)
 <DETE>hainbeste trebesia eta erraztasuna<DETE> musikarako ikusiz (il97_10_b_u)
 etxean badaude <DETE>zenbait robotak<DETE> zenbait gauzak egiten dutenak (tr05_12_or9_f)
 etxean badaude zenbait robotak <DETE>zenbait gauzak<DETE> egiten dutenak (tr05_12_or9_f)
 duela zenbait urte amestutako <DETE>zenbait gauzak<DETE> egia bihurtu dira (tr05_12_or9_f)
 eta beste <DETE>zenbait gauzak<DETE>, aitzitik ez (tr05_12_or9_f)
 bizitzan dauden <DETE>hainbat ezusteak<DETE> badaukatela zerikusirik (tr05_12_or14_j)
 ze desberdinak diren lurrean zehar dauden <DETE>hainbat kulturak<DETE> eta ohiturak (tr05_12_or8_f)

3.3.3.14. Zenbatzaile zehaztugabea + IZE + DETatzizkia + ADJ

3.3.3.15. Zenbatzaile zehaztugabea + IZE + ADJ + DETatzizkia

Erregela: DETE9C

3.3.3.16. Zenbatzaile zehaztugabea + IZE + DETatzizkia + ADJ + DETatzizkia

Erregela: DETE9D

gu harritutak geunden <DETE>hainbeste gauzak ezberdinak<DETE> ikusteagatik (ae98_4_af_a)
 <DETE>zeinbat teknika berriak<DETE> debekatu egin zituztelako (il95_10_m_h)

3.3.3.17. Determinatzaile nolakotzaile/galdetzaile/zehaztugabea + IZE + DETatzizkia

Erregela: DETE10A

Vignemal, <DETE>edozein mendia<DETE> baino politagoa da niretzat (ae98_2_ca_a)
 Gero gerta daiteke <DETE>edozein gauzak<DETE> (ae98_5_ab_a)
 eta ikusten da <DETE>zer giroa<DETE> bizitzen zen-eta bizitzen da (ir90_6_p_k)
 han landatzen zuen <DETE>edozer barazkirik<DETE>, gero hirian saltzeko (il92_8_c_c)
 bizitza zaintzen duena, <DETE>edozein gaisotasuna<DETE> hartzeaz libre dagoela (il92_8_c_d)
 <DETE>Edozein laneguna<DETE> izango balitz bezala... (il92_8_g_d)
 kazetariak hobe jakinerazten digute <DETE>edozein berria<DETE> (il92_9_l_g)
 Baina <DETE>zenbat dirua<DETE> mugitzen da futbolaren bidez? (il97_10_b_l)
 <DETE>Edozein ekintzak<DETE> ez duela onartuko dio (il95_10_a_f)
 eta gainera <DETE>edozein atsoak<DETE> zuek haina dakiela erantzuten zuten (il95_10_a_i)

3.3.3.18. Determinatzaile nolakotzaile/galdetzaile/zehaztugabea + IZE + DETatzizkia + ADJ + DETatzizkia

Erregela: DETE10B

3.3.3.19. Zenbakia + IZE + DETatzizkia

Erregela: DETE11

<DETE>hiru txartela<DETE> emaidazu, mesedez (ae98_2_bh_a)
 autobus geltokira joan ginen eta <DETE>zazpi txartela<DETE> erosi genuen (ae98_4_ay_a)
 <DETE>Bost laguna<DETE> ginen eta kotxean pixka bat "apretatuta" geunden (ae98_4_q_a)
 <DETE>5 gola<DETE> sartu nuen eta irabazi genuen (ae98_4_u_a)
 <DETE>Bi posibilitatea<DETE> dauzkate (ir90_6_k_d)

egunak aprobetatzatzen ditut gure bizia <DETE>lau egunak<DETE> direlako (ae98_4_ap_a)
 orain dela 18 urte, 2 hilabete, 3 egun, <DETE>10 orduak<DETE> eta 20 minutu (ae98_4_u_a)
 Españan <DETE>35 milioi entrenatzaileak<DETE> daude (il97_10_b_l)

3.3.3.20. Zenbakia + IZE + DETatzizkia + ADJ

Erregela: DETE12A

3.3.3.21. Zenbakia + IZE + ADJ + DETatzizkia

Erregela: DETE12B

aurkitu nituen <DETE>bi txakur polita<DETE> (ae98_5_bb_a)
 lehenengo tabernan <DETE>bi gausa interesgarriak<DETE> ikusi genituen (ae98_5_ax_a)
 <DETE>4 herri desberdina<DETE> pasatu genuen (ae98_6_v_a)
 Aldizkari hau <DETE>hiru zati desberdinak<DETE> edukiko ditu (il92_7_g_a)
 beraien artean <DETE>bi ezberdintasun garrantzitsuak<DETE> aipatuko ditut (il92_8_y_e)

<DETE>Hiru joera desberdinak<DETE> daude hau ikusteko (il95_9_j_a)
Erakunde horiek (EPPO) <DETE>bi bide desberdinak<DETE> jarraitzen ditu (il95_10_s_f)

3.3.3.22. Zenbakia + IZE + DETatzizkia + ADJ + DETatzizkia

Erregelak: DETE12C, DETE12D, DETE12E

3.3.4. DETO (DSko osagaien ordena okerra)

3.3.4.1. Zenbatzaile Zehaztugabea + IZE

Erregela: DETO1

<DETO>Gutxi pertsonak<DETO> ginen, baino oso polita zen (ae98_2_aa_a)
<DETO>asko posibilitate<DETO> dituzu egun zoriontsua izango duela (ae98_2_bu_a)
Orain dela <DETO>asko urte<DETO> ni goiz goizean altxatu nintzen (ae98_2_bi_a)
<DETO>Askok perretxiko<DETO> jan ditugu (ae98_2_bs_a)
<DETO>Askok jende<DETO> ezagutzen nintzen eta askenian denak lagunak izan ginen (ae98_4_ac_a)
ni arrapatu <DETO>asko perretxikoa<DETO> (ae98_2_ab_a)
Ni deitu nire lagunak ekarri ezazu saiheskia eta <DETO>asko ardoa<DETO> (ae98_2_ab_a)
Nire frenillo kendu zuen eta <DETO>asko odola<DETO> kampa atera zuen (ae98_2_aj_a)
Berak <DETO>asko ura<DETO> zituen (ae98_2_aw_a)
<DETO>Askok jendea<DETO> genuen klasean (ae98_2_bd_a)
<DETO>Askok denbora<DETO> pasa zuen (ae98_2_bi_a)
<DETO>asko egun zoriontsuak<DETO> egongo dela (ae98_2_bu_a)
Txoznetan <DETO>asko poteak<DETO> artzen genuen (ae98_2_i_a)
<DETO>asko gausak<DETO> hitz egin zuen (ae98_2_bd_a)
<DETO>Askok opariak<DETO> zizkiegun, baina asko jan zuten (edan ere bai) (ae98_2_aa_a)
Elizan <DETO>asko lagunak<DETO> joan ziren (ae98_2_p_a)
orduan <DETO>asko dirurik<DETO> behar da (ae98_2_ñ_a)
Txoznetan asko dantza nintzen eta <DETO>pixka bat kalimotzxo<DETO> edan nuen (ae98_2_al_a)
<DETO>pilo dirua<DETO> irabasiko dut (ae98_4_as_a)
Lo egin baino lehen <DETO>asko lizarrak<DETO> ikusi genuen (ae98_4_am_a)
piska bat urduri nengoen <DETO>asko gausak<DETO> egin behar nuela (ae98_5_ag_a)
langabezia nengoen eta <DETO>diru nahiko<DETO> bilatu nuen arazorik gabe bizitzeko (ae98_7_o_a)
Baditut bai batzuk oso onak eta <DETO>onak nahiko<DETO> ere bai (ae98_8_d_a)
lokomotzio-bideek dauzkaten <DETO>gehien istripuak<DETO> kotxeak direla (il95_9_g_f)
kotxerik ez daukagunok <DETO>samar arraroak<DETO> ematen dugu (il95_9_i_f)

3.3.4.2. Artikulu zehaztugabea + IZE

<DETO>bat ordu<DETO> nire gorputza oso nekatuta dago (ae98_2_ab_a)
bi garagardo edan ditugu eta <DETO>batzuk pintxo<DETO> jan ditugu (ae98_2_j_a)

3.3.5. DETMUG (MUGatasun kontuak)

3.3.5.1. Zenbatzaile Zehaztugabea + IZE + DETatzizkia deklinatuta

Erregela: DETMUG1

Izango du baliotasuna <DETMUG>zenbait kasuetan<DETMUG> (il92_9_n_g)
denok dakigu <DETMUG>zenbat jokalariek<DETMUG> jolastu behar duten partidua bat (il97_10_f_l)
Eta <DETMUG>hainbeste irudiekin<DETMUG> (tr05_12_or7_f)
<DETMUG>horrenbeste iritziekin<DETMUG> urte hau (edo etorkizuna) zer izango den... (tr05_12_or7_f)

3.3.5.2. Determinatzaile Zehaztugabea + IZE + DETatzizkia deklinatuta

Erregela: DETMUG2

<DETMUG>Edozein aldizkarietan<DETMUG> ea astero irtetzen da (ir90_7_a_a)
<DETMUG>Edozein pertsonen<DETMUG> minak gure gupida merezi du (ir90_7_e_a)
<DETMUG>zein momentuetan<DETMUG> komenigarria dena (il92_8_s_b)
"Abe Maritxu Garbia", eguneko <DETMUG>edozein unean<DETMUG> (il97_10_b_m)

3.3.5.3. Zenbakia + IZE + DETatzizkia deklinatuta

Erregela: DETMUG3

Artaxonan bizi nintzen nire gurasoak eta nire <DETMUG>5 anaiarrebarekin<DETMUG> (ae98_4_t_a)
Hori bai, <DETMUG>bi kalimotxoekin<DETMUG> kristona harrapatzen genuen (ae98_5_y_a)
<DETMUG>bi multzo nagusietan<DETMUG> sailkatzen dira (il95_12_e_e)

3.3.6. DETN (DETerminatzailea edo zenbatzailea Nahastu)

Arratsea gu ondartzat <DETN>asko beroa<DETN> zenituen (ae98_2_aw_a)
Hor <DETN>asko beroa<DETN> egin zuen (ae98_2_bñ_a)

hor <DET>asko haizea<DET> egin zuen (ae98_2_bñ_a)
 Egun honetan <DET>asko tritea<DET> izan zen (ae98_2_bñ_a)
 eta <DET>senidea osoari<DET> gonbidatu digute (ae98_2_az_a)
 baina bidaia <DET>luzea asko<DET> zen (ae98_4_m_a)
 <DET>asko ahalegin<DET> nuela eta askenean lortu nuen (ae98_5_ag_a)
 han <DET>gende osoa<DET> mozkortuta, dantzatzen, ohiukatzen etabar (ae98_5_k_a)
 oso ospetsuak izan ginen eta gero <DET>jende osoa<DET> ezagutzen gintuzte (ae98_7_ñ_a)
 Udan ez zuen <DET>batere euria<DET> egin (il92_9_g_b)

