

eman ta zabal zazu



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

EUSKAL FILOLOGIA SAILA

Ezagutza-base lexikala eraikitzeko *Euskal Hiztegiko*

definizioen azterketa sintaktiko-semantikoa

Hitzen arteko erlazio lexiko-semantikoak: definizio-patroiak,
eratorpena eta postposizioak

Mikel Lersundi Ayestaranek

Euskal Filologian Doktore titulua eskuratzeko aurkezturiko

TESIA

Donostia, 2004ko abendua

eman ta zabal zazu



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

EUSKAL FILOLOGIA SAILA

Ezagutza-base lexikala eraikitzeko *Euskal Hiztegiko*

definizioen azterketa sintaktiko-semantikoa

Hitzen arteko erlazio lexiko-semantikoak: definizio-patroiak,
eratorpena eta postposizioak

Mikel Lersundi Ayestaranek Ibon
Sarasola Errazkinen eta Eneko Agirre
Bengoaren zuzendaritzapean egindako
tesia, Euskal Herriko Unibertsitatean
Euskal Filologian Doktore titulua
eskuratzeko aurkeztua.

Donostia, 2004ko abendua

Aita, bukatu det.

Lastima zuk ezingo dezula irakufi. Badakit tesi onek irakurle bat galdu duela, baña aurkezpen-egunean izango zera, ezta? Zenbatetan galdetu ote didazu *baña zeatz-meatz ze itten ai zea?* Ona egiten aritu naizenaren emaitza.

Azkenean alde batera geratu da *Jautarkolen sorta marduldua* liburuan artutako bidea. Baita *Pasaiako Hizkera* lanekoa ere. Zenbait izkera bereziei buruzko informaziorik ez dezu emen topatuko. Alde batera utzi ditut *esparagiñak, txiringak, puñilak, nikitzak, berebilak, urutizkiñak, molapuntak*,... ain maite dituzun itzak, eta konputagañuen arloan euskerak dituen beaiei efepeatu, eta oyei eldu diyet. *Maite* kontutan ari gerala... ain *maite* ez dezun ikuña jafi det lan onetan: *ha*. Amaika aldiz ofi buruz itz egin izan degu, eta ziur orain ere itz egingo genukela, baña distantziya-kontuak direla-eta solasaldi ori beste baterako utzi bearko degu.

Gogoan det Ixa Taldean asi nintzen eguna nire lenengo maratoyaren ondorengo izan zela. Garai artan lagun-talde ttikiya zena gaur egun talde aundiya biurtu da. Asiera artan makiña-kontuak zirela-eta, itsumutíl —eta *itsuneska*— ainbat lankide izan nituen. Izkuntzalari-taldeak ere makiña kontu erakutsi dizkit. Taldeari esker jayo da txosten au.

Estatu Batuetan ere, lagun miñ píloa egin nuen. Gogoan al dituzu? Laguntza edefa izan nuen txiringa-ibilbideetan, bakartasun-egoeretan... Edefa kuadríla!

Telebisiyuan egindako lanak ere izugafi lagundu nau une latzetan. An bizitako esperientziyak, lafitasunak,... indar biurtu zaizkit tesi-garapenean. Bertan egindako lagunak beti urbil izan ditut.

Gogoratzen al dituzu *Euskal Filologia* ikasi nueneko irakasleak eta ikaskideak. Ziur gogoan dituzula Lisboan giñenean sortutako *sesei* ikasle ayen kontakizunak. Bai une polittak, ezta?

Dakizunez, tesi-lan onetan bi zuzendari izan ditut eta aiei zor diet sortutako fruitua. Ibon ezaguna duzu, eta bestea ezagutzera iritsi ez zeran Eneko da. Azken onekin eztabaida-une eta une on ugari konpartitu ditut, eta tesi-lan zatitxo aundiya ari zor diyet. Ibon'ek, Eneko'k, Jose Mari'k, Izaskun'ek eta Xabier'ek leen ziribofoa irakufi didate eta ayen iradokizun jakintsuak azken emaitza onetan islatzen sayatu naiz.

Zer moduz dabil inguruko jendea? Azken lurbira aski gogofa egin zaigu. Lagun eta familiarteko asko joan zarete epe laburfean: Mañu, Eneko, Poli, Juanjoxe, Manueltxo, Josu, zu zeu... Buf, gañera zuetako bostek egun batetik bestera alde egin duzue. Eskumiñak guziei. Zuen falta izugafi nabaritu degu, baña, zorionez, lagun asko izan ditugu inguruan urte gogofa goxatzeko.

Ama ondo dago. Etxean utsune aundiya nabari du, baña bizitza-legea dela esaten du. Aren indafak ni ere indartu nau lan onen azken txanpan. Familiya osoa lagun izan det; lagunak ere ez dira urun ibili. Ezin aztuko ditut *adiskidetasunez* betetako ostiralak. Tristurak baztertzeko gai izan diren ostiral bereziyak.

Eta etxekoei buruz, zer esango dizut ba. Dakizunez, aur-zain gaude, eta tesi-maratoi onetan Amaia izan det uneoro laguntzaie, animatzaie eta indargafi.

Ziur jun zeraten guziok ere, txosten au bukatzen lagundu nauzuela.

Besterik ez, aita. Iruon partez muxu aundi bat eta befiz ikusi arte.

Efenteya'n, 2004'ko neguila'ren 13'a.

Lan hau egiteko, Eusko Jaurlaritzaren diru-laguntza baliatu dut.

Aurkibidea

I.	Sarrera	19
I.1.	Lexikoiak	20
I.2.	Ezagutza-base lexikalak, hiztegi ezagutza-baseak eta ontologiak	23
I.3.	Euskararen Ezagutza-Basea eta <i>Euskal Hiztegia</i>	26
I.4.	EHren prestaketa	29
I.5.	Erlazio lexiko-semantikoen erauzketa	30
I.6.	Azterketa eta deskribapen linguistikoak	33
I.7.	Tesi-lanaren egitura	34
I.8.	Tesiarekin lotutako publikazioak	35
II.	Aurrekariak	39
II.1.	Lexikoa sortzeko hurbilpenak eta iturriak	39
II.2.	Eratorpena	45
II.3.	Patroiak eta postposizioak	48
III.	Hiztegiaren aurreprozesua, Murriztapen Gramatikak eta azterketa estatistikoa ...	53
III.1.	Sarrera	53
III.2.	EHren egitura orokorra	55
III.3.	TEI ekimena	57
III.4.	Hiztegiaren bertsio tabulatua	59
III.4.1.	Beharrezko zaizkigun eremuak	59
III.4.2.	Fitxategien egokitzapena	63
III.5.	Definizioen analisi morfologikoa	64
III.6.	Sarrera-hitzen analisia	67
III.7.	Erlazioak erauzteko tresna: <i>Murriztapen Gramatika</i>	70
III.8.	Hiztegiaren azterketa estatistikoa	72
III.9.	Ondorioak	76
IV.	Definizioen guneak erauzten: sinonimoak, genusak, eta erlatore bereziak	81
IV.1.	Sarrera	81
IV.2.	Sinonimoak	86
IV.2.1.	Definizioen hezurdura	86
IV.2.2.	Semantika	89
IV.3.	Genusak	90
IV.3.1.	Definizioen hezurdura	90
IV.3.2.	Semantika	93
IV.3.3.	Genus orokorrak	95
IV.4.	Erlatore bereziak	96
IV.4.1.	Ia-Sinonimia	98
IV.4.2.	Antonimia	99
IV.4.3.	Hiperonimia / hiponimia	100
IV.4.4.	Ezaugarria	101
IV.4.5.	Jasale	102
IV.4.6.	Berezko ekintza	102
IV.4.7.	Meronimia	103

IV.4.8.	Berezko objektua	105
IV.4.9.	Gabezia, eza	106
IV.4.10.	Betekizun / betegai	106
IV.4.11.	Aditzen genus orokorrak	106
IV.4.12.	Elipsia	108
IV.4.13.	Erlatoreak adierazten duena	109
IV.5.	Izen eta adjektiboak	110
IV.6.	Azpisarrerak	111
IV.7.	Metodologia eta ebaluazioa	113
IV.7.1.	Izenak	114
IV.7.2.	Aditzak	115
IV.7.3.	Adjektiboak	117
IV.7.4.	Izenak eta adjektiboak	119
IV.7.5.	Azpisarrerak	121
IV.7.6.	Laburpena	123
IV.8.	Ondorioak eta etorkizuneko lanak	123
V.	Eratorpena	129
V.1.	Sarrera	129
V.2.	Morfologia	131
V.2.1.	Eratorpenaren eta flexioaren arteko ezberdintasunak	132
V.3.	Atzizki lexikalak	134
V.3.1.	Alfabeto-hurrenkera (Villasante, 1974)	136
V.3.2.	Hitz berrien kategoria eta atzizkien esanahia (Azkue 1923-1925; Lafitte, 1944)	136
V.3.3.	Oinarriaren kategoria (Ithurri, 1979)	137
V.3.4.	Lanen ondorioak	137
V.4.	Eratorpena <i>Euskal Hiztegia</i> n	137
V.5.	Euskarazko eratorpen-atzizkiak EHn	139
V.5.1.	Bi elementuren arteko erlazioak	141
V.5.2.	Hiru elementuren arteko erlazioak	144
V.6.	Atzizkien interpretazioak nola zehaztu definizioak erabiliz	144
V.7.	Emaitza orokorrak	148
V.7.1.	Estaldura	149
V.7.2.	Doitasuna	152
V.8.	Erlazioak EuroWordNeten txertatzea	153
V.8.1.	Lortutako erlazioak WordNet-en txertatzea	153
V.8.2.	Adieren desanbiguazioa	154
V.9.	Etorkizuneko lana	157
V.9.1.	Atzizki lexikal gehiago lantzea	157
V.9.2.	Oinarriaren kategoria kontuan hartzea	158
V.9.3.	Sinonimiaren hedadura	159
V.9.4.	Erlazioak EuskalWordNet-era lotzea	160
V.10.	Ondorioak eta etorkizuneko lana	161
VI.	Patroi lexikalak eta postposizioak	167
VI.1.	Sarrera	167
VI.2.	Patroi lexikalak EHn	171
VI.3.	Postposizioen interpretazioak zehazteko oinarriak	173
VI.4.	Interpretazioen zerrenda zehazteko metodoa	177
VI.4.1.	Metodoa <i>instrumentalari</i> aplikatuta	179

VI.5. Postposizio-atzizkiak eta interpretazioak	186
VI.6. Postposizio konposatuak	186
VI.6.1. Anbiguoak	187
VI.6.2. Ez-anbiguoak	188
VI.7. EHko definizioetan dauden postposizioen desanbiguazioa	189
VI.7.1. EHko definizioetan postposizioak bilatu	189
VI.7.2. Itzulpenak egin	190
VI.7.3. Definizioak lotu	191
VI.7.4. Desanbiguazioa	192
VI.7.5. Emaitzak	194
VI.8. Ondorioak eta etorkizuneko lana	196
VII. Ondorioak eta etorkizuneko lanak	201
VII.1. Oinarriak eta prestakizunak	201
VII.2. Analizatzaile morfologikoa	202
VII.3. Hitzen arteko erlazioak eta gramatika	202
VII.3.1. Sarrera-hitza eta definizioko hitz baten arteko erlazioak	202
VII.3.2. Sarrera-hitzaren barruko erlazioak	203
VII.3.3. Definizio barruko hitzen arteko erlazioak	205
VII.4. Emaiza orokorrak	206
VII.5. Etorkizuneko lanak	207

Eranskinak

A. Erlatore bereziak	239
B. Eratorpena	269
C. Patroi lexikalak	347
D. Postposizio-atzizkiak eta interpretazioak	353
E. Postposizio konposatuak	363
F. Ingelesezkako preposizioak	373
G. Espainierazko preposizioak	385
H. Euskarazko postposizio konposatuak	395
I. Gramatikak	409

Taulen aurkibidea

1. taula.	adar sarrera-hitzaren kategoria, adiera-zenbakia, definizioa eta adibideak.....	22
2. taula.	Tesi-txosten honekin zerikusia duten argitalpenak, dagozkien kapituluekin lotuta.....	35
3. taula.	Tesi-txosten honekin zerikusia duten argitalpenak, dagozkien kapituluekin lotuta.....	36
4. taula.	Hiztegian ditugun kategorien laburdurak, haien adieraziak eta kopuruak.....	59
5. taula.	Hiztegian agertzen diren kategoria-konbinazioak eta kopurua.....	60
6. taula.	Hiztegiko definizioetan agertzen diren laburdurak eta haien adieraziak.....	67
7. taula.	Hiztegiko sarrerak, adierak eta azpisarrerak: kopuruak.....	72
8. taula.	Hiztegian definizioa duten adierak kategoriaka.....	73
9. taula.	Definizioetan dauden hitz usuenen kopurua.....	73
10. taula.	Hiztegiko definizioetan agertzen diren token-laukote, -hirukote eta -bikote usuenak.....	75
11. taula.	Postposizio-token segida usuenak kontuan hartuta ditugun patroï berriak.....	76
12. taula.	Sinonimoen bitartez emandako definizio-kopurua.....	90
13. taula.	Izenen artean topatutako genus usuenen zerrenda.....	96
14. taula.	Aditzen artean topatutako genus usuenen zerrenda.....	96
15. taula.	Izen eta adjektibo direnen artean topatutako genus usuenen zerrenda.....	96
16. taula.	Bikote-segida usuak.....	97
17. taula.	–i eman izena erlatoarearen laburpen-aula.....	98
18. taula.	Ia-sinonimia erlazioa markatzen duten erlatoareak.....	99
19. taula.	Antonimia erlazioa markatzen duten erlatoareak.....	100
20. taula.	Hiperonimia erlazioa markatzen duten erlatoareak.....	100
21. taula.	Ezaugarri erlazioa markatzen duten erlatoareak.....	102
22. taula.	Jasale erlazioa markatzen duten erlatoareak.....	102
23. taula.	Berezko ekintza erlazioa markatzen duten erlatoareak.....	103
24. taula.	Meronimia / holonimia erlazioa markatzen duten erlatoareak.....	105
25. taula.	Berezko objektu erlazioa markatzen duten erlatoareak.....	106
26. taula.	Gabezia adierazten duen erlatoarea.....	106
27. taula.	Betekizun / betegai erlazioak markatzen dituen erlatoarea.....	106
28. taula.	Hitz anitzeko sinonimo / hitz anitzeko hiperonimo erlazioa markatzen duten erlatoareak.....	108
29. taula.	Elipsidun erlatoareak.....	109
30. taula.	Erlatoareak adierazten duen erlazioa markatzen duten erlatoareak.....	109
31. taula.	Sinonimoen, genusen eta antonimoen informazioa.....	110
32. taula.	Erlatoaren informazioa.....	110
33. taula.	Azpisarreraren sinonimoen eta genusen informazioa.....	111
34. taula.	Erlatoaren informazioa.....	112
35. taula.	Izenen laginetik ateratako emaitzak.....	114
36. taula.	Izenen lagineko definizioei buruzko datuak.....	114
37. taula.	Izenen artean ditugun erlazioak.....	115
38. taula.	Aditzen laginetik ateratako emaitzak.....	116
39. taula.	Aditzen lagineko definizioei buruzko datuak.....	116
40. taula.	Aditzen artean ditugun erlazioak.....	117
41. taula.	Adjektiboen laginetik ateratako emaitzak.....	117
42. taula.	Adjektiboen lagineko definizioei buruzko datuak.....	117
43. taula.	Adjektiboen artean ditugun erlazioak.....	119
44. taula.	Izen eta adjektiboen laginetik ateratako emaitzak.....	119
45. taula.	Izen eta adjektiboen lagineko definizioei buruzko datuak.....	119
46. taula.	Izen eta adjektibo diren sarreraren definizio-patroien arabeko kategoria.....	120

47. taula.	Izen eta adjektibo diren sarrera-hitzetan dauden erlazioak.....	121
48. taula.	Azpisarreraren definizio-motak.....	121
49. taula.	Azpisarreraren laginetik ateratako emaitzak.....	122
50. taula.	Azpisarreraren lagineko definizioei buruzko datuak.....	122
51. taula.	Azpisarreretan markatu ditugun erlazioak.....	122
52. taula.	Kategoria guztietako emaitzen batez-bestekoak.....	123
53. taula.	Batez besteko orokorrak.....	123
54. taula.	Erlazio-kopurua.....	126
55. taula.	Markadun definizio-kopurua.....	126
56. taula.	Eratorpena.....	132
57. taula.	Flexioa.....	132
58. taula.	Flexio- eta eratorpen-elementuen erregularitasun maila.....	133
59. taula.	–tzarekin sortutako hitz eratorrien interpretazioak.....	135
60. taula.	Analizatzaileari gehitu dizkiogun eratorpen-atzizkiak eta konposizio- elementuak.....	140
61. taula.	Bi elementuren arteko erlazioak dituzten atzizki lexikalen ezaugarriak.....	142
62. taula.	Bi elementuren arteko erlazioak dituzten atzizki lexikalen ezaugarriak.....	143
63. taula.	Hiru elementuren arteko erlazioak dituzten atzizki lexikalen ezaugarriak.....	144
64. taula.	–tza atzizkidun hitz eratorrien balizko adierak.....	145
65. taula.	sasitza hitzaren balizko interpretazio guztiak.....	145
66. taula.	lumatza sarrera-hitzaren analisi morfologikoa.....	146
67. taula.	–tzadun eratorriak: eratorriaren eta oinarriaren kategoriaren arabera kopuruak.....	147
68. taula.	–tzadun eratorrien ezagututako interpretazioen kopurua.....	147
69. taula.	–tzadun eratorrien sinonimiaren hedadura.....	147
70. taula.	–tzadun sinonimoen interpretazioak.....	148
71. taula.	Aztertutako atzizki lexikalak eta kopuruak.....	149
72. taula.	Aztertutako atzizki lexikalak eta kopuruak.....	150
73. taula.	Eratorri-kopurua: erlazio-kopurua, sinonimo-kopurua.....	150
74. taula.	–arazi aditz-arazlearekin lortutako emaitzak.....	150
75. taula.	–tu/-duren datu estatistikoak: estaldura, doitasuna.....	152
76. taula.	Eratorpenarekin lortutako emaitzen doitasuna.....	152
77. taula.	Aztertutako laginaren estaldura-kopurua.....	153
78. taula.	Hizkuntzen arteko loturak.....	156
79. taula.	Loturen estaldura 100 adibideekin.....	156
80. taula.	Adieren desanbiguazioaren doitasuna.....	157
81. taula.	Landu gabe utzi ditugun eratorpen-atzizkiak eta konposizio-elementuak.....	158
82. taula.	–tasun atzizki lexikalaren interpretazioak.....	160
83. taula.	Patroiak eta erlazioak.....	172
84. taula.	Patroiak laginetan.....	173
85. taula.	Euskarazko instrumental postposizio-atzizkiaren interpretazio-zerrenda.....	180
86. taula.	Postposizio-atzizki instrumentala eta erdarazko preposizioen arteko lotura interpretazio jakin bati dagokionez.....	181
87. taula.	Instrumentalaren interpretazioak eta LCSen arteko lotura.....	181
88. taula.	Dorren rol tematikoekin egin dugun lotura.....	182
89. taula.	Dorren primitiboekin lotura.....	182
90. taula.	Dorren rolen eta gure interpretazioen arteko lotura.....	183
91. taula.	LCSetatik esportatuko ez ditugun interpretazioak.....	183
92. taula.	LCSetako preposizioak gura pasa ondoren instrumentala eta preposizioen arteko ebakidura.....	184
93. taula.	Postposizio konposatuak landu ondorengo ebakidura berriak.....	184
94. taula.	Interpretazio-zerrenda.....	185
95. taula.	Postposizio-atzizkiak eta interpretazioak.....	186
96. taula.	denbora / lekua anbiguotasuna postposizio konposatuetan.....	187

97. taula.	leku moten arteko anbiguotasuna postposizio konposatuetan.	187
98. taula.	ezaugarriaren inguruko anbiguotasuna postposizio konposatuetan.	188
99. taula.	Posposizio konposatuak eta interpretazioak	188
100. taula.	Instrumentalak eta byk, eta instrumentalak eta conek dituzten interpretazio- ebakidurak.	192
101. taula.	Instrumentala by eta conen interpretazio-ebakidura.	193
102. taula.	Egindako esperimentuan instrumentalari esleitutako interpretazioak eta ingelesarekiko lotura.	193
103. taula.	Egindako esperimentuan instrumentalari esleitutako interpretazioak eta espainierarekiko lotura.	193
104. taula.	Instrumentalaren interpretazioak eta erdarazko baliokideak.	194
105. taula.	Hiztegi bakoitzaren emaitzak eta guztiak batera ateratakoak.	195
106. taula.	Eredu den laginean dauden instrumentalaren interpretazio kopuruak.	195
107. taula.	Hiztegian dauden definizioak eta markatutakoen kopurua.	206
108. taula.	Hiztegiko definizioak dituzten erlazioen arabera.	207
109. taula.	Hiztegiko definizioetan dugun erlazio-kopurua.	207

Laburdurak

EBL: Ezagutza-Base Lexikala

EDBL: Euskararen Datu-Base Lexikala

EEB: Euskararen Ezagutza-Basea

EH: *Euskal Hiztegia*

EHEB: Euskal Hiztegiaren Ezagutza-Basea

ELK: Egitura Lexiko Kontzeptualak

HAUL: Hitz Anitzeko Unitate Lexikal

HEB: Hiztegi Ezagutza-Basea

LCS: Lexical Conceptual Structure

LNP: Lengoaia Naturalaren Prozesamendua

MRD: Machine Readable Dictionary —makinek irakurtzeko moduko hiztegia—

OEH: *Orotariko Euskal Hiztegia*

Aita, askotan esan izan didazu, baina...

erran	H1A0	ad.	Esan.
esan	H1A0	ad.	Hitzen bidez, ahoz nahiz izkribuz, gogoan dena adierazi.
mentura	A0	iz.	Zoria, ustekabean edo arrazoi jakinik gabe gertatzen denaren alegiazko zergatikoa; ustekabeko gertaria.

Erraiten ohi dugu. Benturaz egun hillen gara, benturaz bihar, eztugu oren baten segurantzarik: beraz lehen baiño lehen, behar dugu bide onean iarri, enganatuak gertha ezkaitezin. Hunela erraiten dugu, eta erraitea ongi egiten dugu: ordea obratzea bethi ere geroko utzten dugu: heriotzea zahartzerakotzat egotzten eta ordenatzen dugu. Ezin sinhets dezakegu, adinak berak akhaba arteiño, akhabatuko garela. Erran bai nahi duzun bezanbat: benturaz bihar hillen, benturaz gaur, benturaz ez zahartuko. Ordea zahartuko baikiña bezala geure gauza guztiak egin. Geure nahiarekin batean, geure gogoak hanbat segurantza emaiten deraku, non zahartu hutsezko ponturaiño, ezkarela hillen iduritzen baitzaiku.

(Axular, *Gero*: §42)

I. Sarrera

Lengoaia Naturalaren Prozesamenduan (LNP) aurrera egingo bada, tresna gero eta indartsuagoak beharko dira itzulpen automatikoa egin ahal izateko, testuen laburpenak sortzeko, informazioa berreskuratu ahal izateko, oro har, idatzizko testuak makinek uler ditzaten. Automatizazioa gero eta indartsuagoa den gizarte honetan euskarak bere tokitxoa izan dezan, beste hizkuntzetan sortzen eta garatzen diren oinarri eta baliabide berak sortu beharko dira. Lan horretan ari da azken urteotan Ixa Taldea.¹ Ixa Taldean dagoeneko hizkuntzaren zenbait atal landu dira: ortografia, morfologia, sintaxia eta semantika besteak beste. **Esku artean duzuen tesi-lan honen helburua semantika alorrean kokatzen da, euskarazko hitzen² arteko erlazio lexiko-semantikoak zehaztea baita.**

Erlazio lexiko-semantiko horiek oso beharrezkoak dira aipatutako testuen laburpenak, informazio-berreskurapena eta itzulpen automatikoa egiteko garaian.

Adibide gisa, koka gaitezen itzulpen automatikoaren eremuan. Euskarazko zezenari adarra hautsi zaio perpausa espainierara zuzen itzuli nahi badugu, behar beharrezkoa da zezen eta adarren artean erlazorik ba ote dagoen ezagutzea, bestela makinak perpaus honetan adar horrek duen adiera zein den —cuerno, antena, rama, ramal edo afluente— ez du jakingo, eta, itzulpena egiteko garaian,

¹ <http://ixa.si.ehu.es/Ixa> (2004-11-30ean atzitua).