3.3.7. DETGK (DETG eta DETK sintagma berean)

Txoznan egon ginen eta lagun bat eta ni <DETK>moskorra haundi<DETK> hartu ginen (ae98_2_ao_a)
 eta <DETK>moskorra haundi<DETK> harrapatu genuen (ae98_4_c_a)
 eta askenean karajillo <DETK>moskorra haundi<DETK> arrpatu genuen (ae98_4_ba_a)

3.3.8. DET? (era isolatuan zuzentzat eman daitezkeen arren testuinguruan erroredunak direnak)

<DET?>Lauzpabost egunak<DET?> pasatu nituen eta gero Iruñeara etorri nintzen (ae98_2_au_a)
 <DET?>Hamabi orduak<DET?> pasatu eta gero hiri bat allegatu genuen (ae98_2_as_a)
 ni <DET?>lau txuletak<DET?> egin nituen eta asterketa atera (ae98_2_k_a)
 <DET?>10 garagardoak<DET?> erosi nituen (ae98_2_d_a)
 <DET?>Zortzi pertsonak<DET?> geunden eta bi kotxeak hartu genuen Urdaxera joatea (ae98_3_k_a)
 guk <DET?>lau anaiak<DET?> ginen (ae98_3_g_a)
 nire etxean Hiriberrian <DET?>8 pertsonak<DET?> bizi ginen (ae98_3_g_a)
 <DET?>Bizpahiru frogak<DET?> egin eta gero nirentzat zela lan postua esan zidaten (ae98_5_g_a)
 Bapatean <DET?>hiru neskak<DET?> agertu ziren primeran (ae98_5_bf_a)
 La Habanan <DET?>6 egunak<DET?> pasatu genituen (ae98_6_v_a)
 Mexikora joan ginen. <DET?>Hamasei egunak<DET?> pasa ginen (ae98_7_z_a)
 <DET?>atso bezala<DET?> bizitzea guztatzen ez zitzaidana (il92_8_g_d)
 <DET?>Laguna bezala<DET?> oso leial eta fidagarria dirudizu (il92_8_l_c)
 joan den astean <DET?>hiru emakume turkiarrak<DET?> hil bait dituzte (il92_8_g_e)
 normalean <DET?>idazkariak bezala<DET?> lan egiten dugu (il92_9_l_h)
 Ay <DET?>ze lagun<DET?> daukadan, beti dago protesta egiten (il95_9_i_e)
 Bestetik, <DET?>gatazka sozialak ugari<DET?> izan dira kanpaina honetan (il95_10_m_g)
 Ez zen kamioi normala, <DET?>bi trailerrak<DET?> eramaten dituztenetako bat baizik (il97_10_b_n)
 burura etortzen zitzaizkion <DET?>hamaika irudiak<DET?>. Oierren haurtzarokoak (il97_10_d_o)
 komunikabideek <DET?>hainbeste orri eta argazkiak<DET?> eskaini dizkiote (ir90_7_d_a)
 lehenengo begiratu <DET?>beharra<DET?> duena nola da haurtzaindegia (ir90_6_k_d)

gero kalera joan nintzen jolastera edo <DET?>beste gausa<DET?> egitera (ae98_2_bi_a)
 nahiago dut <DET?>beste gauza<DET?> egin (ae98_2_n_a)
 Diborzio eginez geroztik <DET?>beste gizona<DET?> harrapatzea hasi nintzen (ae98_4_ak_a)
 ez da klasetara joaten baina hori <DET?>beste gaia<DET?> da (il92_8_q_b)
 Baina <DET?>beste aukera<DET?> dago haitezula honetan gorderik dauden oroigarriak ezagutu (il95_12_e_d)
 baina ezin dut <DET?>beste plana<DET?> daukalako data horretan (tr05_1_or2_b)
 baina <DET?>beste aukera<DET?> edukiko dugu (tr05_1_or2_b)

<DET?>Lana asko<DET?> egin dut eta hori ez da ona (ae98_2_w_a)
 gero <DET?>asko lan<DET?> egin dut (ae98_2_bt_a)
 momentu horretan <DET?>elurra asko<DET?> bota du (ae98_2_as_a)
 <DET?>asko lan<DET?> egin nuen (ae98_6_r_a)
 Nik, <DET?>asko negar<DET?> egiten dut (ae98_7_m_a)
 <DET?>Asko lan<DET?> egitea eta gutxi jatea bukatu da guretzat (ir90_6_l_h)
 janari gutxi eman arren ondo doa, eta <DET?>asko lan<DET?> egiten du (il92_8_p_a)
 Biok (Adolfok eta zaldiak) <DET?>asko lan<DET?> egiten ba_zuten ere (il92_8_p_a)

<DET?>Eguraldia oso ona<DET?> zegoen ni zazpietan esnatu nintzen (ae98_3_p_a)
 Inoiz, zillarrezko <DET?>gauzak hain politak<DET?> ikusi nituen (ae98_7_z_a)
 ikasia amaitu ondoren zineman <DET?>lana<DET?> bila hasi zen (il97_10_f_u)

3.3.9. DET (zalantzazko kasuak)

guk ez ditugu <DET>dirurik asko<DET> (ae98_2_as_a)
 Hamabietatik ordu bitara <DET>potek asko<DET> hartu genituen (ae98_4_al_a)
 Hasieran ez genuen <DET>lanik asko<DET> (ae98_6_e_a)

ORT_AORG (A organikoa)**Bidaia**

Ez ginen joan lanera eta <AORG>bidai<AORG> oso luzea eta oso polita egin genuen (ae98_2_ap_a)
Karibetik <AORG>bidai<AORG> bat egin nahi genuen (ae98_2_e_a)
Orain dela bi urte <AORG>bidai<AORG> bat egin nuen (ae98_3_c_a)
esan zidan <AORG>bidai<AORG> bat zeukanela bi pertsonarentzat (ae98_3_d_a)
<AORG>Bidai<AORG> honetan aberatz bat ezagutuko nuke (ae98_4_ar_a)
Haiek <AORG>bidai<AORG> bat denontzat prestatu zuten (ae98_4_av_a)
<AORG>Bidai<AORG> bukatu bano lehen istripu bat ala percanze bat genuen (ae98_5_b_a)
klase guztia <AORG>bidai<AORG> bat prestatu genuen (ae98_5_g_a)
nola gustatuko litzaidake <AORG>bidai<AORG> bat egin (ae98_5_g_a)
8,30etan geratu ginen <AORG>Bidai<AORG> bat egitera "Ordesara" (ae98_6_d_a)
<AORG>Bidai<AORG> osoa hitz eginez, far eginez, abestuz egon ginen (ae98_6_d_a)
<AORG>Bidai<AORG> erditik bokata batzuk ez genituen jan baizik eta kampion gas atera genuen (ae98_6_d_a)
Tensioa pasa eta gero gure <AORG>bidairekin<AORG> jarraitzea erabaki genuen (ae98_6_l_a)
Niretzat <AORG>bidai<AORG> onena izan zen (ae98_6_v_a)
Egon ginen <AORG>bidai<AORG> osoa hitz egiten (ae98_6_y_a)
ikas <AORG>bidai<AORG> bat egin nuen nere klaseko jendearekin (ae98_7_af_a)
Hori izan zen nere lehen <AORG>bidai<AORG> luzeagoa (ae98_7_af_a)
seguraski nere biziko <AORG>bidai<AORG> hoberena (ae98_7_af_a)
joan nintzen lagun batzuekin <AORG>bidai<AORG> bat egitera (ae98_7_o_a)
gaurko Euskal Herria ez da aspaldiko <AORG>bidai<AORG> hartan ezagutu zenuena (il92_9_m_e)
ez zen mugitu <AORG>bidai<AORG> osoan (il92_9_k_c)
<AORG>Bidai<AORG> luze hori egiteko hari pilo bat behar zuen (il92_9_j_b)
zaldiz egin genituen <AORG>bidai<AORG> hauek (il95_12_i_b)
nire andrearekin <AORG>bidai<AORG> bat egitea erabaki nuen (il97_10_f_o)
azken 20 urteak oso garrantzitsuak izan ziren eta <AORG>bidai<AORG> (il97_10_k_u)

Gauza

Denmok batera <AORG>gauz<AORG> pilo prestatu genuen dirua ateratzeko (ae98_5_g_a)
eta <AORG>gauz<AORG> bat esango nuke pena merezi duela momentu hori bizitzea (ae98_5_ab_a)
ikus daiteke <AORG>gauz<AORG> asko ez daudela zuzen (ae98_6_x_a)
Gero eta <AORG>gauz<AORG> gehiago ikasten ditut (ae98_6_x_a)
Nire pentsamenduak egun osoa <AORG>gauz<AORG> bat zeuden (ae98_7_c_a)
Egunagatik <AORG>gauz<AORG> guztiak ondo ikusten dira, argi gehiago dago (il92_8_j_a)
eta <AORG>gauz<AORG> pilo bat eta abar (il92_8_j_a)
aintzinean, <AORG>gauz<AORG> guztiak elkarri hitzegiten zioten (il92_9_b_a)
Baina <AORG>gauz<AORG> guztien artean hizketan omen ez zekiena goroldioa zela (il92_9_b_a)
Mutilak edozein gauza egiteko <AORG>gauz<AORG> da atentzioa deitzeko (il92_9_h_d)
Futbola milaka pertsona mugitu ditzakeen <AORG>gauz<AORG> bakarra izan daiteke (il97_10_h_l)
Kezkatzen zaidan beste <AORG>gauz<AORG> bat komunikabideetan futbolak duen tratamendua da (il97_10_h_l)
eta beste <AORG>gauz<AORG> batzuegatik kezkatzea (il97_10_h_l)

Animalia

Galdetu zigun zergatik ez genuen <AORG>animali<AORG> bat etxean (ae98_5_a_a)
Nire ahizpa eta biok beti saiatu gara <AORG>animali<AORG> bat edukitzen (ae98_5_a_a)
<AORG>Animali<AORG> guztien artean gizakiengan sortzen duen laguntasuna (il92_9_b_c)
ikusi ahal izan dugu <AORG>animali<AORG> lagungarri hau (il92_9_b_c)
Nere amari ez zitzaion batere gustatzen <AORG>animalirik<AORG> (il92_9_k_c)
lan egiteko <AORG>animali<AORG> bakarra bizirik zegoenez (il92_9_i_b)
inguruan dabilta <AORG>animali<AORG> guztiak (il95_10_j_d)
nahiz eta <AORG>animali<AORG> basati izan, oso soziablea da (il95_10_j_d)

Pertsona / Pertsonai

Lore bat horma batean, ekaitz bat, <AORG>pertson<AORG> arraro bat, edozein gauza (ae98_6_a_a)
beste <AORG>pertson<AORG> batekin patatak biltzen hasi nintzen (ae98_6_e_a)
beste <AORG>pertson<AORG> batzuen larritu gabe (ir90_7_d_a)
Bakarrik hiru <AORG>pertsonai<AORG> (ir90_7_g_a)
ez zen <AORG>pertsonai<AORG> famatua (il92_8_o_e)
Gure <AORG>pertsonai<AORG> honen gehiena gustatzen zaidana ausarta dela da (il92_8_f_b)
nahiz eta hemengo kartzelen barruan ere <AORG>pertsonai<AORG> inportanteak izan (il95_10_s_a)
irakurleak <AORG>pertsonai<AORG> ezberdinekin identifikatzea daukadalako (il97_10_k_n)
detektibea <AORG>pertsonai<AORG> bitxia izan da ingeles literaturan (il97_10_k_u)
Nire <AORG>pertsonairik<AORG> preziatuenak emakume helduak dira (il97_10_k_n)