² Ezer baino lehen, eta *hitz* terminoarekin arazoak sor daitezkeenez, azaldu nahi dugu guretzat *hitzak* kontzeptuak izendatzeko erabiltzen diren elementuak direla. Kontzeptua maila abstraktoan geratzen da, eta *hitza* dugu kontzeptu hori azalean izendatzeko erabiltzen den errepresentazioa (Lewandowski, 1992).

horietako bat ausaz aukeratuko du. Adibideak erakusten duenez, nabaria da hitzen arteko erlazioak ezagutzea beharrezkoa dela LNPren zenbait alorretarako.

Baina hitzen arteko erlazio lexiko-semanticetan sakondu baino lehen, ikus dezagun LNPren erabilgarri diren hitz-zerrendak —lexikoiak— sortzeko garaian nolako metodoak eta zein oinarri erabili ohi diren.

I.1. Lexikoiak

Makinek, testuen laburpenak, informazio-erazketa eta itzulpen automatikoa bezalako lanak modu egokian burutu ditzaten, **lexikoiak** behar dira. Wilks *et al.*-ek (1996) **lexikoa** eta **hiztegia** bereizten dituzte: **lexikoa** konputagailu-programa batekin batera erabiltzeko sarrera-hitz multzo egituratua da; eta, **hiztegia**, berriz, hitz baten adierari buruzko deskribapena eta informazio lexikala dakarren testu inprimatua. Definizio hauen arabera, badirudi lexikoa makinek euren lana burutu dezaten behar duten hitz multzo egituratua dela; eta hiztegia, berriz, pertsonak hitz bati buruzko informazioa lortu ahal izateko testu inprimatua.

Lexikoi hori egituratzeko garaian hizkuntzaren hainbat alorri erreparatu ahal dakiok: morfologia, sintaxia edo semantika. **Gure helburua lexikoa semantikoki antolatzea da.**

Lexikoi hori **eskuratzeko** garaian bi **hurbilpen** nagusi erabili dira: **arauemailea** eta **deskriptiboa**. **Hurbilpen arauemailean**, hasiera-hasieratik marko zurrana zehazten da, eta informazioa marko horretan txertatzen da. **Hurbilpen deskriptiboan**, aldiz, aurretik ez dugu inongo marko zehatzik, eta ezaugarri multzoa osatuko duten elementuak aztertutako datuetan agertutakoak dira.

Hurbilpen hauei oso lotuta dago **gizakiaren eskuhartzea**. Eskuhartzearen arabera metodoa **eskuzkoa**, **automatikoa** edo **erdi-automatikoa** da. Zein metodo erabili erabakitzeko garaian, kontuan izan behar dira, alde batetik, zeintzuk diren erabiliko diren iturriak; eta, bestetik, helburu den aplikaziorako zein informazio zehaztu edo markatu behar den. Hala, eskuzko metodoek hurbilpen arauemailea darabilte, automatikoei deskriptiboa, eta erdi-automatikoei bai arauemailea eta bai deskriptiboa.

Metodo bakoitzak bere abantailak eta oztopoak ditu. Eskuzko garapenak **introspekzioan** oinarritzen dira. Lexikoa osatzeko garaian, gizakiek munduari buruz eta hizkuntzari buruz duten jakinduria erabiliz gero, sortutako datuen zuzentasuna bermatuko dugu. Era honetan garatutako proiektuek duten arazo nagusia jende eta denbora ugari behar izatea da; eta jende ezberdin asko garai ezberdinetan proiektu baten barruan lan egiterakoan, koherentzia arazoak sor daitezke.

Garatutako lan erdi-automatiko eta automatikoei, berriz, **oinarri** gisa egitura gabeko edo egituratutako oinarri lexikalak izan ohi dituzte, hala nola, **corpusak**, **hiztegiak**, eta **bien arteko konbinazioak**.

Corpusak egitura gabeko testutzat hartu ohi dira, baina erabilerari buruzko informazioa, hitzak errealitatean dituzten maiztasunak, egitura sintaktiko zenbaitek dituzten maiztasunak, eta halako informazioa lortzeko oso erabilgarriak dira. Hitzak ez dira isolaturik agertzen, baizik eta testuinguru jakin batean, eta testuinguru horrek berak informazio baliagarria ematen du.

Corpusen garrantzia ezberdina da dituzten marken arabera. Corpora hutsa —esaldi multzoa, alegia— baldin bada, erabilgarria da; baina are erabilgarriagoak dira lematizatutako corpusak, hitzen kategoriak markatuta dituzten corpusak, morfologikoki analizatuta dauden corpusak, egitura sintaktikoak markatuta dituzten corpusak, semantikoki —bai adiera mailan, eta bai hitzen arteko erlazio mailan— markatutako corpusak, eta abar. Beraz, corpus batean gero eta informazio gehiago izan, gero eta ikerkuntza-alor gehiagotan erabili ahalko da.

Corpusak ez dira beti **elebakarrak**, eta corpus **elebidunek** gero eta garrantzia handiagoa dute LNPN. Corpus elebidun batek bi hizkuntza —gutxienez— parekatzeko aukera ematen du. Corpus elebidun hauek **lerrokatuta** baldin badaude, gainera, hizkuntza bateko esaldia beste hizkuntzako esaldi baliokidearekin —edo esaldi baliokideekin— parekatuta dago.

Egituratutako baliabide lexikalen multzoan, berriz, hainbat testu mota daude: makinak irakurtzeko moduko hiztegi (MRD)³ elebakarrak eta elebidunak, tesaueroak eta entziklopediak. Guztiak baliabide lexikal egituratuak dira, baina duten egitura eta informazioa ezberdinak dira.

³ MRD, *Machine Readable dictionary*ren laburdura da. LNPN asko erabiltzen den terminoa dugu.

Hiztegiak testu berezituak dira. Euren gaia hizkuntza da —edo hizkuntzak, hiztegi elebidunen kasuan—. Hitzei buruzko informazioa dakarte, eta baita hitz hauen adierei buruzkoa. Baina hitzei buruzko informazioa izateaz gain, munduari buruzko informazioa ere badute. Hau nabariagoa da hiztegi **elebakarretan** hiztegi **elebidunetan** baino. Hiztegi elebakarrak eta elebidunak iturri aski interesgarriak dira lexikoiak sortzeko garaian, hitzei eta adierei buruzko informazio-iturri egituratuak direlako.

Hiztegietatik hainbat informazio lor daiteke. Aurreprozesu handirik gabe, **sarrera-hitz** guztiak erauziz hitz-zerrenda sor daiteke. Hala, LNPN garrantzi handia duten hitz anitzeko zenbait unitateren erreferentzia izan dezakegu, batik bat azpisarrerak ere kontuan izanez gero.⁴

Sarrera-hitzez gain, **kategoriak** ere badaude. Gainera, sarrera-hitzen eta kategorien arteko lotura dago. Sarrera-hitz horien **adierak** ere badaude. Adiera horiek zeintzuk diren **definizioekin** zehaztu ohi dira, eta adiera bakoitzari dagozkion **adibideak** ere badaude. Informazio hau guztia datu-base batean sartuz gero, dagoeneko informazio aberatsa izango dugu. 1. taulan dugu *Euskal Hiztegi* (EH) adar sarrera-hitzari zuzenean lotuta dagoen informazioa: sarrera-hitza (*Sar.*), kategoria (*Kat.*), adiera-zenbakia (*Adr.*), definizioa eta adibideak.

Sar.	Kat.	Adr.	Definizioa	Adibideak
adar	izen	A1	Zenbait animaliak bekoki aldean dituzten hezurrezko luzakin bakoitza.	Orein adar-zabala. Errinozeroaren sudur adarra. Oreinaren adar abartuak. Ahuntzak adarrak dituen aldera: alderantziz. Adarrak okerrak ditu. Adar bihurriak. Adarrak moztu. Idia adarrerean eta gizona hitzerean.
adar	izen	A2	Adar hustu eta zulatua, musika tresna bezala erabiltzen dena.	Adar hotsa entzunda.
adar	izen	A3	Adarkia.	Adarrezko edontzia. Adar erre usaina.
adar	izen	A4	Zuhaitzaren enborretik ateratzen diren besoetako bakoitza.	Adarretara igo. Erroak utzirik adarrei lotzea.

1. taula. adar sarrera-hitzaren kategoria, adiera-zenbakia, definizioa eta adibideak.

Hala ere, LNPN barruan adiera kontuek egunez egun garrantzi handiagoa dutenez, hitzen arteko loturak zehazteko gero eta lan gehiago egiten dira. Hau da, aurreko

⁴ Hitz anitzeko unitate lexikalen (HAUL) ezagutza gai garrantzitsua da LNPN. Makinek zuriunetik zuriunera arteko unitateak hartu ohi dituzte kontuan, baina HAUL hauekin kontzeptu hori aldatu beharra dago, unitate beraren barruan zuriuneak izango direlako. *adarra jo —iseka* adierarekin— unitatea baldin badugu, ezin analizatuko ditugu *adarra* eta *jo* separatuta. Dena unitate gisa hartu beharko dugu analisi egokia lortzeko asmoz. Halako unitateak hiztegi elebakarretan *azpisarrerak* izan ohi dira. Eta, *Euskal Hiztegi* (EH), *adarra jo* azpisarrera adar sarrera-hitzaren barruan dago.

paragrafoan aipatu bezala, sarrera-hitzak, kategoriak, adierak, definizioak eta adibideak lotu ditugu, baina datu-basean dauden sarrera-hitzen artean ez dugu inolako loturarik, eta ezta definizio barruan dauden hitzen artean ere.

Hasierako adibidearekin jarraituz (zezenari adarra hautsi zitzaion), adarren itzulpen zuzena lortu ahal izateko ezinbestekoa da animaliarekin zerikusia duela jakitea (1. taulan *A1* adiera), eta zezena animalia dela. Hiztegian badira beste adar batzuk: zuhaitzarekin zerikusia duena (*A4*), musika-tresna dena (*A2*), adarkia (*A3*), baina horiek ez dira zezenari adarra hautsi zitzaion adibidean dugun adarren adiera. Perpaus horretako adar hitzaren adiera desanbiguatuta, itzulpen zuzena emateko gai izango gara.

Gure kasuan, oinarri gisa euskarazko hiztegi elebakarra izan dugu, *Euskal Hiztegia* (EH) (Sarasola, 1996), eta bertatik informazioa automatikoki erauzi dugu. Hala ere, lan guztia ez da automatikoa izan, erauzitako informazioa antolatze marko linguistikoa sortu baitugu. Beraz, erabili dugun metodoa erdi-automatikoa izan da.

Hitzen arteko erlazio lexiko-semantikoak erauzteko, lehenik hitzen artean dagoen azaleko erlazioa zein den esan behar da eta, erlazio bat baino gehiago badago, desanbiguatu egin behar da. Behin hori egin ondoren, definizio horretan agertzen diren hitzen adierak zeintzuk diren jakin behar da, aurretik desanbiguatua izan den erlazioa ez baita bi hitzen artekoa —sakonean, behintzat—, baizik eta bi adieren edo bi kontzeptuen artekoa.

Beraz, hasieran hitzak lotuko baditugu ere, azken helburua kontzeptuen arteko lotura zehaztea da. Eta kontzeptuen arteko erlazioak dituen datu-basea dagoeneko ez da datu-base arrunta. Halako informazioa duten datu-baseak *hiztegi ezagutza-baseak* (HEB), *ezagutza-base lexikalak* (EBL), eta *ontologiak* dira.

1.2. Ezagutza-base lexikalak, hiztegi ezagutza-baseak eta ontologiak

Ezagutza-base lexikalek (EBL) hitzei eta adierei buruzko informazioa dute. EBL batean konputagailuak hizkuntza ulertu ahal izateko hitzei buruz jakin beharreko guztia egon beharko luke (Yokoi, 1995). Hauen ezaugarri garrantzitsuen herentzia

izaten da, adierak *klase / azpiklase* hierarkien inguruan antolatzen dira-eta (Copestake, 1990).

Hiztegi ezagutza-baseek (HEB), berriz, hiztegietatik erauzitako informazioa jasotzen dute (Artola, 1993). Erauzitako informazioaren artean, hemen ere, adieren hierarkiak dira aipagarriak. HEB batean, hiztegian dugun informazio inplizitua esplizitu bihurtzen da, eta, hala, programa batek —edo giza erabiltzaileak— erabili ahalko du. EBL batean dugun informazioa ez da hiztegi batean dugun bera, hiztegietako informazioaz gain, bestelako informazioa ere egoten baita.