Brownen arrakasta <AORG>pertsonai<AORG> berean datza (il97_10_k_u)

Hizkuntza

<AORG>Hizkuntz<AORG> bat gero eta galdutagoa denean (il92_9_n_e)
 Batez ere lku berean beste <AORG>hizkuntz<AORG> ofiziala dagoenean (il92_9_n_e)
 hauek <AORG>hizkuntz<AORG> batetik bestera itzultzen grazia galtzen dute (il92_9_a_f)
 ulergaitzak bihurtzen dira <AORG>hizkuntz<AORG> batetik beste batera pasatzerakoan (il92_9_b_g)
 hitz jokuetan oinarritzen direnak <AORG>hizkuntz<AORG> batetik beste batera (il92_9_b_f)
 horren aurrean daukagun jarrera <AORG>hizkuntz<AORG> politika bat suposatzen da (il95_10_g_e)
 eta aurre egiteko behar du <AORG>hizkuntz<AORG> politika bat (il95_10_g_e)
 elkarbizitzea da <AORG>hizkuntz<AORG> politikaren helburua (il95_10_g_e)

Familia

jende gutxi bizi ziren, beste bi edo hiru <AORG>famili<AORG> (il92_8_n_a)
 Behin <AORG>famili<AORG> oso bat suaren aurrean gosez elkarrekin hil zela (il92_9_j_b)
 ijito <AORG>famili<AORG> guztiak sakabanatuta izan behar ziren (il95_9_g_a)
 <AORG>Famili<AORG> guztietan nabaritzen dira aldaketa hauek (tr05_12_or13_i)
 <AORG>Famili<AORG> bakoitzak diru pilo bat gastatzen du (tr05_12_or13_i)

Gastronomia

<AORG>gastronomi<AORG> elkarte bat egin genuen (il92_7_g_a)
 Beste aldetik <AORG>gastronomi<AORG> ohiturak nondik datozte (il92_7_g_a)
 kultur, kirol edo <AORG>gastronomi<AORG> elkartearen idazkaria naizen honek (il92_7_h_a)

Anaia

gainera beste <AORG>anai<AORG> du (ae98_2_ae_a)
 <AORG>Anai<AORG> bat naukan eta oso onak ziren gure harremanak (ae98_5_ñ_a)
 ez du lagunik, ezta bere <AORG>anairekin<AORG> harremanik ere (il92_9_l_d)

Egun bat, udan, <AORG>Donostira<AORG> joan ginen (ae98_4_t_a)
 niri bururatu zitzaidan <AORG>idei<AORG> bat (ae98_6_q_a)
 allegatzerakoan <AORG>denborik<AORG> galdu gabe bota genituen gura kainak (ae98_6_q_a)
 gure <AORG>Euskaltzaindiren<AORG> bigarren frakzio-taldea ekintza (ae98_7_f_a)
 <AORG>Heriotz<AORG> hauen artean ez dago ezberdintasunik, (ir90_7_c_a)
 <AORG>heriotz<AORG> baten berria diruko iturri bat izan daitekela (ir90_7_h_a)
 Onar iezadazu <AORG>alukeri<AORG> bat izan dela (ir90_7_g_a)
 <AORG>Alukeri<AORG> bat, bai, (ir90_7_g_a)
 ikasleek heziketa ona lortzeko <AORG>tresneri<AORG> hauek sarritan erabili behar dutela (ir90_6_u_a)
 esku batean <AORG>kapel<AORG> beltz eta bestea poltsikoan zuela topo egin zuen (ir90_6_s_l)
 kontutan hartuz <AORG>ezkontz<AORG> hura ez zela batere aproposa (ir90_7_d_a)
 ondo iruditzen zaigu <AORG>idei<AORG> hau (il92_7_j_a)
 Gure <AORG>kultur<AORG> gastronomikoa zabaltzen du atzerrian (il92_7_g_a)
 <AORG>Donostira<AORG> etorri ginenean (il92_9_f_c)
 lehiotik begiratu zuenean <AORG>paisai<AORG> eder bat ikusi zuen (il92_8_q_a)
 ehin <AORG>Polonin<AORG> zegoenean (il92_8_t_e)
 esan dezaket <AORG>arrazakeri<AORG> mota asko daudela (il92_8_d_e)
 <AORG>portzentai<AORG> hori ehuneko hogeitaseira igo egiten da (il92_9_m_h)
 hauteskundeetako <AORG>emaitz<AORG> kaskarrak onartzeko (il92_9_a_f)
 noizbehinka <AORG>izkanbil<AORG> batzuk paioekin eduki arren (il95_9_g_a)
 baina <AORG>lelokeri<AORG> bat esango zidan jakinda bururatu zitzaidan aitaren batean sartzea (il95_9_t_d)
 Kotxearen orokorreko erabilerak oso <AORG>emaitz<AORG> txarrak ekarri dizkigu (il95_9_t_f)
 Ospitalean zaurituta eraman ninduten, ni eta zaharra <AORG>anbulantzi<AORG> berberan (il95_10_r_j)
 edozein lekutara joateko <AORG>distantzi<AORG> handiak dituzte (il95_12_i_b)
 <AORG>Erreportai<AORG> honetan ikusi dugunez (il95_12_i_d)

Euskara Teknikoko ikasleen testuetan etiketatutako determinatzaile-erroreak

3.3.1. DETK

3.3.1.1. IZE + DET kendu

LAZTANEK GENEETAN <DETK>ERAGIN<DETK> DUTE (ET-12_g_e)

Hauetan, jaioberrien <DETK>estres<DETK> eta antsietatea aztertzeo arratoiak erabili ziren (ET-12_g_e)

estresarekin loturiko hormonon <DETK>kontzentrazio<DETK> txikiagoa zen arratoi hauetan (ET-12_g_e)

Arnas arazo gehienak BGBK-ak (biriketako <DETK>gaixotasun buxatzaile kroniko<DETK>) dira (ET-12_a_m)

aldi bereab <DETK>medikazio<DETK> banatzen dugu (ET-12_a_m)

Gaur egun, mundu mailan 7-5 milioi pertsona dute <DETK>Alzheimer<DETK> (ET-12_g_k)

Percepción: pertzepzioa/<DETK>hautemate<DETK> (ET-12_r_q)

Proceso: <DETK>prozesu<DETK> (ET-12_r_q)

Productividad: <DETK>emankortasun<DETK>/produktibitatea (ET-12_r_q)

Orientación educativa y psicopedagógica: Psikopedagogi eta <DETK>hezkuntza-orientabide<DETK> (ET-12_r_q)

Departamentos de Orientación: Orientabide <DETK>departamentu<DETK>. (ET-12_r_q)

Salud mental: <DETK>Adimen-osasun<DETK> (ET-12_r_q)

Diseño y desarrollo curricular: Curriculu <DETK>diseinu<DETK> eta garapena (ET-12_r_q)

3.3.1.2. IZE + ADJ + DET kendu

Atención temprana : <DETK>Laguntza goiztiar<DETK> (ET-12_r_q)

Fruta eta barazkiak jan oso maiz, baita zuntz <DETK>kantitate handi<DETK> duten zerealeak (ET-12_or2_j)

3.3.1.4. DETK aditz-esapideetan

lasai <DETK_HAUL>arnas hartu<DETK_HAUL> eta deskonektatzeko momenturen bat (ET-12_p_ñ)

3.3.1.6. DETK izenki-predikatuei

bere deskribapenari <DETK_IP>profil esaten <DETK_IP> zaio (ET-12_l_d)

Nire informazioa <DETK_IP>lagungarri izango<DETK_IP> zaizulakoan (ET-12_k_f)

Praktiketako egun bat azaltzea oso <DETK_IP>luze egin <DETK_IP> daiteke (ET-12_e_m)

ondorioz askotan <DETK_IP>zail egiten<DETK_IP> zaie nahi dutena ongi azaltzea (ET-12_i_f)

Bien arteko erlazioari <DETK_IP>ganbio tipo deritzo<DETK_IP> (ET-12_q_o)

hezkuntza <DETK_IP>derrigorrezko den<DETK_IP> zerbait dela erabat ados nago (ET-12_r_q)

<DETK_IP>euskaldun berri naiz<DETK_IP> (ET-12_w_r)

3.3.1.7. DETK hitza deklinatzean

<DETK_DEKL>Irtenbidei<DETK_DEKL> begirata badakizu arlo honetan lan egiteko EGA edukitzea ona dela (ET-12_a_f)

Egunero sartzean <DETK_DEKL>gauen<DETK_DEKL> gertaturikao entzuten dugu (ET-12_a_m)

ebakuntza <DETK_DEKL>baten<DETK_DEKL> pasrte hartzeko... (ET-12_e_m)

eta gero <DETK_DEKL>bazkarik<DETK_DEKL> iristean berauek banatu (ET-12_e_m)

Goizeko <DETK_DEKL>>6:45tan<DETK_DEKL> iratzargailuak jotzen duen (ET-12_h_m)

<DETK_DEKL>8:00tan<DETK_DEKL> ordurarte gertatutako gorabehera guztiak azaltzen dizkigute (ET-12_h_m)

<DETK_DEKL>8:30tan<DETK_DEKL> gelaz gela, gaixo bakoitzak behar dituen medikazioak banatzen (ET-12_h_m)

<DETK_DEKL>9:00tan<DETK_DEKL> konstanteak hartzen ditugu (ET-12_h_m)

<DETK_DEKL>10:00tarako<DETK_DEKL> konstante guztiak hartuta eta grafiketara pasatuta (ET-12_h_m)

<DETK_DEKL>11:00tan<DETK_DEKL> sendaketak egiten dira (ET-12_h_m)

Lana niretzat goizeko <DETK_DEKL>8tan<DETK_DEKL> hasten da (ET-12_j_m)

Goizeko <DETK_DEKL>10tan<DETK_DEKL> gosaria ailegatzen da (ET-12_j_m)

<DETK_DEKL>12tako<DETK_DEKL> medikazioa banatu, 1etan bazkaria (ET-12_j_m)

eguna 2004-ko urriaren 13an eguerdiko <DETK_DEKL>13-tan<DETK_DEKL> amaituko da (ET-12_k_i)

zientzia eta teknologia fakultatean eta kimika-fisika <DETK_DEKL>sailan<DETK_DEKL> (ET-12_k_i)

Honen <DETK_DEKL>lekukotako<DETK_DEKL> bat dira ikastetxeak (ET-12_p_p)

Orduan bai <DETK_DEKL>gizabanakori<DETK_DEKL> askatasuna kenduko liokeela (ET-12_r_q)

3.3.2. DETG (DETerminatzaile atzizkia Gehitu)

3.3.2.1. Determinatzaile atzizkia gehitu sintagman

Orduan <DETG>guztia hau<DETG> jakinda zure erabakia zure eskuetan dago, (ET-12_i_f)

3.3.2.3. DETG postposizio-egituretan

<DETG_POS>Beraiei eskerrak<DETG_POS> ikusten dugu aurrerago erizain bezala balio izango dugun (ET-12_e_m)

3.3.2.4. DETG izenki-predikatuei

laztanek, beste mekanismo batzuetan ere eragina izan dezaketela <DETG_IP>posiblea da <DETG_IP> (ET_12_n_e)

3.3.3. DETE (DETerminatzailea Errepikatu)

<DETE>Informazioa egokiaz<DETE> jabetzeko ezinbestekoak izango dira (ET_12_p_ñ)

honen baitan <DETE>pertsonak bakoitzak<DETE> egiten dituen aukeraketak (ET_12_p_p)

Behin .ZIP fitxategia zuen makinara jaitsi ondoren, <DETE>fitxategia hau<DETE> deskonprimitu (ET_12_or2_j)

3.3.4. DETO (DSko osagaien ordena okerra)

Orduan <DETO>guztia hau<DETO> jakinda zure erabakia zure eskuetan dago (ET_12_i_f)

3.3.5. DETMUG (MUGatasun kontuak)

Erizaintza ikasketak iraundutako <DETMUG>hiru urteetan<DETMUG> zehar (ET_12_g_m)

Bukatzeko, azken <DETMUG>3 asteetan<DETMUG> pediatrian egon nintzen (ET_12_n_m)

<DETMUG>Zenbait politikoen<DETMUG> eskuetan gelditzen bada (ET_12_p_p)

<DETMUG>40 balore nagusiez<DETMUG> osatua (ET_12_q_o)

baso guztiak berdinak ez direla ikusiko dugu; <DETMUG>zenbait adibideen<DETMUG> bidez (ET_12_y_u)

ez zait bidezkoa iruditzen <DETMUG>zenbait ikasleek<DETMUG> haiekin inoiz egotea nahi ez izatea (ET_12_g_m)

3.3.8. DET? (era isolatuan zuzentzat eman daitezkeen arren testuinguruan erroredunak direnak)

<DET?>erizaina bezela<DET?> aritzeko eskari handia dago (ET_12_j_f)

Euskal Herrian geratzen bazara, <DET?>irtenbide profesionalak gehiago<DET?> izango dituzu (ET_12_w_r)

3.3.9. DET (zalantzazko kasuak)

glukokortikoedetan egiten dute <DET>eragin<DET> egoera latz baten kontrako erantzuna murrizten (ET_12_a_e)

ORT_AORG (A organikoa)

ezinbestekoa da zerbitzu arloan ahalik eta <AORG>hizkuntz<AORG> gehien menperatzea (ET_12_g_f)

<AORG>animali<AORG> basatien inguruko ikerketa piloa egin ahal izango dituzu (ET_12_t_r)

E eranskina:

Determinatzaile-erroreak detektatzeko erregelak

(1)	Errore mota eta kategoria/azpikategoria: 3.3.1.1. DETK: <i>Izena + determinatzailea</i> kendu motako sintagma.
(2)	Errorearen deskripzioa: Determinatzailea ezinbestekoa da argumentuetan; IS orok behar du determinatzaile ageri bat, Ø determinatzailea zilegiztatzen duen zenbatzailerik ez badarama.
(3)	Errore motaren adibide batzuk: *Ondoren <DETK>afari<DETK> prestatu genuen. (1. erregela) *<DETK>Donostiako<DETK> da. (2. erregela) *<DETK>Txupinazo<DETK> eta gero hasi zen. (3. erregela) *<DETK>Garagardo<DETK> edan ondoren. (4. erregela) *Gainera beste <DETK>anai<DETK> du. (5. erregela) *Gure <DETK>helburu<DETK> lortu dugu. (6. erregela) *Nire <DETK>irakasle<DETK> ezin zuen hori ulertu. (7. erregela) *Bere <DETK>amodio<DETK> aitortu zidan. (8. erregela) *Bere <DETK>irakasle<DETK> ezin zuen hori ulertu. (9. erregela) *Haiei <DETK>jende<DETK> jatorra da. (10. erregela) *Zure <DETK>lagun<DETK> atsegina zen. (11. erregela) *Nire <DETK>auto<DETK> oso handia da. (12. erregela) *Haren <DETK>lankide<DETK> oso jatorra da. (13. erregela) *Gure <DETK>kanpamendu<DETK> ez genuen jarri. (14. erregela) *Bere <DETK>soineko<DETK> ez da polita. (15. erregela) *Bere <DETK>urtebetetze<DETK> gaur da. (16. eta 17. erregelak)
(4)	Erroreari dago(z)kion zuzenketa(k): Ondoren afaria prestatu genuen. Donostiakoa da. Txupinazoa eta gero hasi zen. Garagardoa edan ondoren. Gainera beste anai bat du. Gure helburua lortu dugu. Nire irakasleak ezin zuen hori ulertu. Bere amodioa aitortu zidan. Bere irakasle batek ezin zuen hori ulertu. Haiei jendea jatorra da. Zure lagun hura atsegina zen. Nire autoa oso handia da. Haren lankide bat oso jatorra da. Gure kanpamendua ez genuen jarri. Bere soinekoa ez da polita. Bere urtebetetzea gaur da.
(5)	Errorearen iturria: Euskara-ikasleen corpora eta Zubiri eta Zubiri (1995).
(6)	Oharra(k): Honelako kasu gehienak automatikoki detektatzea zaila denez, egitura jakin batzuekin bakarrik saiatu gara mementoz. HAUL asko HAUL seguru bezala tratatzen ez direnez, DETK1 erregelari HAULEn zerrendan ez-ziurrak diren adibide batzuk kontuan hartu ditugu (ikus errorearen gramatikako LIST HAUL zerrenda). Horiez gain, HAULEn zerrendan ez dauden arren, hitz segida finko hauek sartu ditugu, alarma faltsuak ekiditearren: ARGÍ UTZÍ/GERATU, ARRAZOÍ IZAN, AUTO-STOP EGIN, BAIETZ/EZETZ ERANTZUN, ARRO EGON, PORRU EGINDA EGON, UMEZURTZ GERATU, EZEREZ ESAN...

(7)	Erregela(k): (1. erregela) ADD (&OKER_DETK1_0) TARGET IZE IF (NOT -2 DET-ZEHAZTUA OR NMGP) (NOT -1 BESTE OR DET-ZEHAZTUA OR DET-ZEHAZTUGABEA OR DET-NOLGAL OR DET-NOLARR OR DET-ERKARR OR DET-ERKIND OR IOR OR INS OR LITRO OR METRO OR MILIOI OR ASKO OR ("ehunka") OR ("milaka") OR ("milioika") OR ("bilioika")) (0 IZE + ARR + ABS + MG) (NOT 0 AORG OR ENTI_ORG OR ENTI_PER OR ZEN_DEK OR IZAN OR ARTE OR SEG-MIN-ORDU-ASTE-HIL-URTE OR LITRO OR METRO OR ORDUBETE OR HILABETE OR ERDI OR LAURDEN OR USTE OR EZIN OR AHAL OR BEHAR OR NAHI OR HAUL OR ("gabe") OR ("laster") OR ("erantzule") OR ("gustu") OR ("elbarri") OR ("itsu") OR PILO OR ARGİ OR ARRAZOI OR AUTO-STOP OR BAIETZ OR EZETZ OR MEDIO OR ARRO OR PORRU OR UMEZURTZ OR EZEREZ OR KONP OR ABL OR ABU OR ABZ OR ALA OR DAT OR DES OR ERG OR GEN OR INE OR INS OR MOT OR PAR OR SOZ OR ({POS-BUK) OR (@ADLG)) (1 ADT OR ADI) (NOT 1 NAHI OR BEHAR OR MUGATUA OR IZAN OR AUKERATU OR HAUTATU OR BIHURTU OR BILAKATU OR IZENDATU OR DEITU OR SARTU) (NOT 2 IZENA OR ABIZENA); (2. erregela) ADD (&OKER_DETK1A_1) TARGET IZE IF (0 IZE + GEL) (NOT 0 ("aurka") OR ("kontra") OR ("alde")) (1 IZAN OR EGON); (3. erregela) ADD (&OKER_DETK1B_2) TARGET IZE IF (NOT -1 DET-ZEHAZTUA OR DET-ZEHAZTUGABEA OR DET-NOLARR OR DET-NOLGAL) (0 IZE + ARR + ZERO) (NOT 0 DET OR SEG-MIN-ORDU-ASTE-HIL-URTE OR AORG OR PKOMA OR POSTPOSIZIOAK-1 OR POSTPOSIZIOAK-2 OR POSTPOSIZIOAK-3 OR POSTPOSIZIOAK-4 OR POSTPOSIZIOAK-5 OR POSTPOSIZIOAK-6 OR POSTPOSIZIOAK-6-1 OR POSTPOSIZIOAK-7 OR POSTPOSIZIOAK-8 OR POSTPOSIZIOAK-9 OR POSTPOSIZIOAK-10 OR POSTPOSIZIOAK-10-1 OR POSTPOSIZIOAK-11 OR POSTPOSIZIOAK-11-1 OR POSTPOSIZIOAK-12 OR POSTPOSIZIOAK-13 OR POSTPOSIZIOAK-14 OR POSTPOSIZIOAK-15) (1 ETA) (2 ADB-GERO); (4. erregela) ADD (&OKER_DETK1C_0) TARGET IZE IF (NOT -2 DET-ZEHAZTUA OR NMGP) (NOT -1 BESTE OR DET-ZEHAZTUA OR DET-ZEHAZTUGABEA OR DET-NOLGAL OR DET-NOLARR OR DET-ERKARR OR DET-ERKIND OR IOR OR INS OR LITRO OR METRO OR MILIOI) (0 IZE + ARR + ABS + MG) (NOT 0 AORG OR ENTI_ORG OR ENTI_PER OR IZAN OR ABL OR ABU OR ABZ OR ALA OR DAT OR DES OR ERG OR GEN OR INE OR INS OR MOT OR PAR OR SOZ OR ({POS-BUK) OR ARTE OR SEG-MIN-ORDU-ASTE-HIL-URTE OR LITRO OR METRO OR ORDUBETE OR HILABETE OR ERDI OR LAURDEN OR USTE OR EZIN OR AHAL OR BEHAR OR NAHI OR HAUL OR ("laster") OR ("erantzule") OR PILO OR ARGİ OR AUTO-STOP OR BAIETZ OR EZETZ OR MEDIO OR ARRO OR PORRU OR UMEZURTZ OR EZEREZ OR (@ADLG)) (1 ADI) (NOT 1 AUKERATU OR HAUTATU OR BIHURTU OR BILAKATU OR IZENDATU OR DEITU OR SARTU) (2 ONDOREN);
-----	---

(5. erregela)

```
ADD (&OKER_DETK4A_1) TARGET BESTE IF (NOT -1 NAHI-BEHAR)
(0 BESTE)
(NOT 0 ABL OR DAT OR PAR OR GEN)
(1 IZE + ARR + MG)
(NOT 1 AORG OR ABL OR ABU OR ABZ OR ALA OR
DAT OR DES OR ERG OR GEN OR INE OR INS OR
MOT OR PAR OR SOZ OR
SEG-MIN-ORDU-ASTE-HIL-URTE)
(2 ADT OR ADI)
(NOT 2 DET-ZEHAZTUA OR DET-ZEHAZTUGABEA OR
DET-ERKARR OR ADJ OR ETA OR PUNTUAZIO_MARKA);
```