Ontologiak, berriz, mundu errearen kontzeptualizazioak dira, mundu erreari buruzko inferentziak egiteko gaitasuna dutenak. Definizio lauso hau aukeratu dugu, Adimen Artifizialaren arloan definizio zehatzagoek eztabaida pizten dutelako, eta, tesi-lan honi dagokionez, ontologiaren ezaugarri bat delako guretzat garrantzitsua: hierarkia darabiltela bizkarrezur gisa. Ontologiak aplikazio askotarako eraiki izan dira —*software*aren berrerabilgarritasuna, medikuntzako sistema adituak, lengoaia naturalaren sorkuntza, ulermena, itzulpena, eta abar—, eta normalean eremu espezifikoetarako eraiki ohi dira.

Hona ontologiak definitzeko egin den saiakeretako bat:

Ontologies are often equated with taxonomic hierarchies of classes, class definitions, and the subsumption relation, but ontologies need not be limited to these forms. (...) The word "ontology" seems to generate a lot of controversy in discussions about Artificial Intelligence. It has a long history in philosophy, in which it refers to the subject of existence. It is also often confused with epistemology, which is about knowledge and knowing. (...) In the context of knowledge sharing, I use the term ontology to mean a specification of a conceptualization. That is, an ontology is a description (like a formal specification of a program) of the concepts and relationships that can exist for an agent or a community of agents. This definition is consistent with the usage of ontology as set-of-concept-definitions, but more general. And it is certainly a different sense of the word than its use in philosophy (Gruber, 1993).⁵

Gruber (1993), Onyshkevych eta Nirenburg (1994) eta Guarino (1997) bat datoz ontologiak oso heterogeneoak eta norberearen beharren arabera eginak direla esaterakoan. Hala ere, ontologia guztiek dute kontzeptu zerrenda bat, eta kontzeptu

⁵ <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html> (2004-11-30ean atzitua).

horien arteko hierarkia *klase / azpiklase* erlazioak egituratzen du. Hori da ontologiaren ezaugarriarik garrantzitsuenetakoa.

Aurrekoak ikusita, badirudi EBLak eta ontologiak berdintsuak direla, baina ba al dute ezberdintasunik?

Ontologietan nabaria da kontzeptuen arteko *klase / azpiklase* hierarkiak duen garrantzia. HEBetan ere erlazio honek egituratzen du ezagutza-basea, nahiz eta adieren arteko erlazioari *hiperonimia / hiponimia* deitu. Autore batzuk ontologia eta EBLen arteko diferentziak azpimarratzen saiatu dira eta ez ditugu guk horiek ukatuko: kontzeptu guztiak lexikalizatuta egon behar duten edo ez, hierarkiaren goiko aldea antolatzeke irizpideak linguistikoak soilik diren edo ez, eta abar. Hala ere, beraien arteko mugak ez daude garbi. Adibidez, EAGLES taldearen lexikoari buruzko behin-behineko barne-txostenean (EAGLES, 1998), *WordNet*⁶ ontologiatik hurbil ikusten dute, nahiz eta EBL gisa aurkezten den beti.

WordNet can best be characterized as somewhere in between a semantic network and a conceptual ontology. The synsets are conceptual units rather than lexical semantics units. The relations are better seen as semantic inferencing schemes than as lexicalization patterns.⁷

Communications of the Association for Computing Machinery aldizkariaren 1995eko azaroko alean, EDR⁸ (Yokoi, 1995; EDR, 1995), CYC⁹ (Lenat, 1995) eta *WordNet* (Miller *et al.*, 1993) konparatzen dira. Aipatzen den diferentzia nabariena orientazioa da: CYCek sen ona eta munduari buruzko inferentziak jaso nahi ditu, eta EDR eta *WordNet* inferentzia linguistikoetarako daude prestatuta. Baina diferentzia hori ez da hain garrantzitsua, EAGLESen barne txostenean (EAGLES, 1998) aitortzen den bezala *EuroWordNet*¹⁰ buruz ari direnean:

EuroWordNet is different from AI-ontologies such as CYC or Sensus / Pangloss in that its focus is on the linguistically-motivated relations rather than the semantic inference schemes only. In this respect it provides information on the exact semantic relation between the

⁶ <http://www.cogsci.princeton.edu/~wn> (2004-11-30ean atzitua). II. kapituluaren azalduko dugun bezala, sinonimia erlazioaren inguruan antolatutako ezagutza-basea da.

⁷ <http://www.ilc.cnr.it/EAGLES96/rep2/node20.html> (2004-11-30ean atzitua).

⁸ <http://www.ijnet.or.jp/edr> (2004-11-30ean atzitua). *Japan Electronic Dictionary Research Institute* sortutako ezagutza-basea.

⁹ LNPaz gain adimen artifizialarentzat erabilgarria den ontologia. <http://www.cyc.com> (2004-11-30ean atzitua).

¹⁰ II. kapituluaren azalduko dugun bezala, *EuroWordNet*en oinarria *WordNet* da.

lexicalized words and the expressions of languages (this may still be useful for making inferences as well).¹¹

Diferentzia nagusia orientazioan dagoela uste dugu guk ere: ontologietan munduari buruzko informazioa dugu, kontzeptuen arteko erlazioek ez dute zertan motibazio linguistikorik eduki. Bestalde, EBLek hizkuntzaren ulermenerako eta sormenerako beharrei erantzun nahi diete, baina, azken finean, jakina da LNParent muturrera iristeko hizkuntzan agertzen diren arazo guztiak gainditu beharko direla, sen ona barne. Beraz, EBLetan munduari buruzko informazioak egon behar du. Adibide garbi bat *hiperonimia* erlazioa da. Izan ere, ontologietan eta EBLetan gordetzen den informazio semantikoa gainjarri egiten da; biak egitura isolatu bezala diseinatuko balira, ezagutza bera bi aldiz errepresentatu beharko litzateke, adibidez, *hiperonimiari* dagokion ezagutza.

I.3. Euskararen Ezagutza-Basea eta *Euskal Hiztegia*

Ixa Taldearen xedeetako bat *Euskararen Ezagutza-Basea* (EEB) sortzea da, eta horretarako bi bide jorratzen ari dira. Batetik, dagoeneko beste hizkuntzetarako garatuta dagoen *EuroWordNet* oinarri hartuta *EuskalWordNet* (ikus II. kapitulua) sortzea. Lan hau eskuzkoa da eta introspekzioan oinarritzen da. Bestetik, tesi-lan honetan ikusiko duguna: **euskara iturri duen EH hartu eta *Euskal Hiztegiaren Ezagutza-Basea* (EHEB) sortzeko ahalik eta erlazio lexiko-semantiko gehien automatikoki erauzi**. Lan honek alderdi introspektiboa eta deskriptiboa biltzen ditu. **Erauzketa bera erabat automatikoki egin da, baina elementuen ezagutza ahalbidetzen dituzten erregelak eskuz landu dira**. Tesi-lan honetan EHEB eta *EuskalWordNet* lotzeko ahaleginik egin bada ere, etorkizuneko lana izango da erabateko loturak gauzatzea. Horretarako kontzeptu mailako loturak eta erlazioak lortu behar ditugu, eta adiera-desanbiguazioan eta erlazioen desanbiguazioan sakondu.

Beraz, EHEBa sortzen hasteko EH bera eta bertako definizioak nolakoak diren aztertu behar dugu bertatik informazioa nola erauzi jakin ahal izateko.

EH (Sarasola, 1996) hiztegi elebakarra da, eta hiztegi elebakarrek dituzten hainbat ezaugarri ahalbidetzen dute hitzen arteko erlazio lexiko-semantikoak

¹¹ <http://www.ilc.cnr.it/EAGLES96/rep2/node20.html> (2004-11-30ean atzitu).

erauztea. Alde batetik, hitz bat definitzerakoan, hiztegi elebakarren definizioen zirkularitatea dela-eta, hiztegian sarrera duten beste hitzak erabiltzen dira —teoria mailan, behintzat—; eta, bestalde, definizio hauek berek —nahita edo nahi gabe— definizio-patroi edo eskema bertsuak erabiltzen dituzte hitza definitzerakoan.

Smith-ek eta Maxwell-ek (1980) dioten bezala, hitza definitzeko hiru modu erabiltzen dira hiztegietan:

- **Sinonimo** baten bitartez: sarreraren esanahi berbera duen hitza erabiliz (ikus IV.2. atala).

akabatu A1 ad. **Bukatu, amaitu.**

- Definizio klasiko baten bitartez: Aristotelesek definitutako *genus + differentia specifica*. Sarrera lexikala hitz generiko baten bitartez (*genus*) eta bereizgarri dituen ezaugarri batzuen bitartez (*differentia specifica*) definitzen da. Sarrera lexikala eta genusaren artean erlazio *hiperonimikoa* dago (ikus IV.3. atala).

aireontzi A0 iz. Hegalda daitekeen zernahi **ibilgailu.**

Adibidean, hitz generikoa edo *hiperonimoa* ibilgailu da; eta sarrera lexikala, aireontzi, termino zehatza edo *hiponimoa*.

- **Erlatore berezien** bitartez: sarrera lexikala eta definizioeko gune semantikoa harremanean jartzeko bide sintaktikoak dira. Erlatore berezi honek erakutsiko du zein harreman lexiko-semantiko dagoen sarrera-hitza eta definizioaren gune semantikoaren artean (ikus IV.4. atala).

ezpara A0 iz. Habea, Tabanidae familiako intsektuei
ematen zaien izena.

Erlatore berezi hauek dira definizioaren gune semantikoa eta sarrera-hitza erlazionatzen dituzten hitzak —edo hitz multzoak—. Adibideko erlatore berezia -i ematen zaien izena da, nahiz eta erlatore gisa definizioeko gune sintaktikoa (izena) besterik ez dugun markatuko.¹² Erlatore berezi honek sarrera-hitza (ezpara) eta definizioeko gune semantikoa (intsektu) erlazionatzen ditu.

¹² Definizioetan elementuak markatzeko III.7. atalean ikusiko dugun *Murritzapen Gramatika* erabili dugu, eta gramatika honek dituen ezaugarriak direla-eta, erlatore gisa hitz bakarra besterik ezin da markatu.

Aurrekoaz gain, gerta daiteke definizioaren gune sintaktikoa hutsa izatea semantika aldetik —Bruce *et al.*-en (1992) hitzetan, *disturbed heads*—, eta definizioan dagoen beste hitzen batek izatea garrantzi semantiko handiagoa. Honelako hitzei ere erlatore berezi deituko diegu, eta azpian zein erlazio lexiko semantiko duten bakoitzari dagokion atalean ikusiko dugu.

epaitondo A0 iz. Motzondoa, enborra moztu ondoren lurrari atxikia gelditzen den zuhaitz **zattia**.

Aurreko adibidean, deskribapenaren gunea zati da, baina benetan informazio interesgarria ematen duena zuhaitz da.

Definizio-patroi orokor hauek ikusi ondoren, hona **EHko definizioetan aurki ditzakegun definizio-egitura nagusiak**. Zenbaitetan sarrera-hitz baten definizio-eremuan definizio bat baino gehiago daude. Hala gertatzen da definizio-eremu horretan “;” dugunean.

foruzaletasun A0 iz. Foruen zaletasuna; foruzaleen dotrina.

Kasu honetan foruzaletasun hitzak bi definizio ditu: bata zaletasun gisa, eta bestea dotrina gisa.

Bestalde, definizio-eremuan definizioak egiteko modu bat baino gehiago aurkitzen dugu. Zenbaitetan, sarrera-hitza sinonimoen bitartez agertzen da definituta; beste zenbaitetan deskribapen baten bitartez —honi deitu ohi zaio definizio—; baina biak ager daitezke elkarren ondoan, sinonimoak eta deskribapenak. Definizio-patroi berezi hauek kasuz kasu aztertuko ditugu bakoitza dagokion atalean, baina eskema orokorra hau dugu:

- | | | | | |
|-----|---------------|---------------|-----------|--------------------------|
| (1) | sarrera-hitza | adiera-ikurra | kategoria | Deskribapena. |
| (2) | sarrera-hitza | adiera-ikurra | kategoria | Deskribapena, sinonimoa. |
| (3) | sarrera-hitza | adiera-ikurra | kategoria | Sinonimoa, deskribapena. |
| (4) | sarrera-hitza | adiera-ikurra | kategoria | Sinonimoa. ¹³ |

Hona aurreko definizio-patroiei dagozkien adibideak. Adibideek dituzten zenbakiak bat datoz definizio-patroien zenbakiekin.