(6. erregela)

```
ADD (&OKER_DETK5A1_0) TARGET IZE IF (-1 IOR + PERARR + GEN)
(0 IZE + ARR + ABS + MG)
(NOT 0 AORG OR GEL OR GEN OR ("berri") OR
("zain") OR ("esku") OR ({POS-BUK}))
(1 ADT OR ADI)
(NOT 1 ETA OR EDO OR DET OR ({POS-BUK}) OR
IZAN OR AUKERATU OR HAUTATU OR BIHURTU OR
BILAKATU OR IZENDATU OR DEITU OR EGIN OR
SARTU);
```

(7. erregela)

```
ADD (&OKER_DETK5A2_0) TARGET IZE IF (-1 IOR + PERARR + GEN)
(NOT -1 DES)
(0 IZE-ARR-ZERO)
(NOT 0 AORG OR DEK OR ("berri") OR
("zain") OR ("esku") OR ({POS-BUK}))
(1 ADT OR ADI)
(NOT 1 ADJ OR DET OR IZE OR ETA OR EDO OR
({POS-BUK}));
```

(8. erregela)

```
ADD (&OKER_DETK5A3_0) TARGET IZE IF (-1 DET-ERKIND-ERKARR + GEN)
(0 IZE + ARR + ABS + MG)
(NOT 0 AORG OR GEL OR GEN OR ("berri") OR
("zain") OR ("esku") OR ({POS-BUK}))
(1 ADI OR ADT)
(NOT 1 ETA OR EDO OR DET OR ({POS-BUK}) OR
IZAN OR AUKERATU OR HAUTATU OR BIHURTU OR
BILAKATU OR IZENDATU OR DEITU OR EGIN OR
SARTU);
```

(9. erregela)

```
ADD (&OKER_DETK5A4_0) TARGET IZE IF (-1 DET-ERKIND-ERKARR + GEN)
(NOT -1 DES)
(0 IZE-ARR-ZERO)
(NOT 0 AORG OR DEK OR GEN OR ({POS-BUK}))
(1 ADT OR ADI)
(NOT 1 ADJ OR DET OR IZE OR ETA OR EDO OR
({POS-BUK}));
```

(10. erregela)

```
ADD (&OKER_DETK5B1_0) TARGET IZE IF (-1 DET-ERKIND-ERKARR AND SING-PLUR + GEN)
(NOT -1 DES)
(0 IZE + ARR + ZERO)
(NOT 0 AORG OR GEL OR GEN)
(1 ADJ + ABS + NUMS)
(2 ADT);
```

(11. erregela)

```
ADD (&OKER_DETK5B2_0) TARGET IZE IF (-1 IOR + PERARR + GEN)
                                     (NOT -1 DES)
                                     (0 IZE + ARR + ZERO)
                                     (NOT 0 AORG OR GEL OR GEN)
                                     (1 ADJ + ABS + NUMS)
                                     (2 ADT);
```

(12. erregela)

```
ADD (&OKER_DETK5C1_0) TARGET IZE IF (-1 IOR + PERARR + GEN)
                                     (NOT -1 DES)
                                     (0 IZE + ARR + ZERO)
                                     (NOT 0 AORG OR GEL OR GEN)
                                     (1 OSO + ADB)
                                     (2 ADJ + ABS + NUMS)
                                     (3 ADI OR ADT);
```

(13. erregela)

```
ADD (&OKER_DETK5C2_0) TARGET IZE IF (-1 DET-ERKIND-ERKARR AND SING-PLUR AND GEN)
                                     (NOT -1 DES)
                                     (0 IZE-ARR-ZERO)
                                     (NOT 0 AORG + GEL + GEN)
                                     (1 OSO + ADB)
                                     (2 ADJ + ABS + NUMS)
                                     (3 ADI OR ADT);
```

(14. erregela)

```
ADD (&OKER_DETK5D1_0) TARGET IZE IF (-1 PERARR + GEN)
                                     (0 IZE + ARR + ABS + MG)
                                     (NOT 0 AORG OR GEL OR GEN OR HAUL)
                                     (1 EZ)
                                     (2 ADT OR ADL);
```

(15. erregela)

```
ADD (&OKER_DETK5D2_0) TARGET IZE IF (-1 DET-ERKIND-ERKARR + GEN)
                                     (0 IZE + ARR + ABS + MG)
                                     (NOT 0 AORG OR GEL OR GEN OR HAUL)
                                     (1 EZ)
                                     (2 ADT OR ADL);
```

(16. erregela)

```
ADD (&OKER_DETK6A_0) TARGET IZE IF (NOT -1 DET-ZEHAZTUA OR DET-ZENB-ZG)
                                     (0 IZE + ARR + ABS + MG)
                                     (NOT 0 AORG OR GEL OR GEN)
                                     (1 ("atzo") OR ("gaur") OR ("bihar"))
                                     (NOT 1 ({POS-BUK}))
                                     (2 ADT OR ADI);
```

(17. erregela)

```
ADD (&OKER_DETK6B_0) TARGET IZE IF (NOT -1 DET-ZEHAZTUA OR DET-ZENB-ZG)
                                     (0 IZE + ARR + ZERO)
                                     (NOT 0 AORG OR GEL OR GEN)
                                     (1 ("atzo") OR ("gaur") OR ("bihar"))
                                     (NOT 1 ({POS-BUK}))
                                     (2 ADT OR ADI);
```

(1)	Errore mota eta kategoria/azpikategoria: 3.3.1.2. DETK: <i>Izena + adjektiboa + determinatzailea</i> kendu motako sintagma.
(2)	Errorearen deskripzioa: Izenaz eta adjektiboz osaturiko determinatzailek gabeko Izen-Sintagma erroreduna da, Ø determinatzailea zilegiztatzen duen zenbatzailek ez badarama.
(3)	Errore motaren adibide batzuk: *<DETK>Erresaka handi<DETK> nuen. (1. erregela) *<DETK>Arazo larri<DETK> da hori. (2. erregela) *Beste <DETK>hegazkin handi<DETK> hartu nuen. (3. erregela)
(4)	Erroreari dago(z)kion zuzenketa(k): Erresaka handia nuen. Arazo larria da hori. Beste hegazkin handi bat hartu nuen.
(5)	Errorearen iturria: Euskara-ikasleen corpora.
(6)	Oharra(k): Honek kasu gehienak automatikoki detektatzea zaila denez, egitura jakin batzuekin bakarrik saiatu gara mementoz.
(7)	Erregela(k): (1. erregela) ADD (&OKER_DETK2A_1) TARGET IZE IF (NOT -2 DET-ZEHAZTUA OR DET-ZEHAZTUGABEA OR DET-NOLARR OR DET-NOLGAL) (NOT -1 DET-ZEHAZTUA OR DET-ZEHAZTUGABEA OR DET-NOLARR OR DET-NOLGAL) (0 IZE + ARR + ZERO) (NOT 0 ABL OR ABU OR ABZ OR ALA OR DAT OR DES OR ERG OR GEN OR INE OR INS OR MOT OR PAR OR SOZ OR SEG-MIN-ORDU-ASTE-HIL-URTE OR AHAL OR EZIN) (1 ADJ + ARR + ABS + MG) (NOT 1 GEHIEGI OR ASKO OR GUTXI OR GUTI OR HAUL OR (@ADLG)) (2 ADT OR ADI) (NOT 2 IZAN OR MUGATUA OR AUKERATU OR HAUTATU OR BIHURTU OR BILAKATU OR IZENDATU OR DEITU OR EGIN OR SARTU); (2. erregela) ADD (&OKER_DETK2B_1) TARGET IZE IF (NOT -1 BESTE OR DET-ZENB-ZG OR DET-ZEHAZTUA) (0 IZE + ARR + ZERO) (1 ADJ + ARR + ABS + MG) (NOT 1 (@ADLG)) (2 IZAN); (3. erregela) ADD (&OKER_DETK4B_2) TARGET BESTE IF (NOT -1 NAHI-BEHAR) (0 BESTE) (NOT 0 ABL OR DAT OR PAR) (1 IZE + ARR + ZERO) (2 ADJ + ARR + ABS + MG) (NOT 2 (@ADLG)) (3 ADT OR ADI) (NOT 3 DET-ZEHAZTUA OR DET-ZEHAZTUGABEA OR DET-ERKARR OR ADJ OR ETA OR PUNTUAZIO_MARKA);

(1)	Errore mota eta kategoria/azpikategoria: 3.3.3.1. DETE: <i>Izena + determinatzaile atzizkia + adjektiboa + determinatzaile atzizkia</i> motako sintagma.
(2)	Errorearen deskripzioa: Izena eta adjektibo batez osaturiko IS batean, determinatzaile atzizki singularra (-a) edota plurala (-ak) biek (izenak eta adjektiboak) badaramate, sintagma erroreduna da; IS batek ezin ditu hartu bi determinatzaile.
(3)	Errore motaren adibide batzuk: *<DETE>Mina handia<DETE> sentitzen nuen oinean. (1. erregela) *<DETE>Mozkorra handiarekin<DETE> azaldu zen. (2. erregela) *Oso <DETE>etxea polita<DETE> da. (3. erregela) *<DETE>Lekua berriak<DETE> ezagutu nahi nituen. (4. erregela) *<DETE>Lorategiak loretsuak<DETE> ikustera noa. (5. erregela)
(4)	Erroreari dago(z)kion zuzenketa(k): Min handia sentitzen nuen oinean. Mozkor handiarekin azaldu zen. Oso etxe polita da. Leku berriak ezagutu nahi nituen. Lorategi loretsuak ikustera noa.
(5)	Errorearen iturria: Euskara-ikasleen corpora.
(6)	Oharra(k): -
(7)	Erregela(k): (1. erregela) ADD (&OKER_DETE1A_1) TARGET IZE IF (NOT -1 DET-ZENB-ZG OR DET-NOLGAL OR DET-ZEHAZTUA) (0 IZE + ARR + ABS + NUMS) (NOT 0 AORG) (1 ADJ + ABS + NUMS) (NOT 1 ({POS-BUK})) (2 ADI OR ADT) (NOT 2 IZAN); (2. erregela) ADD (&OKER_DETE1B_1) TARGET IZE IF (NOT -1 DET-ERKIND OR DET-ERKARR OR IOR) (0 ARR + ABS + NUMS) (NOT 0 AORG) (1 ADJ + ARR AND NUMS AND KASUAK) (2 ADI OR ADT) (NOT 2 IZAN); (3. erregela) ADD (&OKER_DETE1C_1) TARGET IZE IF (-1 OSO) (0 IZE + ABS + MUGATUA) (NOT 0 AORG) (1 ADJ AND ABS AND NUMS AND MUGATUA) (2 ADI OR ADT); (4. erregela) ADD (&OKER_DETE1D1_1) TARGET IZE IF (NOT -1 DET-NOLGAL OR DET-ZEHAZTUA OR ETA OR EDO) (0 IZE + ARR + ERG + MUGATUA) (1 ADJ + ABS + NUMP) (2 ADI OR ADT) (NOT 2 IZAN OR UKAN); (5. erregela) ADD (&OKER_DETE1D2_1) TARGET IZE IF (NOT -1 DET-NOLGAL OR DET-ZEHAZTUA OR ETA OR EDO) (0 IZE + ARR + ABS + MUGATUA) (1 ADJ + ABS + NUMP) (2 ADI OR ADT) (NOT 2 IZAN OR UKAN);