(1) galai A0 iz. Dama bati edo emakume bati laguntzen dion gaztea.

(2) gaztanbera A0 iz. Gazura kendu zaion esne gatzatua, mamia.

¹³ Definizio eremuan deskribapen bat azaldu edo ez, ez da kontu aleatorioa, lexikografoak oso gogoan izandako eskemak baizik. Deskribapena agertzen duten hitzak dira garrantzi handien dutenak, ondoren deskribapena eta sinonimoa dutenak, ondoren sinonimoa eta deskribapena, eta, azkenik sinonimoa besterik ez dutenak. Sinonimoa besterik ez duten hauetan agertzen den hitza hiztegian bilatuz gero, normalean deskribapen baten bitartez definitua agertuko da.

(3) gorni A0N1 iz. Gorrina, landareen gaitza.

(4) fotografia A1 iz. Argazkia.

Definizio horietatik hitzen arteko erlazio lexiko-semantikoak erauzteko, EH bera modu berezian prestatu dugu.

I.4. EHren prestaketa

EH (Sarasola, 1996) paperezko hiztegi gisa jaio zen. Guk makinek irakurtzeko moduko formatu batean behar dugu bertan dagoen informazioa interpretatu nahi izanez gero. Horretarako, badugu formatu ez-egituratu batean dagoen *Microsoft Word* bertsioa, baina guk formatu egituratu bat behar dugu, emaitzak onak izan daitezen. Lehenengo pauso gisa, hiztegia III.3. atalean ikusiko dugun TEI ekimenaren arabera markatu dugu. TEIk proposatutako gidalerroei esker, bereizita agertuko dira hiztegian dauden eremu guztiak. Eta bereizketa honi esker, guri beharrezko zaizkigun elementuak bakarrik hautatuko ditugu: sarrera-hitza, kategoria, adiera-ikurra eta definizio-eremua. Albo batera utziko ditugu sarrera-hitz baten ondoren agertuko zaizkigun bestelako elementuak.

Kategoria-eremua garrantzitsua da azterkizun izango diren elementuak hautatzeko garaian. Aztertu ditugunak *izenak*, *adjektiboak*, eta *aditzak* dira. Kategoria bakoitzak definizio-patroi bereziak ditu, eta honexegatik, bakoitzarekin multzo bat osatu dugu. Baina hiztegian agertzen diren sarrera-hitz guztiek ez dute kategoria bakarra, eta, zenbaitetan, badira adiera berean *izen* eta *adjektibo* kategoria duten hitzak. IV.4.11. atalean ikusiko dugun bezala, hauekin multzo berezia egin dugu, eta definizio-patroien arabera saiatu gara zehazten —nahiz eta bi kategoria izan— ea lexikografoak kategoria bateko edo besteko eskema erabili ote duen. Beste multzo berezia *azpisarrerak* ditugu (ikus IV.6. atala). Azpisarrerak askotan kategoria-eremurik gabe agertzen dira, eta automatikoki sarrera-hitzaren kategoria esleitu diegu. Kategoria hori ez da beti zuzena, eta aurreko multzoan bezala, kategoria bakoitzeko definizioak aplikatu dizkiegu sarrera-hitzaren kategoria zein den erabakitzeke. Hauen artean izenen, adjektiboen eta aditzen definizio-patroiak daude.

Hiztegian bertan zer dugun jakiteko, hiztegian dauden elementuen azterketa estatistikoa egin dugu, bertan usuen agertzen diren hitzak eta elementu-segidak

zeintzuk diren jakiteko (ikus III.8. atala). Elementu horiei *token* deitu ohi zaie, eta hauen barruan ditugu hitz-formak, postposizioak —kasu gramatikalak, postposizio-atzizkiak eta postposizio konposatuak—, parentesiak, gidoiak, puntuazio-ikurrak, eta abar. Definizio-patroi bat ezagutu ahal izateko, *token*-segida horiek guztiak hartu ditugu kontuan.

Token-segida horiek ezagutzeko, definizioen analisi morfologikoa egin dugu (ikus III.5. atala). Analisi horiei esker, dagoeneko markatuta ditugu definizioetan agertzen diren hitz-formen lemak, hauek hartzen dituzten postposizioak, mugatasunak, agertzen diren puntuazio-ikurrak, parentesiak, komatxoak, eta abar. Hiztegiko definizioen barruan ezkutatuta dagoen informazio inplizitua dagoeneko esplizitu dugu, eta makinak badaki definizio baten barruan zein *token* dauden.

Hiztegiko elementuen azterketa estatistikoa eginda, erlazioak erauzteko garaia da.

I.5. Erlazio lexiko-semantikoen erauzketa

IV. kapituluaren sarrera-hitza eta definizioko hitz baten arteko erlazioak erauzi ditugu. Garrantzitsuena, erlazio hierarkikoa da. Definizio-eremuan deskribapena dagoenean, deskribapen horren gunea da sarrera-hitzaren *gurasoa*. Aurretik aipatu dugun bezala, Aristotelesen genusa (ikus IV.3. atala). Sarrera-hitzarekin erlazioan dagoen hitza ezagutzeko garaian, testuingurua hartzen dugu erreferentzia gisa, eta lan-tresna gisa testuingurua kontuan hartzen duen gramatika erabili dugu: *Murritzapen Gramatika* (*Constraint Grammar*) (Karlsson *et al.*, 1995; Tapanainen, 1996) (ikus III.7. atala).

Hierarkia marka dezaketen beste elementuak erlature bereziak dira. Erlature bereziek genusaren lekua hartzen dute definizioan, baina esanahi gabeak edo esanahi ahulekoak direnez, definizioko beste elementu bat jartzen dute sarrera-hitzarekin harremanetan. Erlature berezi hauek markatzen duten erlazioa gehienetan ez da hierarkikoa (ikus IV.4. atala).

Aurrekoak bezala, sinonimoak ere sarrera-hitzaren sare semantikoa zehazteko baliagarri diren elementuak dira. Sinonimoen arteko erlazioa ez da hierarkikoa, baizik eta maila berekoa (ikus IV.2. atala).

Baina kategoria guztiek ez dituzte erlazio berak (Miller *et al.*, 1993). Adibidez, adjektiboen kasuan ez dugu erlazio hierarkikorik izango, eta erlazonatutako hitz guztiak sare semantikoaren elementuak dira. Zenbait kasuetan sare semantiko baten barruan nolabaiteko *gradazioa* badagoenez, badirudi erlazioa hierarkikoa dela, baina erlazio hierarkikoaz baino gehiago *familiartasunaz* hitz egiten da. bikain adjektiboaren definizioa hartuta, badirudi *onen hiperonimoa* dela:

bikain H2A0 izond. Guztiz **ona**, dohain onez txit hornitua.

Baina adjektiboen kasuan ez dugu erlazio hierarkikorik, eta *onen* familia berekoak dira: on, bikain, txar, eratsu, onxko **eta** pasagarri.

txar	A1	izond. Funtsezko akats bat duena, bere egitekoa ongi betetzen ez duena, ona ez dena.
erasu	A0	izond. Eraz betea, ona , egokia.
onxko	A0	izond. Ona , aski ona .
pasagarri	A1	izond. Onar daitekeena, baina ez oso ona , nolanahikoa.

VI. kapituluan Aristotelesek *differentia* deitzen duen elementuaren azterketa egin dugu. Modu horretan definizioko hitzen arteko erlazioak zehaztu ditugu eta aurretik sortuta dugun sare semantikoa osatu. *Differentia* hori erauzteko erabili izan diren teknikak bi dira: definizio barruko **patroi lexikalak** detektatzea eta hitzen arteko loturak egiten dituzten **preposizioak edo postposizioak**¹⁴ aztertzea.

Patroi lexikalek erlatore berezien antza dute; patroiaren barruan genusa agertzen da, eta genus honen eta definizioko beste hitz baten arteko erlazioa zehazten dute. Patroi lexikal jakin hauek, lotzen dituzten elementuei buruzko informazio interesgarria emango dute (ikus VI.2. atala).

haritz A1 iz. Pagoaren familiako **zuhaitza**, (...).

Aurreko adibidean, haritzen genusa dugu zuhaitz, eta -en familiako patroi lexikalaren bitartez zuhaitz eta pago lotzen dira.

Hitzen arteko loturak gauzatzen dituzten beste elementuak **postposizioak** dira. Postposizio hauek sintagma ezberdinen arteko loturak dira, eta interpretazio jakina dutenez, interpretazio horrek markatuko du hitzen artean dagoen erlazioa (ikus VI.3. atala).

¹⁴ Euskararen kasuan postposizioak ditugu, baina inguruko beste hizkuntzetan lotura-elementu horiek preposizioak dira.

ildo A1 iz. Goldea~~z~~ lurra irauli~~z~~ egiten den
 irekidura luzea.

Adibide honetan postposizio-atzizki *instrumentala* bitan agertzen da. ildoren genusa irekidura *da*, eta irekidura hori irauliz egiten *da*, eta iraultzeko, goldea erabiltzen *da*. egin eta irauliren arteko erlazioa *modua* *da*, eta irauli eta golderen artekoa *instrumentua* (ikus VI.4.1. atala).

Lan hau gauzatu ahal izateko, postposizio horien interpretazio multzoa zein den zehaztu dugu. Postposizio baten interpretazioa desanbiguatzeko beste hizkuntzetako preposizioak erabiliko ditugu. Desanbiguazio hori gauzatu ahal izateko, postposizioek eta preposizioek interpretazio-zerrenda bateratua beharko dute. Postposizio batek hainbat interpretazio izan ditzake, baina beste hizkuntza batean horren baliokide den perpausa baldin badugu, eta postposizio horren lekuan preposizioa, horien interpretazioa bien arteko interpretazio-ebakiduratik¹⁵ aterako da. Beraz, oso garrantzitsua izango da euskarazko postposizioei esleitutako interpretazioak beste hizkuntzetako preposizioei esleitutako interpretazioekin bateragarriak izatea. Hiztegiak corpus paralelo —berezi— gisa erabiliko ditugu, eta VI.7. atalean ikusiko dugun teknikari esker corpus paralelo lerrokatu gisa, gainera. Lanaren atal honetan ia-ia itzulpen automatikoa egiten ari garela esan ahalko dugu, eta itzulpen-lan horietarako preposizioen eta postposizioen interpretazio bateragarriak izatea oso garrantzitsua da.

Eta eraikitako sare semantikoa handitzeko asmoz hiztegi-sarrera bera analizatu dugu. Sarrera-hitza eratorria baldin bada, oinarri eta eratorriaren arteko erlazioa zehazten saiatuko gara V. kapituluan. Eratorpenarekin egindako lana *EuskalWordNet*ekin lotu dugu partzialki. Horretarako erlazioak kontzeptuen artekoak bihurtu ditugu *EuroWordNet*en dauden beste hizkuntzetako hitzei esker. Guk ingelesa eta espainiera erabili ditugu, eta kontzeptu mailako lotura ematen dugunez, euskararentzat lortu dugun informazioa, beste hizkuntzentzat ere baliagarria da.

¹⁵ *Ebakidura* hitzarekin, bi hitzek dituzten balizko kontzeptuen artean kontzeptu bera zein diren adierazi nahi dugu.

I.6. Azterketa eta deskribapen linguistikoak

Honaino azaldutako prozesu guztia garatu ahal izateko, egin behar izan ditugun azterketa eta deskribapen linguistikoak honakoak dira:

- A. eranskinean erlatore bereziak paratu ditugu. Bertan daude aztertu ditugun erlatore berezi guztiak eta hauek sortzen dituzten erlazioak. Hauen laburpena tesiaren IV. kapituluan egin dugu. Kapitulu honetan sinonimoen eta genusen azterketa egin dugu, sortzen dituzten erlazio lexiko-semantikoak azalduz.
- B. eranskinean eratorpenaren azterketa azaldu dugu. Eratorpena tesi-txosten honen V. kapituluan azaldu dugu, baina eratorpen-atzizki bakoitzari buruzko informazioa handia denez, zati handiena eranskinean ematea erabaki dugu. B. eranskinean, beraz, eratorpen-atzizki bakoitzaren interpretazio-bilduma dago. Eratorpen-atzizkien interpretazio bakoitzak hiztegiko definizioetan zein isla duen aztertu dugu, eratorriaren eta oinarriaren arteko erlazioa zehazteko.