(1)	Errore mota eta kategoria/azpikategoria: 3.3.3.3. DETE: <i>Izena + determinatzaile atzizkia + artikulua zehaztugabea</i> motako sintagma.
(2)	Errorearen deskripzioa: Determinatzaile atzizkidun izen baten eskuinean artikulua zehaztugabe bat badator, sintagma erroreduna da; IS batek ezin ditu hartu bi determinatzaile.
(3)	Errore motaren adibide batzuk: *Euskal Herria <DETE>nazioa bat<DETE> izan dela. (1. erregela) *<DETE>Lokala batzuk<DETE> mihatu dituzte. (2. erregela) *<DETE>Opariak batzuk<DETE> erosi dizkiogu. (3. erregela) *<DETE>Sagardotegia batean<DETE> kriston giro dugu. (4. erregela) *<DETE>Laguna batzuei<DETE> esan diet. (5. erregela) *<DETE>Lankidea batek<DETE> esan dit.(6. erregela)
(4)	Erroreari dago(z)kion zuzenketa(k): Euskal Herria nazio bat izan dela. Lokal batzuk mihatu dituzte. Opari batzuk erosi dizkiogu. Sagardotegi batean kriston giro dugu. Lagun batzuei esan diet. Lankide batek esan dit.
(5)	Errorearen iturria: Euskara-ikasleen corpora eta Zubiri eta Zubiri (1995).
(6)	Oharra(k): -
(7)	Erregela(k): (1. erregela) ADD (&OKER_DETE2A_1) TARGET IZE IF (0 IZE + ARR + ABS + NUMS) (NOT 0 AORG) (1 DET-ZEHAZTUA + NMGS + ABS + MG) (2 ADT OR ADI) (NOT 2 ETORRI); (2. erregela) ADD (&OKER_DETE2B_1) TARGET IZE IF (0 IZE + ARR + ABS + NUMS) (NOT 0 AORG) (1 DET-ZEHAZTUGABEA + NMGP + ABS + NUMP); (3. erregela) ADD (&OKER_DETE2C_1) TARGET IZE IF (0 IZE + ARR + ABS + NUMP) (NOT 0 AORG) (1 DET-ZEHAZTUGABEA + NMGP + ABS + NUMP); (4. erregela) ADD (&OKER_DETE2D1_1) TARGET IZE IF (0 IZE + ARR + ABS + NUMS) (NOT 0 AORG) (1 DET-ZEHAZTUA + NMGS + KASUAK); (5. erregela) ADD (&OKER_DETE2D2_1) TARGET IZE IF (0 IZE + ARR + ABS + NUMS) (NOT 0 AORG) (1 DET-ZEHAZTUGABEA + NMGP + KASUAK); (6. erregela) ADD (&OKER_DETE2D3_1) TARGET IZE IF (0 IZE + ARR + ABS + NUMS) (NOT 0 AORG) (1 DET-ZEHAZTUA + NMGS + ERG);

(1)	Errore mota eta kategoria/azpikategoria: 3.3.3.4. DETE: <i>Izena + determinatzaile atzizkia + adjektiboa + artikulua zehaztugabea</i> motako sintagma.
(2)	Errorearen deskripzioa: Izenaz, adjektiboaz eta artikulua zehaztugabe batez osaturiko IS batean determinatzaile atzizkia eta artikulua zehaztugabea, biak, jarritz gero, sintagma erroreduna da. Horrelako egitura duten sintagmetan artikulua zehaztugabea bakarrik jarri behar da, sintagmaren eskuineko muturrean.
(3)	Errore motaren adibide batzuk: *<DETE>Ametsa polit bat<DETE> egin dut. (1. erregela) *<DETE>Jaia handi batzuk<DETE> egin nituen. (2. erregela) *<DETE>Laguna on bati<DETE> eman diot. (3. erregela) *<DETE>Txoria txiki batzuen<DETE> habia da. (4. erregela)
(4)	Erroreari dago(z)kion zuzenketa(k): Amets polit bat egin dut. Jai handi batzuk egin nituen. Lagun on bati eman diot. Txori txiki batzuen habia da.
(5)	Errorearen iturria: Euskara-ikasleen corpusa.
(6)	Oharra(k): -
(7)	Erregela(k): (1. erregela) ADD (&OKER_DETE3A1_2) TARGET IZE IF (0 IZE + ARR + ABS + NUMS) (NOT 0 AORG) (1 ADJ + ARR + ZERO) (2 DET-ZEHAZTUA + NMGS + ABS); (2. erregela) ADD (&OKER_DETE3A2_2) TARGET IZE IF (0 IZE + ARR + ABS + NUMS) (NOT 0 AORG) (1 ADJ + ARR + ZERO) (2 DET-ZEHAZTUGABEA + NMGP + ABS); (3. erregela) ADD (&OKER_DETE3D1_2) TARGET IZE IF (0 IZE + ARR + ABS + NUMS) (NOT 0 AORG) (1 ADJ + ARR + ZERO) (2 DET-ZEHAZTUA + NMGS + KASUAK); (4. erregela) ADD (&OKER_DETE3D2_2) TARGET IZE IF (0 IZE + ARR + ABS + NUMS) (NOT 0 AORG) (1 ADJ + ARR + ZERO) (2 DET-ZEHAZTUGABEA + NMGP + KASUAK);

(1)	Errore mota eta kategoria/azpikategoria: 3.3.3.5. DETE: <i>Izena + adjektiboa + determinatzaile atzizkia + artikulua</i> zehaztugabea motako sintagma.
(2)	Errorearen deskripzioa: Izenaz, adjektiboz eta artikulua zehaztugabe batez osaturiko izen-sintagmetan determinatzaile atzizkia eta artikulua zehaztugabea, biak, jarritz gero, sintagma erroreduna da. Horrelako egitura duten sintagmetan artikulua zehaztugabea bakarrik jarri behar da, sintagmaren eskuineko muturrean.
(3)	Errore motaren adibide batzuk: *Deitu nion <DETE>berri ona bat<DETE> kontatzeko. (1. erregela) *<DETE>Katu txikiak batzuk<DETE> ikusi genituen. (2. erregela) *<DETE>Lagun ona bati<DETE> eman diot. (3. erregela) *<DETE>Lagun ona batzuei<DETE> eman diet. (4. erregela)
(4)	Erroreari dago(z)kion zuzenketa(k): Deitu nion berri on bat kontatzeko. Katu txiki batzuk ikusi genituen. Lagun on bati eman diot. Lagun on batzuei eman diet.
(5)	Errorearen iturria: Euskara-ikasleen corpusa.
(6)	Oharra(k): -
(7)	Erregela(k): (1. erregela) ADD (&OKER_DETE3B1_2) TARGET IZE IF (0 IZE + ARR + ZERO) (1 ADJ + ABS + NUMS + MUGATUA) (2 DET-ZEHAZTUA + NMGS + ABS); (2. erregela) ADD (&OKER_DETE3B2_2) TARGET IZE IF (0 IZE + ARR + ZERO) (1 ADJ + ABS + MUGATUA) (2 DET-ZEHAZTUGABEA + NMGP + ABS); (3. erregela) ADD (&OKER_DETE3E1_2) TARGET IZE IF (0 IZE + ARR + ZERO) (1 ADJ + ARR + ABS + NUMS) (2 DET-ZEHAZTUA + NMGS + KASUAK); (4. erregela) ADD (&OKER_DETE3E2_2) TARGET IZE IF (0 IZE + ARR + ZERO) (1 ADJ + ARR + ABS + NUMS) (2 DET-ZEHAZTUGABEA + NMGP + KASUAK);

(1)	Errore mota eta kategoria/azpikategoria: 3.3.3.6. DETE: <i>Izena + determinatzaile atzizkia + adjektiboa + determinatzaile atzizkia + artikulua zehaztugabea</i> motako sintagma.
(2)	Errorearen deskripzioa: Izenaz, adjektiboz eta artikulua zehaztugabe batez osaturiko izen-sintagmetan determinatzaile atzizkia eta artikulua zehaztugabea, biak, jarritz gero, sintagma erroreduna da. Horrelako egitura duten sintagmetan artikulua zehaztugabea bakarrik jarri behar da, sintagmaren eskuineko muturrean.
(3)	Errore motaren adibide batzuk: *<DETE>Afaria ederra bat<DETE> prestatu zigun. (1. erregela) *<DETE>Etxea ederra batzuk<DETE> eraikiko dituzte. (2. erregela) *<DETE>Ametsa politak batzuk<DETE> egin ditut. (3. erregela) *<DETE>Laguna ona bati<DETE> eman diot. (4. erregela) *<DETE>Txoria txikia batzuen<DETE> habia da. (5. erregela)
(4)	Erroreari dago(z)kion zuzenketa(k): Afari eder bat prestatu zigun. Etxe eder batzuk eraikiko dituzte. Amets polit batzuk egin ditut. Lagun on bati eman diot. Txori txiki batzuen habia da.
(5)	Errorearen iturria: Euskara-ikasleen corpusa.
(6)	Oharra(k): -
(7)	Erregela(k): (1. erregela) ADD (&OKER_DETE3C1_2) TARGET IZE IF (0 IZE + ARR + ABS + NUMS) (1 ADJ + ARR + ABS + NUMS + MUGATUA) (2 DET-ZEHAZTUA + NMGS + ABS); (2. erregela) ADD (&OKER_DETE3C2_2) TARGET IZE IF (0 IZE + ARR + ABS + NUMS) (1 ADJ + ABS + NUMS) (2 DET-ZEHAZTUGABEA + NMGP + ABS); (3. erregela) ADD (&OKER_DETE3C3_2) TARGET IZE IF (0 IZE + ARR + ABS + NUMS) (1 ADJ + ABS + NUMP) (2 DET-ZEHAZTUGABEA + NMGP + ABS); (4. erregela) ADD (&OKER_DETE3F1_2) TARGET IZE IF (0 IZE + ARR + ABS + NUMS) (1 ADJ + ARR + ABS + NUMS) (2 DET-ZEHAZTUA + NMGS + KASUAK); (5. erregela) ADD (&OKER_DETE3F2_2) TARGET IZE IF (0 IZE + ARR + ABS + NUMS) (1 ADJ + ARR + ABS + NUMS) (2 DET-ZEHAZTUGABEA + NMGP + KASUAK);

[illegible]

[illegible]

(1)	Errore mota eta kategoria/azpikategoria: 3.3.3.10. DETE: <i>Izena + determinatzaile atzizkia + adjektiboa + determinatzaile atzizkia + erakuslea</i> motako sintagma.
(2)	Errorearen deskripzioa: Determinatzaile atzizkidun izenaz, determinatzaile atzizkidun adjektiboaz eta erakusle batez osaturiko sintagma erroreduna da; IS batek ezin du hartu determinatzaile bat baino gehiago.
(3)	Errore motaren adibide batzuk: *<DETE>Parkea erraldoia hura<DETE> gustatu zitzaidan. (1. erregela) *<DETE>Kontraesana garbia honek<DETE> nahasten nau. (2. erregela) *<DETE>Laguna ona hauek<DETE> etorriko dira. (3. eta 4. erregelak) *<DETE>Lekua polita horretara<DETE> joango gara. (5. erregela)
(4)	Erroreari dago(z)kion zuzenketa(k): Parke erraldoi hura gustatu zitzaidan. Kontraesan garbi honek nahasten nau. Lagun on hauek etorriko dira. Leku polit horretara joango gara.
(5)	Errorearen iturria: Euskara-ikasleen corpusa.
(6)	Oharra(k): -
(7)	Erregela(k): (1. erregela) ADD (&OKER_DETE4E1_2) TARGET IZE IF (0 IZE + ARR + ABS + NUMS) (NOT 0 AORG OR GEL) (1 ADJ + ABS + NUMS) (2 DET-ERKARR + ABS + NUMS); (2. erregela) ADD (&OKER_DETE4E2_2) TARGET IZE IF (0 IZE + ARR + ABS + NUMS) (NOT 0 AORG OR GEL) (1 ADJ + ABS + NUMS) (2 DET-ERKARR + ERG + NUMS); (3. erregela) ADD (&OKER_DETE4E3_2) TARGET IZE IF (0 IZE + ARR + ABS + NUMS) (NOT 0 AORG OR GEL) (1 ADJ + ABS + NUMS) (2 DET-ERKARR + ABS + NUMP); (4. erregela) ADD (&OKER_DETE4E4_2) TARGET IZE IF (0 IZE + ARR + ABS + NUMS) (NOT 0 AORG OR GEL) (1 ADJ + ABS + NUMS) (2 DET-ERKARR + ERG + NUMP); (5. erregela) ADD (&OKER_DETE4H_2) TARGET IZE IF (0 IZE + ARR + ABS + NUMS) (1 ADJ + ABS + NUMS) (2 DET-ERKARR + KASUAK);

(1)	Errore mota eta kategoria/azpikategoria: 3.3.3.11. DETE: <i>Izena + determinatzaile atzizkia + zenbatzaile orokorra + determinatzaile atzizkia</i> motako sintagma.
(2)	Errorearen deskripzioa: Izen batez eta zenbatzaile orokor batez osaturiko sintagmetan zenbatzaile orokorrak bakarrik darama determinatzaile atzizkia; izenak ere determinatzailea badarama, sintagma erroreduna da.
(3)	Errore motaren adibide batzuk: *Eta <DETE>gerrak guztiak<DETE> bukatu dira. (1. erregela) *Zuri <DETE>idazlana guztia<DETE> zuzendu dizu. (2. erregela)
(4)	Erroreari dago(z)kion zuzenketa(k): Eta gerra guztiak bukatu dira. Zuri idazlan guztia zuzendu dizu.
(5)	Errorearen iturria: Euskara-ikasleen corpora eta Zubiri eta Zubiri (1995).
(6)	Oharra(k): -
(7)	Erregela(k): (1. erregela) ADD (&OKER_DETE6A1_1) TARGET DET IF (0 DET-ZEHAZTUGABEA + ZERO) (1 IZE + ARR + ABS + NUMP) (NOT 1 AORG OR ABL OR ALA OR DAT OR DES OR DESK OR GEL OR GEN OR INE OR INS OR MOT OR SOZ) (2 DET-ORO + ABS AND NUMP AND KASUAK); (2. erregela) ADD (&OKER_DETE6A2_1) TARGET DET IF (0 DET-ZEHAZTUGABEA AND ZERO) (1 IZE AND ARR AND ABS AND NUMP) (NOT 1 AORG OR ABL OR ALA OR DAT OR DES OR DESK OR GEL OR GEN OR INE OR INS OR MOT OR SOZ) (NOT 2 PH) (3 DET-ORO AND ABS AND NUMP AND ABL OR ALA OR DAT OR DES OR DESK OR GEL OR GEN OR INE OR INS OR MOT OR SOZ);

(1)	Errore mota eta kategoria/azpikategoria: 3.3.3.12. DETE: <i>Izena + determinatzaile atzizkia + zenbatzaile zehaztugabea</i> motako sintagma.
(2)	Errorearen deskripzioa: Izenaz eta zenbatzaile zehaztugabe batez osaturiko sintagmetan izena determinatzaile atzizkiduna bada, sintagma erroreduna da.
(3)	Errore motaren adibide batzuk: *Orain dela <DETE>urtea asko<DETE> Irlandara joan nahi nuen.
(4)	Erroreari dago(z)kion zuzenketa(k): Orain dela urte asko Irlandara joan nahi nuen.
(5)	Errorearen iturria: Euskara-ikasleen corpusa eta Zubiri eta Zubiri (1995).
(6)	Oharra(k): Batzuetan anbiguoak izan daitezke honelako adibideak: " <i>lana asko egin dut aurten</i> " / " <i>lan asko egin dut aurten</i> "; eta testuingururik gabe ezin esan lehenengoa erroreduna denik. Bestalde, izena singularrean bakarrik hartzen dugu kontuan, pluralean anbiguotasuna are handiagoa izan daitekeelako.
(7)	Erregela(k): (1. erregela) ADD (&OKER_DETE7A_1) TARGET IZE IF (0 IZE + ARR + ABS + NUMS) (NOT 0 AORG) (1 DET-ZEHAZTUGABEA + ABS + MG); (2. erregela) ADD (&OKER_DETE7B_1) TARGET IZE IF (0 IZE + ARR + ABS + NUMS) (NOT 0 AORG) (1 ASKO + ADB + GRA);

(1)	Errore mota eta kategoria/azpikategoria: 3.3.3.13. DETE: <i>Zenbatzaile zehaztugabea + izena + determinatzaile atzizkia</i> motako sintagma.
(2)	Errorearen deskripzioa: Izen-sintagma batean zenbatzaile zehaztugabe baten ondoren datorren izena mugagabea ez bada (singularra edo plurala bada), sintagma erroreduna da.
(3)	Errore motaren adibide batzuk: *<DETE>Hainbeste oinazea<DETE> dauka. (1. erregela) *<DETE>Hainbat ezusteak<DETE> izan ditu. (2. erregela)
(4)	Erroreari dago(z)kion zuzenketa(k): Hainbeste oinaze dauka. Hainbat ezuste izan ditu.
(5)	Errorearen iturria: Euskara-ikasleen corpora eta Zubiri eta Zubiri (1995).
(6)	Oharra(k): -
(7)	Erregela(k): (1. erregela) ADD (&OKER_DETE9A_1) TARGET DET IF (0 DET-ZENB-ZG + MG) (NOT 0 ERG OR ABL OR ALA OR DAT OR DES OR DESK OR GEL OR GEN OR INE OR INS OR MOT OR SOZ OR ZE OR ZEIN) (NOT 1 AORG) (1 IZE + ARR + ABS + NUMS) (NOT 2 ADJ); (2. erregela) ADD (&OKER_DETE9B_1) TARGET DET IF (0 DET-ZENB-ZG + MG) (NOT 0 ERG OR ABL OR ALA OR DAT OR DES OR DESK OR GEL OR GEN OR INE OR INS OR MOT OR SOZ OR ZE OR ZEIN) (1 IZE + ARR + ABS + NUMP) (NOT 2 ADJ);

(1)	Errore mota eta kategoria/azpikategoria: 3.3.3.15. DETE: <i>Zenbatzaile zehaztugabea + izena + adjektiboa + determinatzaile atzizkia</i> motako sintagma.
(2)	Errorearen deskripzioa: Zenbatzaile zehaztugabe (<i>hainbat</i> eta <i>zenbait</i>), izen eta adjektibo batez osaturiko sintagman adjektiboa determinatzaile atzizkiduna bada, sintagma erroreduna da.
(3)	Errore motaren adibide batzuk: *<DETE>Zenbait istorio interesgarria<DETE> kontatu dizkigu. *<DETE>Hainbat mendi altua<DETE> igo ditugu.
(4)	Erroreari dago(z)kion zuzenketa(k): Zenbait istorio interesgarri kontatu dizkigu. Hainbat mendi altu igo ditugu.
(5)	Errorearen iturria: Euskara-ikasleen corpora.
(6)	Oharra(k): Mementoz <i>hainbat</i> eta <i>zenbait</i> zenbatzaile zehaztugabeak bakarrik hartu ditugu kontuan, alarma faltsu ugari izateko arriskua ekiditearren.
(7)	Erregela(k): ADD (&OKER_DETE9C_2) TARGET DET IF (0 DET-ZENB-ZG + MG) (NOT 0 ABL OR ALA OR DAT OR DES OR DESK OR GEL OR GEN OR INE OR INS OR MOT OR SOZ) (1 IZE + ARR + ZERO) (2 ADJ + ARR + ABS + NUMS) (NOT 2 (@ADLG));

(1)	Errore mota eta kategoria/azpikategoria: 3.3.3.16. DETE: <i>Zenbatzaile zehaztugabea + izena + determinatzaile atzizkia + adjektiboa + determinatzaile atzizkia</i> motako sintagma.
(2)	Errorearen deskripzioa: Zenbatzaile zehaztugabe (<i>hainbat</i> eta <i>zenbait</i>), izen eta adjektibo batez osaturiko sintagman izena eta adjektiboa determinatzaile atzizkidunak badira, sintagma erroreduna da.
(3)	Errore motaren adibide batzuk: *<DETE>Zenbait istorioa interesgarria<DETE> kontatu dizkigu. *<DETE>Hainbat mendia altua<DETE> igo ditugu.
(4)	Erroreari dago(z)kion zuzenketa(k): Zenbait istorio interesgarri kontatu dizkigu. Hainbat mendi altu igo ditugu.
(5)	Errorearen iturria: Euskara-ikasleen corpora eta Zubiri eta Zubiri (1995).
(6)	Oharra(k): -
(7)	Erregela(k): ADD (&OKER_DETE9D_2) TARGET DET IF (0 DET-ZENB-ZG + MG) (NOT 0 ABL OR ALA OR DAT OR DES OR DESK OR GEL OR GEN OR INE OR INS OR MOT OR SOZ) (1 IZE AND ARR AND ABS AND MUGATUA) (2 ADJ + ABS + NUMS);

(1)	Errore mota eta kategoria/azpikategoria: 3.3.3.17. DETE: <i>Determinatzaile zehaztugabea, nolakotzailea edo galdetzailea + izena + determinatzaile atzizkia</i> motako sintagma.
(2)	Errorearen deskripzioa: Izen-sintagma batean determinatzaile nolakotzaile, galdetzaile edota zehaztugabe baten ondoren datorren izena determinatzaile atzizkiduna bada, sintagma erroreduna da.
(3)	Errore motaren adibide batzuk: *Vignemal <DETE>edozein mendia<DETE> baino politagoa da niretzat. *Ez dakit <DETE>zein liburua<DETE> irakurri.
(4)	Erroreari dago(z)kion zuzenketa(k): Vignemal edozein mendi baino politagoa da niretzat. Ez dakit zein liburu irakurri.
(5)	Errorearen iturria: Euskara-ikasleen corpora eta Zubiri eta Zubiri (1995).
(6)	Oharra(k): -
(7)	Erregela(k): ADD (&OKER_DETE10A_1) TARGET DET IF (0 DET-NOLARR + MG + ZERO) (NOT 0 ERG OR ABL OR ALA OR DAT OR DES OR DESK OR GEL OR GEN OR INE OR INS OR MOT OR SOZ OR ZE OR ZEIN) (1 IZE + ARR + ABS + NUMS + MUGATUA) (NOT 1 AORG) (NOT 2 ADJ);