Hitz eratorriak ezagutzeko garaian, analizatzaile morfologikoari hainbat aldaketa egin behar izan dizkiogu. Erabilitako analizatzaile morfologikoak Ixa Taldeak garatutako EDBLetik hartzen du informazioa, eta bertan zenbait atzizki lexikal landuta egon arren, ez daude denak. Bertan atzizki lexikalak txertatu ahal izateko, hauek jasaten dituzten erregela morfo-fonologiko guztiak aztertu beharko lirateke, eta gainera, datu-basean oraingoz modu zabalean landu ez den oinarriari buruzko informazioa gehitu —lan honen albo-ondorioetako bat da datu-base horretako zenbait elementuen informazioa gehitzea, adibidez, izenen biziduntasuna—. ¹⁶ Beraz, lan egiteko beste bide bat hautatu dugu. Analizatzaile morfologikoari atzizki lexikal berriak gehitu dizkiogu, eta, hori egiteko, hiztegiko sarreraren amaiera usuenen azterketa egin dugu. Ondoren, horietatik atzizki lexikalak —edo hauen aldaerak— zeintzuk diren erabaki dugu, eta horiek txertatu dizkiogu analizatzaileari horiei buruzko informazioa eman dezan. Zerrenda horretan konposizio-elementuak ere agertu zaizkigu, eta maiztasun handikoak diren heinean, txertatzea erabaki dugu hitz konposatuaren esanahiaren berri izan dezakegulakoan. Hau guztia V. kapituluan azaldu dugu.

¹⁶ Eratorpen atzizkiek jasaten dituzten aldaketa morfo-fonologikoak aztertu ditu Izagirre-k (2004).

- C. eranskinean patroï lexikalak aztertu ditugu eta hauek sortzen dituzten erlazio motak azaldu. Eranskin honen laburpena VI. kapituluari jarri dugu.
- D. eranskinean euskarazko postposizio-atzizkien eta kasu gramatikalen azterketa paratu dugu. Hauen interpretazioak zeintzuk diren aztertu eta eman ditugu. Azterketa honen ondorioa VI. kapituluari emandako laburpena.
- E. eranskinean postposizio konposatuak aztertu ditugu. Hauek dituzten interpretazioen arabera bi taldetan sailkatu ditugu: anbiguoak eta ez-anbiguoak. Hauen laburpena VI. kapituluari ikus daiteke.
- F., G. eta H. eranskinak D. eta E. eranskinekin lotuta daude. F. eranskinean ingelesezko preposizioak daude interpretazioen arabera banatuta. G. eranskinean espainierazkoak modu berean antolatuta; eta H. eranskinean euskarazkoak modu berean antolatuta. Hiru eranskin hauetan hizkuntzak lotzen saiatzen gara postposizioen eta preposizioen interpretazioen bitartez, eta itzulpen automatikorako erabilgarri izango diren *lotura zuzenen* berri eman dugu. Hau guztia VI. kapituluari azaldu dugu.
- Tesian zehar definizio-patroï ezberdinen berri eman dugu, EHn agertzen diren definizio motak aztertu baititugu. Definizio horietan erlazioen bat duten elementuak markatzeko *Murriztapen Gramatika* erabili dugu, eta egindako erregela guztiak I. eranskinean daude.

I.7. Tesi-lanaren egitura

II. kapituluari aurrekariak aipatu ditugu, bertan ikusiko dugu arlo honetan beste hizkuntzetan orain arte egin izan diren lan garrantzitsuenak zeintzuk izan diren. III. kapituluari hiztegiaren datu-azterketa egin dugu; bertan agertzen diren elementuen maiztasuna, hiztegiaren beraren ezaugarriak eta agertzen diren definizioen hainbat ezaugarri eman ditugu. IV. kapituluari sarrera-hitza eta definizioaren elementu baten arteko erlazioak aipatu ditugu, hala nola, sinonimoak, genusak eta erlatore bereziak. V. kapituluari, sarrera-hitza albo batera utzi gabe, eratorpena aztertu dugu, hau da, sarrera-hitz eratorriek haien oinarriekin duten erlazioa zein den zehaztu dugu. VI. kapituluari, berriz, sarrera-hitza utzi eta definizioan agertzen diren elementuen arteko erlazioak zeintzuk diren zehaztu dugu; horretarako definizio berruko patroï lexikalak

eta lotura-elementu diren postposizioen interpretazioak erabili ditugu. Eta, VII. kapituluaren lan guztiaren ondorio nagusiak eta etorkizuneko lanak bildu ditugu.

I.8. Tesiarekin lotutako publikazioak

Atal honetan tesiaren joanean —eta honekin zerikusia duten— argitalpenak paratu ditugu.

Egileak	Urtea	Titulua	Non Argitaratua	Kapitulua
Lersundi M. eta Agirre E.	2003	Semantic interpretations of postpositions and prepositions: a multilingual inventory for Basque, English and Spanish	Workshop on "The linguistic dimensions of prepositions and their use in computational linguistics formalisms and applications". Tolouse, France.	VI.
Agirre E., Ansa O., Arregi X., Artola X., Diaz de Ilarraza A., Lersundi M.	2003	A Conceptual Schema for a Basque Lexical-Semantic Framework	Conference on Computational Lexicography and Text Research (Complex), Budapest	orokorra
Lascrain V., Agirre E., Lersundi M., Popelínsky L.	2003	Disambiguation of case suffixes in Basque	TALN workshop on Automatic Treatment of minority and small languages	VI.
Agirre E., Lersundi M., Martinez D.	2002	A Multilingual Approach to Disambiguate Prepositions and Case Suffixes	ACL Workshop: Word Sense Disambiguation: recent successes and future directions	VI.
Aldezabal I., Aranzabe M., Atutxa A., Lersundi M.	2002	Levin-ek English verb classes and alternations (1993) liburuan proposatzen dituen ingeleserako alternantziak euskararekin parekatuz	Barne txostena. UPV/EHU / LSI / TR 13-2002	VI.
Agirre E., Lersundi M.	2001	Extracción de relaciones léxico-semánticas a partir de palabras derivadas usando patrones de definición	Proceedings of the Annual SEPLN meeting, Jaen	V.
Aldezabal I., Ansa O., Arrieta B., Artola X., Hernández G., Lersundi M.	2001	EDBL: a General Lexical Basis for the Automatic Processing of Basque	IRCS Workshop on linguistic databases. Philadelphia (USA).	VI.
Agirre E., Ansa O., Arregi X., Arriola J.M., Artola X., Diaz de Ilarraza A., Lersundi M., Martinez D., Sarasola K., Soroa A., Urizar R.	2001	Euskal Hiztegiatik hitzen arteko erlazio semantikorantz: sinonimoen, genusen eta erlatore berezien bilaketa	Barne txostena: UPV/EHU/LSI/TR 2-2001.	IV.

2. taula. Tesi-txosten honekin zerikusia duten argitalpenak, dagozkien kapituluekin lotuta.

Egileak	Urtea	Titulua	Non Argitaratua	Kapitulua
Diaz de Ilarraza A., Lersundi M., Mayor A., Sarasola K.	2000	Etiquetado semiautomático del rasgo semántico de animicidad para su uso en un sistema de traducción automática	Actas del XVI Congreso de la SEPLN Vigo	IV.
Agirre E., Ansa O., Arregi X., Artola X., Diaz de Ilarraza A., Lersundi M., Martinez D., Sarasola K., Urizar R.	2000	Extraction Of Semantic Relations From A Basque Monolingual Dictionary Using Constraint Grammar	Proceedings of Euralex Stuttgart (Germany).	IV.
Agirre E., Ansa O., Aranzabe M., Arregi X., Arriola J.M., Artola X., Diaz de Ilarraza A., Lersundi M., Martinez D., Sarasola K., Soroa A.	2000	Hiztegieta oinarritutako hitzen adiera-desanbiguazioa euskaraz	Barne txostena: UPV/EHU/LSI/TR 10-2000	IV.
Aldezabal I., Ansa O., Artola X., Ezeiza A., Gojenola K., Insausti J.M., Lersundi M.	1999	Euskararen Datu-Base Lexikala (EDBL): eskema berriaren proposamena	Barne txostena: UPV/EHU/LSI/TR9-99	orokorra
Agirre E., Ansa O., Arregi X., Arriola J.M., Diaz de Ilarraza A., Lersundi M., Soroa A., Urizar R.	1998	Extracción de relaciones semánticas mediante gramáticas de restricciones	Congreso SEPLN, Alicante..	IV.

3. taula. Tesi-txosten honekin zerikusia duten argitalpenak, dagozkien kapituluekin lotuta.

Udazkena... 2003...

udazken	A0	iz.	Udaren ondoko urtaroa, Lurraren iparraldean irailaren 21etik abenduaren 21era hedatzen dena.
kolore	A1	iz.	Gauzakien gainaldearen nolakotasuna, mota jakin bateko argi izpiak sortzean edo islatzean datzana, begiari era jakin batean eragiten diona.

Udazken koloretan,
landen lurrinak zeharkatuz,
hitaz oroït
eta higan nauk.

Zuhaitz biluziaren
gerizpean orpondoa,
orbelaren hilobi,
horiz eta gorriz
dena lokarturik.

Eskuratu hostoa
xume bezain eder,
hain soila bere heriotzean,
zuhaitz guztiaren bizkortasunez
hain betea,
non erortze honen duintasunak
hiri kantatzera naraman.

Berriro zuhaitzari
so natzaïok, kezkatuik ote?...
joate askearen oparotasunean,
betikotasunaren irria zirudik.

Gatibu naukan denborak,
bere baitan egositako
ametsetaz isekaz
bailegoan.

Udazken koloretan,
landen lurrinak zeharkatuz,
hitaz oroït eta higan nauk,
hain soil hire heriotzean,
hain xume, adiorik gabeko
partitzean.

(Benito Lertxundi)

II. Aurrekariak

Sarreran aipatu dugunez, LNPN aurrera egiteko lexikoen beharra azpimarratu izan da. Lexikoi horiek osatzeko garaian, zenbait galdera erantzun beharko ditugu, eta atal honetan beste hizkuntzetan egin izan diren pareko lanak ikusiko ditugu. Alde batetik, beste hizkuntzetan lexikoiak sortzeko erabili izan diren hurbilpenak eta iturriak zeintzuk izan diren ikusiko dugu. Iturriei dagokienez, hiztegietatik egindako erauzketa-lanei garrantzi berezia emango diegu, gurea ere halakoa izan delako. Beste aldetik, eratorpenari, patroi lexikalei eta preposizioei buruzko hainbat lan kontuan izan ditugu, V. eta VI. kapituluetan ikusiko dugunez, tesi-lan honetan informazioa erauzteko garaian garrantzi berezia izan dutelako.

Tesi-lan honetan erabilitako euskarari buruzko lanak zeintzuk izan diren dagozkien ataletan aipatuko ditugu.

II.1. Lexikoia sortzeko hurbilpenak eta iturriak

Sarreran ikusi dugunez, informazioa erauzteko garaian erabiltzen den hurbilpenaren arabera lan mota bat edo beste izango dugu. Hurbilpen nagusiak bi izan dira: **arauemailea** eta **deskriptiboa**. Hurbilpen hauekin lotura estua dute informazioa erauzteko metodoek: eskuzkoa, automatikoa eta erdiautomatikoa.

Hurbilpen arauemailea erabiltzen duten lanak eskuzkoak izan ohi dira. Hauen iturri nagusia introspektzioa da, hizkuntzalariak munduari buruz duen jakinduria eta ezagutza. Era honetan sortutako zenbait lexikoi badaude: *Word Dictionary*, 10.000 sarrera dituen *Linguistic String Projecter*ako (LSP) sortutako lexikoia (Fox *et al.*,

1988); *WordNet* (Miller, 1990) gaur egun 2.0 bertsioa da indarrean, eta 152.059 hitz daude bertan errepresentatuta, eta baita euren arteko erlazio semantikoak dituzten 115.424 *synset*¹⁷ edo sinonimo-multzo ere; *Comlex* (Grishman *et al.*, 1994) ingeleseko 38.000 inguru hitzentzako informazio sintaktikoa dakarren lexikoi konputazionala; CYC Ontologia (Lenat, 1995) 100.000 termino baino gehiago ditu. LDOCEren azken bertsioak, LDOCE3-NLP, 80.000 adiera ditu, eta linguistika konputazionalako ikerkuntzarako laguntza gisa sortu dute *Longmaneko* lexikografoek.