(1)	Errore mota eta kategoria/azpikategoria: 3.3.3.18. DETE: <i>Determinatzaile zehaztugabea, nolakotzailea edo galdetzailea + izena + determinatzaile atzizkia + adjektiboa + determinatzaile atzizkia</i> motako sintagma.
(2)	Errorearen deskripzioa: Izen-sintagma batean determinatzaile nolakotzaile, galdetzaile edo zenbatzaile zehaztugabe baten ondoren datorren izena determinatzaile atzizkiduna bada, sintagma erroreduna da.
(3)	Errore motaren adibide batzuk: *<DETE>Zein plana polita<DETE> egin genuen. *<DETE>Zein ipuinak interesgarriak<DETE> kontatu zizkigun. *<DETE>Zein lagunak atseginak<DETE> ditugun.
(4)	Erroreari dago(z)kion zuzenketa(k): Zein plan polita egin genuen. Zein ipuin interesgarriak kontatu zizkigun. Zein lagun atseginak ditugun.
(5)	Errorearen iturria: Euskara-ikasleen corpora eta Zubiri eta Zubiri (1995).
(6)	Oharra(k): -
(7)	Erregela(k): ADD (&OKER_DETE10B_2) TARGET DET IF (0 DET-NOL + MG) (NOT 1 AORG) (1 IZE + ARR + MUGATUA) (2 ADJ + ABS + MUGATUA);

(1)	Errorre mota eta kategoria/azpikategoria: 3.3.3.19. DETE: <i>Zenbakia + izena + determinatzaile atzizkia</i> motako sintagma.
(2)	Errorrearen deskripzioa: Izen-sintagma batean zenbaki baten ondoren datorren izenak determinatzaile atzizki singularra badarama, sintagma erroreduna da.
(3)	Errorre motaren adibide batzuk: *<DETE>Bi liburua<DETE> irakurri ditut. *<DETE>Bost mendia<DETE> igo genituen.
(4)	Errorreak dago(z)kion zuzenketa(k): Bi liburu irakurri ditut. Bost mendi igo genituen.
(5)	Errorrearen iturria: Euskara-ikasleen corpora eta Zubiri eta Zubiri (1995).
(6)	Oharra(k): -
(7)	Erregela(k): ADD (&OKER_DETE11_1) TARGET DET IF (0 DET-ZEHASTUA + NMGP) (NOT 0 ERG OR ORD OR GEN OR GEL OR DESK OR DES OR DAT OR BNK OR INE) (NOT 1 AORG OR GEL OR DESK OR ZENBAKI OR IZE_ZENB OR BAK) (1 IZE + ARR + ABS + NUMS) (NOT 1 (SNB)) (2 ADI OR ADT);

(1)	Errore mota eta kategoria/azpikategoria: 3.3.3.20. DETE: <i>Zenbakia + izena + determinatzaile atzizkia + adjektiboa</i> motako sintagma.
(2)	Errorearen deskripzioa: Zenbaki, izen eta adjektiboz osaturiko sintagma batean izenak determinatzaile atzizki singularra badarama, sintagma erroreduna da.
(3)	Errore motaren adibide batzuk: *<DETE>4 herria desberdin<DETE> pasatu genuen. *Aurkitu nituen <DETE>bi txakurra polit<DETE>.
(4)	Erroreari dago(z)kion zuzenketa(k): 4 herri desberdin pasatu genuen. Aurkitu nituen bi txakur polit.
(5)	Errorearen iturria: Euskara-ikasleen corpora eta Zubiri eta Zubiri (1995).
(6)	Oharra(k): -
(7)	Erregela(k): ADD (&OKER_DETE12A_2) TARGET DET IF (0 DET-ZEHAZTUA + NMGP) (NOT 0 ORD OR GEL OR DESK OR INE) (1 IZE + ARR + ABS + NUMS) (NOT 1 GEL OR DESK) (2 ADJ + ABS + MG) (NOT 2 ERG OR GRA);

(1)	Errorre mota eta kategoria/azpikategoria: 3.3.3.22. DETE: Zenbakia + izena + determinatzaile atzizkia + adjektiboa + <i>determinatzaile atzizkia</i> motako sintagma.
(2)	Errorerearen deskripzioa: Zenbaki, izen eta adjektibo osaturiko sintagma batean izenak eta izenaren eskuinetara datorren adjektiboak determinatzaile atzizki singularra edota plurala badaramate, sintagma erroreduna da.
(3)	Errorre motaren adibide batzuk: *<DETE>4 herria desberdina<DETE> pasatu genuen. (1. erregela) *<DETE>4 herria desberdinak<DETE> pasatu genuen. (2. erregela) *<DETE>4 herriak desberdinak<DETE> pasatu genuen. (3. erregela)
(4)	Errorreari dago(z)kion zuzenketa(k): 4 herri desberdin pasatu genuen.
(5)	Errorrearen iturria: Euskara-ikasleen corpusa.
(6)	Oharra(k): -
(7)	Erregela(k): (1. erregela) ADD (&OKER_DETE12C_2) TARGET DET IF (0 DET-ZEHAZTUA + NMGP) (NOT 0 ORD OR GEL OR DESK OR INE) (1 IZE + ARR + ABS + NUMS) (NOT 1 GEL OR DESK) (2 ADJ + ABS + NUMS) (NOT 2 ERG OR GRA); (2. erregela) ADD (&OKER_DETE12D_2) TARGET DET IF (0 DET-ZEHAZTUA + NMGP) (NOT 0 ORD OR GEL OR DESK OR INE) (1 IZE + ARR + ABS + NUMS) (NOT 1 GEL OR DESK) (2 ADJ + ABS + NUMP) (NOT 2 ERG OR GRA); (3. erregela) ADD (&OKER_DETE12E_2) TARGET DET IF (0 DET-ZEHAZTUA + NMGP) (NOT 0 ORD OR GEL OR DESK OR INE) (1 IZE + ARR + ABS + NUMP) (NOT 1 GEL OR DESK) (2 ADJ + ABS + NUMP) (NOT 2 ERG OR GRA) (NOT 3 IZAN);

(1)	Errore mota eta kategoria/azpikategoria: 3.3.4.1. DETO: Zenbatzaile zehaztugabea + izena motako sintagma.
(2)	Errorearen deskripzioa: Asko, gutxi edo gehiago zenbatzaile zehaztugabeek eta izen batez osaturiko sintagma batean zenbatzailea izenaren aurretik badator, sintagma erroreduna da.
(3)	Errore motaren adibide batzuk: *<DETO>Gutxi pertsonak<DETO> ginen, baino oso polita zen. *<DETO>Asko perretxiko<DETO> jan ditugu. *<DETO>Gehiago liburu<DETO> irakurri behar ditut.
(4)	Erroreari dago(z)kion zuzenketa(k): Pertsona gutxi ginen, baino oso polita zen. Perretxiko asko jan ditugu. Liburu gehiago irakurri behar ditut.
(5)	Errorearen iturria: Euskara-ikasleen corpusa.
(6)	Oharra(k): -
(7)	Erregela(k): ADD (&OKER_DETO1_1) TARGET ASKO OR GUTXI OR GEHIAGO IF (NOT -2 EZ) (0 ASKO OR GUTXI OR GEHIAGO) (NOT 0 %SIB OR ABL OR ABU OR ALA OR BAN OR DAT OR DES OR DESK OR ERG OR GEL OR GEN OR INE OR INS OR MOT OR PAR) (1 IZE + ARR + ABS) (NOT 1 NUMP OR FALTA OR ("amets") OR ("aurka") OR ("aurre") OR ("balio") OR ("barre") OR ("bat") OR ("berba") OR ("bultza") OR ("dei") OR ("ele") OR ("eraso") OR ("hitz") OR ("hots") OR ("hozka") OR ("huts") OR ("indar") OR ("irain") OR ("jaramon") OR ("jolas") OR ("kaka") OR ("kalte") OR ("kasu") OR ("keinu") OR ("kontra") OR ("korrika") OR ("lan") OR ("lasterka") OR ("lo") OR ("negar") OR ("oihu") OR ("on") OR ("otoi") OR ("otoitz") OR ("parte") OR ("pixa") OR ("porrot") OR ("pot") OR ("poto") OR ("putz") OR ("topa") OR ("topo") OR ("uko") OR ("txiza") OR ("zaunka") OR ("zin"));

(1)	Errore mota eta kategoria/azpikategoria: 3.3.5. DETMUG: Zenbatzaile zehaztugabea edo zenbakia + izena + determinatzaile atzizkia motako sintagma.
(2)	Errorearen deskripzioa: Zenbatzaile zehaztugabe baten ondoren datorren deklinatutako izena mugagabea ez bada, sintagma erroreduna da.
(3)	Errore motaren adibide batzuk: *<DETMUG>Zenbait lagunarentzat<DETMUG> erosi dut. (1. erregela) *<DETMUG>Edozein lagunarentzat<DETMUG> erosi dut. (2. erregela) *<DETMUG>Bost lagunarekin<DETMUG> etorri da. (3. erregela)
(4)	Erroreari dago(z)kion zuzenketa(k): Zenbait lagunentzat erosi dut. Edozein lagunentzat erosi dut. Bost lagunekin etorri da.
(5)	Errorearen iturria: Euskara-ikasleen corpusa.
(6)	Oharra(k): -
(7)	Erregela(k): (1. erregela) ADD (&OKER_DETMUG1_1) TARGET DET IF (0 DET-ZENB-ZG + MG) (NOT 0 ERG OR ABL OR ALA OR DAT OR DES OR DESK OR GEL OR GEN OR INE OR INS OR MOT OR SOZ OR ZE OR ZEIN) (1 IZE + ARR + KASUAK + NUMS) (NOT 2 ADJ); (2. erregela) ADD (&OKER_DETMUG2_1) TARGET DET IF (0 DET-NOLARR + MG + ZERO) (NOT 0 ERG OR ABL OR ALA OR DAT OR DES OR DESK OR GEL OR GEN OR INE OR INS OR MOT OR SOZ OR ZE OR ZEIN) (1 IZE + ARR + KASUAK + NUMS) (NOT 2 ADJ); (3. erregela) ADD (&OKER_DETMUG3_1) TARGET DET IF (0 DET-ZEHAZTUA + NMGP) (NOT 0 ERG OR ORD OR GEN OR GEL OR DESK OR DES OR DAT OR BNK OR INE) (NOT 1 AORG OR GEL OR DESK OR ZENBAKI OR IZE_ZENB OR BAK) (1 IZE + ARR + KASUAK + NUMS) (NOT 1 (SNB) OR ({POS-BUK})) (2 ADI OR ADT);