Nirenburg-ek eta Raskin-ek (1987) aurkezten duten lanean, hasieran kontzeptuen ontologia sortzen dute, eta ondoren kontzeptu horiei hitzak lotzen dizkiete. Metodologia hau azaltzeko garaian, bada oso adierazgarria den esamoldea: *lehenik mundua, eta ondoren hitzak*. Hurbilpen hau itzulpen automatikoa egiten duten sistema zenbaitetan aplikatu izan da. Horien adibide dugu METAL (Hutchins eta Somers, 1992). Eta lan egiteko sistema honi jarraitzen zaizkion proiektuen artean honakoak ditugu: CYC (Lenat eta Guha, 1990), *Upper Model* (Bateman, 1990), ONTOS (Carlson eta Nirenburg, 1990), *WordNet* (Miller, 1990) eta EDR (Yokoi, 1995).

Hurbilpen deskriptiboetara pasa aurretik, eta tesi-lan honetan erabiliko dugunez, hona euskarazko *WordNet*aren (*EuskalWordNet*) ezaugarriak. *EuskalWordNet* *EuroWordNet*en oinarrituta dago. Eta *EuroWordNet*ek, berriz, *WordNet*en diseinu bera du. *WordNet* ingelesez eta ingeleserako sortutako ezagutza-basea da. Ingelesezko adiera bakoitza hitz sinonimo multzo batekin —*synset* deritzonarekin— adierazten da eta adiera guztiak hierarkikoki antolatuta daude. *EuroWordNet* diseinu bereko beste EBL bat da eta Europako zortzi hizkuntzatarako —alemana, espainiera, estoniera, frantsesa, ingelesa, italiera, nederlandera eta txekiera— zabaldu da. *EuskalWordNet* metodo erdiautomatikoak erabiliz sortu da, baina lortutako emaitzak eskuz orrazten dira.

Hurbilpen deskriptiboan, berriz, aurretik sortuta dauden testu-baliabideak erabiltzen dira oinarri gisa, eta bertan dagoen informazioa erauzten saiatzen dira era automatikoak edo erdi-automatikoak erabiliz. Testu-baliabide horien artean ditugu MRDak, datu-base lexikografikoak, datu-banku terminologikoak, morfologikoki,

¹⁷ <http://www.cogsci.princeton.edu/~wn/index.shtml> (2004-11-30ean atzitu).

sintaktikoki eta semantikoki etiketatutako corpus elebakarrak eta elebidunak,... Hurbilpen hau juxtu arauemailearen kontrakoa da: *lehenik hitzak, eta ondoren kontzeptuak*. Hurbilpen honen adibide dugu Bruce *et al.*-ek (1992) sortutako LDOCE hiztegian agertzen diren izenen taxonomia. Bruce *et al.*-ek LDOCE hiztegiko definizioak hartu dituzte, genusak zeintzuk diren ezagutu, eta genus hauek hiztegi berean duten adieraren arabera desanbiguatu. Lortzen duten estaldura %90 da.

Eta iturri gisa corpusak erabiltzen dituztenen artean, berriz, Pereira *et al.*-ek (1993) hierarkia semantiko berria sortu dute multzokatze distribuzionala erabiliz.

Badira **introspektzioa eta deskripzioa uztartzen** dituzten lanak. Lehenik ezaugarri multzo minimoa proposatzen dute euren sistemako lexikoak gutxieneko koherentzia eta egitura izan dezan, eta, ondoren, hurbilpen deskriptiboak erabiliz, aurreko egitura arauemaileari lotu beharreko datu lexikoak automatikoki erauzten dituzten artean daude (Ageno *et al.*, 1992) eta (Hovy eta Knight, 1993) lanak.

Sarreran aipatu dugun bezala, **metodo automatiko eta erdiautomatikoek** erabiltzen dituzten **iturriak corpusak** eta **egituratutako baliabide lexikalak** izan ohi dira. Zenbaitetan bien arteko nahasketa diren metodoak ere erabili izan dira.

Corpus elebakarrei dagokienez, orain arte gehien erabili izan diren corpusak milioi bat hitz inguru dituen *Brown* corpora (Francis eta Kucera, 1982), eta milioi bat eta hiru milioi hitzen arteko kopurua —erabiltzen den materialaren arabera— duen *Wall Street Journal*aren materialak izan dira. Bi hauek *Penn Treebank* proiektuari esker markatu dira maila hauetan: hitzaren kategoria (Marcus *et al.*, 1993) eta azaleko egitura sintaktikoa (Marcus *et al.*, 1994). Aurrekoez gain, 250.000 hitzetako *Brown* corpusaren testu zati bat hartu dute (Miller *et al.*, 1993) eta Princeton-go kategoria-etiketatzailerik automatikoarekin etiketatu dute lehenik (Brill, 1992), eta ondoren eskuz *WordNet*eko adierekin. Ng-k eta Lee-k (1996) eskuz etiketatu dituzte *Brown* corpusean eta *Wall Street Journal*-en agertutako 191 izen eta aditzen 192.800 agerpen.

Euskaraz ditugun corpus garrantzitsuenak *XX. Mendeko Euskararen Corpus Estatistikoa*¹⁸ eta *Ereduzko Prosa Gaur*¹⁹ dira.²⁰ Ixa Taldean ikerketarako erabiltzen dira *Egunkaria* eta *Berria* egunkarien hemerrotekekin sortutako corpusak.

¹⁸ <http://www.euskaracorpora.net> (2004-11-30ean atzitu).

Corpus elebidunei dagokienez, ikertzaileen artean gehien erabili izan dena *Hansard* corpora izan da. Corpus honetan *Canadian Parliamentary Proceedings* daude, eta ingelesa eta frantsesa dira bertan aurkitzen ditugun hizkuntzak. Ingeleseko 3,5 milioi esaldiri dagozkien 97 milioi hitz daude, eta frantseseko 3,7 milioi esaldiri dagozkien beste 97 milioi hitz. Corpus hau lerrokatuta dago, hau da, hizkuntza bateko esaldi bakoitzaren beste hizkuntzako esaldi baliokidea zein den markatuta dago (Brown *et al.*, 1991).

Corpusak egiturarik gabeko iturri lexikaltzat hartzen dira, eta egituratutako baliabide lexikalen artean egitura askozaz nabariagoa duten iturriak ditugu: MRDak —hiztegiak—, tesauroak, eta entziklopediak.

Iturri erabilienak **hiztegiak** —MRDak— izan dira, eta tesi-lan honetan ere hiztegia erabiliko dugu. Hiztegietatik informazioa erauzteko metodoa ez da berria LNPn, eta honi buruzko laburpen interesgarriak ditugu (Castellón, 1992), (Artola, 1993), (Agirre, 1999), (Rigau, 1998) eta (Arriola, 2000) lanetan. Gure kasuan euskara landu nahi dugu, eta horretarako beste hizkuntzetan egindako lana gurera aplikatuko dugu.

Halako lanak 80ko hamarkadan hasi ziren.²¹ Amsler-en hainbat lanetan (Amsler eta White, 1979; Amsler, 1980; 1981) dagoeneko aipatzen da halakorik, eta hauen ondoren, *The Merriam-Webster New Pocket Dictionary* (W7N) (Chodorow *et al.*, 1985; Markowith *et al.*, 1986; Binot eta Jensen, 1987), eta *LONGMAN Dictionary of Contemporary English* (LDOCE) (Michiels eta Noël, 1982; Boguraev eta Briscoe, 1987; Jensen eta Binot, 1988; Boguraev eta Briscoe, 1989; Wilks *et al.*, 1990) hiztegien gainean egindako lanak publikatu ziren.

Lan horietan hiztegia oso izaera bereziko testu-corpustzat hartzen da (Van den Hurk eta Meijs, 1986; Boguraev *et al.*, 1989).²² Berezitasun hori corpus horretatik

¹⁹ <http://www.ehu.es/euskara-orria/euskara/ereduzkoa> (2004-11-30ean atzitua).

²⁰ *XX. Mendeko Corpus Estatistikoa* UZEIk sortutako corpus orekatua da eta 4.658.036 hitz ditu. *Ereduzko Prosa Gaur* corpora EHUK sortu du ereduzkotzat hartutako 86 idazlanekin. Honek ere 4.500.000 hitz baino gehiago ditu.

²¹ Hauen aurretik ere zenbait lan egin ziren, hala nola liburu gisa argitaratutako Sparck-Jones-en (Sparck-Jones, 1986) tesi-lana, edo Litkowsky-ren lanak (1978; 1980). Lan hauek, garai hartako teknologia informatikoaren egoera eta hiztegiak euskarri magnetikoan edukitzeko aukera urria zirela eta, aurrera egiteko hamaika oztopo aurkitu zituzten.

²² Van den Hurk-en eta Meijs-en (1986) lanaren izenburua aski adierazgarria da: *The dictionary as a corpus: Analyzing LDOCE's definition language*. Boguraev *et al.*-ek (1989) lexikoa eskuratzeko bi iturburu aipatzen dituzte, testu-corpus orokorrak eta MRDak; artikulu honetan metodologia bat

atera daitekeen informazio lexikalaren nolakotasunari zor zaio. Corpus orokor batetik hizkuntza-elementuei buruzko informazio asko eta asko atera badaiteke ere, —eta testu-corpusen azterketa gaur egun LNParen barruan indarrean dagoen arloetako bat da—, hiztegi batetik, bereziki hiztegi bateko definizioen analisitik, zuzenean atera daitekeen informazioa nekez aterako genuke corpus orokorrago batetik. Calzolari-k (1990b) dioen bezala, informazio-eskuratze hori ez da osoa, eta ezta hutsik gabekoa ere, baina bai oinarri sendoa jarriko duena.

Boguraev-ek (1987), Boguraev-ek eta Briscoe-k (1989) eta Calzolari-k (1990b) corpusetatik zein hiztegietatik jakintza lexikala ahal den neurrian automatikoki eskuratu nahiak dituen arazoei buruz dihardute.

Calzolari-k (1986; 1990a) hiztegia, hitzen esanahiari dagokionez, testu-corpus hutsa baino askoz ere gehiago dela dio. Hiztegia oso entitate egituratua dugu, jakintza orokor nahiz lexikalaren iturri estimagaitza.

Informazioa erauzteko, hiztegietatik ez ezik, badira egituratutako beste baliabide lexikaletatik informazioa erauzten duten lanak ere: Yarowsky-k (1992), Resnik-ek (1995) eta Grefenstette-k (1993) *Roget's International Thesaurus*a erabili dute.²³ Byrd-ek (1989) *Roget's II: The New Thesaurus*a eta *The New Collins thesaurus*a erabili ditu. Grefenstette-k (1993) *Macquarie's thesaurus*a erabili du. Sanchez-ek (1991) *Diccionario Ideológico de la lengua Española (J. Casares)* tesauro espainiarra erabili du. Eta Utsuro *et al.*-ek (1993) *Bunrui Goi Hyō* tesauro japoniarra erabiltzen dute.

Entziklopediei dagokienez, berriz, Yarowsky-k (1992) lexikoien sorkuntzarako *Grolier's Encyclopaedia* erabili du; eta Gomez *et al.*-ek (1994) *The World Book Encyclopedia* erabili dute.

Baina badira iturri batera edo bestera lotzen ez diren zenbait lan, eta bi motatako iturriak erabiltzen dituzte. Gerta daiteke batetik edo bestetik —eskuz edo automatikoki— sortutako lexikoiek hainbat hutsune izatea. Adibidez, Walker-ek eta Amsler-ek (1986) *The Merriam-Webster New Pocket Dictionary* hiztegiko sarrerak eta *New York Times* egunkariko 3 hilabetetako testuak konparatu dituzte. Emaizetan

finkatzen dute hiztegi-iturriok arakatzeko, hitzen ezaugarri semantikoak eskuratzeko asmoz, eta azken helburua lexiko konputazional baten alderdi semantikoa osatzea da.

²³ Tesauro honek 60.071 hitz ditu 1.000 kategoria semantikotan banatuta.

aipatzen dutenez, egunkari horretako berrietan agertzen ziren hitzen %64a ez zen hiztegian agertzen. Lautik bat hitz deklinatuak dira, beste laurdena izen bereziak, seiren bat marraz loturiko hitz konposatua, hamabiren bat gaizki idatzitako formak, eta, beste laurdena, erabaki gabekoak —hiztegia egin ondoreneko hitz berriak izan daitezkeenak, alegia—.

Aurrekoaz gain, Briscoe-k eta Carroll-ek (1993) diotenez, oinarrian iturri lexikalak dituzten etiketatze sintaktikoa egiten duten sistemen akatsen erdiak azpikategorizatutako balentziaren desagokitasunagatik gertatzen dira. Gainera, informazio lexikoa batzuetan lotuta dago domeinu jakin batzuetara, eta hori ez da hiztegi orokorretan agertzen.

Ezagutza-baseak eraikitzerakoan iturri bakartzat hiztegiak ez direla erabili behar diote Ide-k eta Veronis-ek (1993). Haien artikulua izenburua aski esanguratsua da: *Extracting knowledge bases from machine-readable dictionaries: Have we wasted our time?*. Artikulu honetan ondorioztatzen dutenez, ezagutza-baseak eraikitzekeo hiztegiak oso garrantzitsuak dira, baina, zenbaiten ustearen aurka, ezingo dira erabat automatikoki sortu. Hiztegiek hitzen arteko erlazioei buruzko informazio handia dutenez, ezin ukatuzkoa da oinarri indartsuak direla; baina, hauek dituzten hainbat arazo direla-eta, pertsonaren lana ezinbestekoa izango da. Artikulu berean aipatzen dute hiztegiak laguntza handia izango dutela hiztegia sortzeko garaian ezagutza-base hauek erabilia; eta LNPN lan egiten dutenen eta hiztegiak arteko lana bultzatzeak hiztegi hobeak sortzea eta LNPrako oinarri indartsuagoak garatzea ekarriko duela.

Beste arazoetako bat da hiztegiek ez dutela inoiz corpusek bezain ongi biltzen hitzen maiztasun erlatiboa, eta ezta hainbat testuingurutan hitzek duten erabilera ere. Beraz, egokiena izango da bi iturriek emango duten informazioa elkarrekin erabiltzea. Baliabideak konbinatzen dituzten lanak modu honetan sailkatzen dira:

- Iturri lexikal egituratuak konbinatzen dituztenak —MRDak, ontologiak, tesauroak, eta abar— (Knight eta Luk, 1994).
- Iturri egituratuak eta ez egituratuak baliatzen dituztenak (Klavans eta Tzoukermann, 1996).

Iturri lexikalak erabiltzeko garaian punturik garrantzitsuena hauen eskuragarritasuna izan ohi da. Honekin oso lotuta *copyright*aren edo autore-eskubideen arazoa egon ohi

da. Zenbaitetan corpusak izan badira, baina horien erabilgarritasuna oso mugatua da autore-eskubideekin arazoak daudelako.

Azaldutako lan hauen guztien emaitza zabalduena ezagutza-base hierarkikoak dira, eta halako erlazioak tesi-lan honen IV. kapituluan landu ditugu. Baina hitzen arteko sare semantikoa sortzeko garaian, ez dugu erlazio hierarkikoa bakarrik kontuan izango, sarrera-hitz eratorrien azterketa egin dugu eta baita definizio barruan dauden hitzen artean dauden erlazioen azterketa ere.

II.2. Eratorpena

Eratorpenaren bitartez sarrera-hitza den hitz eratorria eta bere oinarriaren artean dagoen erlazioa zein den zehazten saiatu gara.

Gaur egun, *WordNet*eko 2.0 bertsioan eratorpena aztertu dute, eta eratorriaren eta oinarriaren artean zein erlazio dagoen eskuz zehaztu dute (Fellbaum eta Miller, 2003). Lan honetan diotenez, halako erlazioak oso garrantzitsuak dira, eta eskuz egitearen garrantzia azpimarratzen dute. Eskuz egitearen arrazoirik garrantzitsuen adieren artean desanbiguatuta behar izatea da. Diotenez, gaur egun eratorria eta oinarria ezagutzen duten tresna onak badaude, baina horien semantika ezagutzeko garaian ez omen daude hain erauzketa zuzenak. Habash-en eta Dorr-en (2003) eta Viegas *et al.*-en (1996) lanak aipatzen dituzte lan garrantzitsu gisa, baina adieren kontuarekin egiten omen dute topo. Fellbaum-ek eta Miller-ek (2003) diotenez, eskuz egitea behar-beharrezkoa da adiera mailako loturak lortu nahi izanez gero. Guk, tesi-lan honetan azalduko dugunez, hitz eratorrien definizioak eta *EuskalWordNet*en bitartez beste hizkuntzetako hitzekin dugun lotura erabiliko ditugu adiera erabakitzeko, eta V.7.2. atalean ikusiko dugunez, errorea %3 da.

Beste hizkuntzetako *WordNet*ak eraikitzeke garaian ere, hitzen arteko erlazioak markatzeko eratorpena erabili berri dute turkieraz eta arabieraz.

Bilgin *et al.*-ek (2004) turkieraren atzizki ez anbiguoekin lan egin dute eta horiek oinarriekin dituzten erlazioak zehaztu dituzte. Erlazio horiek kontzeptu mailakoak direnez, eta *EuroWordNet*ekin lotuta daudenez, lotuta dauden hizkuntza guztientzat dira baliagarri.

Black-ek eta El-Kateb-ek (Black eta El-Kateb, 2004) diotenez, arabierarentzat *WordNeta* sortzen ari dira eskuz. Arabierak eratorpen maila oso handia du, eta artikuluan dioten bezala, oso garrantzitsua da hitz eratorria eta oinarria lotzea. Arabieraren eratorpena, ordea, ez da hitz baten aurrean edo ondoren gertatzen, baizik eta aurrean, tartean eta atzean. Zenbaitetan, hitz baten erroa hiru kontsonante dira, eta hauen tartean jartzen diren bokalen arabera, esanahi aldetik ezberdinak diren hitzak sorten dira. Beraz, hauek egiten dutena da ingeleseko *WordNetean* dauden kontzeptuak —*synsetak*— haiek sortutako *synsetekin* lotu, eta *synset* horietan dauden hitzak bere erroarekin lotuta daude —hala behar duenean—.

Santana *et al.*-ek (2004) espainierako aurrizkiekin egin dute lan. Espainierako 98 aurrizkirekin halako taxonomia eraiki dute, eta hauekin hitz eratorrien eta oinarrien arteko erlazio morfolexikoak lortu dituzte. Lan hau egin ahal izateko hainbat hiztegitatik ateratako lexikoarekin osatu dute corpusa (*DRAE*, *VOX*, *DUE*, *LAROUSSE*, *CALVE SM*, *ESPASA CALPE*, *JULIO CASARES*, *MANUEL ALVAR EZQUERRA*). Hauez gain, adjektiboetatik sortutako 1.240 izen berezi sartu dituzte —*marxista* modukoak—, eta aditz-partizipioetatik eratorritako beste 9.000 adjektibo. Denera, 134.109 forma kanoniko —dagoeneko eratorpena aplikatu zaion hitza, edo aplikatu dakiokena— dituzte. Beraz, nahiz eta corpusa eraikitzeko hiztegiak izan dituzten oinarri, guk lan egingo dugun modutik pixka bat urruntzen dira. Hala ere, azkeneko emaitza berdintsua dela pentsa dezakegu, hau da, eratorriaren eta oinarriaren arteko erlazioaren bat lortzen dutela.

Aurreko lanetan eratorpena eskuz landu dute, baina eratorpenari buruzko lan automatikoak ere egin izan dira.

Light-ek (1996) corpusak erabili ditu oinarri gisa, eta hitz eratorri baten ezaugarriak eratorri horrek duen atzizkiaren ezaugarrien arabera zehazten ditu. Lehenik eratorpen-atzizkiek sortzen dituzten eratorrien eta hauen oinarrien ezaugarriak aztertzen ditu —nolako kategoriak dituzten, batik bat—. Gero corpusera jotzen dute (*Brown* corpusa), eta bertako agerpenak kategoriarekin markatzen dituzte. Erabiliko dituzten eratorpen-atzizkiak ez-anbiguoak dira, eta oinarrian agertzen den kategoriari garrantzi handia ematen diote. Izan ere, kategoria-etiketatzailerak ebazten ez dituen anbiguitasunak alde batera utziko ditu. 18 eratorpen-atzizki eta -aurrizki aztertu ditu. 1,1 milioi hitzeko corpusean aplikatu du

eta 2.535 hitzentzat lortu dute informazioa. Emaitei dagokienez, eratorriari ezaugarri semantiko egokia %76an eman diote, eta oinarriari %82an. Estaldura neurtzeko, *re-* aurrizkia aztertu dute eta %85eko estaldura lortu dute.

Penthedourakis-ek eta Vanderwende-k (1993) eta Peñas *et al.*-ek (2000) oinarri gisa egituratutako baliabide lexikalak erabili dituzte.

Penthedourakis-ek eta Vanderwende-k (1993) eratorriaren eta oinarriaren arteko erlazioak lortu nahi dituzte eta EBL batean sartu. Oinarria zein den %94,4an ezagutzen dute ondo, eta kasu polisemikoen adiera zein den %65ean desanbiguatzen dute. Adierak desanbiguatzekeo garaian, eratorriaren definizioa eta aurretik, beste lan batzuetan (Montemagni eta Vanderwende, 1992; Dolan *et al.*, 1993) definizio horietatik erauzitako erlazio lexiko-semantikoak hartzen dituzte kontuan; oinarria izan daitekeen hitza —lehenengo pausoan hitza bakarrik, eta ez honen adiera—, eta eratorriak dituen erlazio lexiko-semantikoak eratorri horien definizioetatik erauzten dituzte oro har, baina zenbait eskuz sortu dituzte. Ondoko pauso batean, sortu duten algoritmoa erabiliz, oinarriaren adiera zein den zehazten saiatzen dira.

Peñas *et al.*-ek (2000) ere, oinarria eta eratorriaren arteko erlazioa zehazteko garaian, hiztegiko definizioak erabili dituzte. Loturak egiteko sarrera-hitza eta definizioan agertzen den hitz bat erabiltzen dituzte, nahiz eta zenbaitetan zuzen-zuzenean ez izan eratorriaren oinarrian dagoen hitza —*admirable* eta *admiración* lotu dituzte, nahiz eta *admiración* ez izan *admirableren* oinarria—. Lan honetan ez dute adiera desanbiguaziorik egiten, eta oinarrien adiera guztiak ontzat ematen dituzte. Honek %10eko errore-tasa sortzen die. Espainierako hitzen arteko erlazioak lortzeko, hiztegi elebakarrak erabili dituzte, baina horiek ez dituzte *EuroWordNetera* lotu, adiera mailako desanbiguaziorik egiten ez dutelako. Beraz, espainierako loturak lortu dituzte, baina ez dituzte beste hizkuntzekin lotu.

Guk eratorpenari V. kapitulua eskaini diogu. Eta eratorpena euskaraz aski aztertuta dagoenez, atzizki lexikalei buruzko informazioa eta eratorpenari berari buruzko informazioa nagusiki iturri hauetatik hartu dugu: Azkarate, 1990b; Azkue, 1923-1925; Ithurri, 1979; Lafitte, 1944; eta Villasante, 1974.

Eta sarrera-hitza alde batera utziz, definizio barruko hitzen arteko erlazioak lortzeko patroiz lexikaleei eta postposizioei edo preposizioei buruzko lanak ikusiko ditugu.

II.3. Patroiak eta postposizioak

Definizioan dauden hitzen arteko sare semantikoa zehazteko ahaleaginean, Vanderwende-ren (1995) eta Harabagiu-ren (1997) lanak izan ditugu oinarri. Vanderwende-k (1995) patroi jakin batzuei egiten die erreferentzia, eta Harabagiu-k (1997), nahiz eta patroiak bilatu, patroi horietan preposizioen interpretazioei ematen die garrantzia.

Vanderwende-k (1995) dioenez, definizioetan agertzen den sintaxiak garrantzia du, baina, zenbaitetan, sintaxiak baino garrantzi gehiago omen du lexikoak, eta hitz jakin batzuk dituzten patroiak errepikatzen direnez, horiek halako erlazio zehatzak sortzen dituzte definizio batean dauden bi hitzen artean. Ateratzen doazen informazioa EBLEan sartzen dute. Guk halako patroiei patroi lexikal deitu diegu (ikus VI.2. atala).

Harabagiu-ren (1997) helburua, berriz, *WordNet 1.5*en dauden erlazioak aberastea da. *WordNet*eko *synset* bakoitzak kontzeptu horren definizio modukoa dakar, eta definizio moduko honi *glosa* deitzen zaio. Glosetan dauden hitzen arteko erlazioak desanbiguatzen saiatzen da, agertzen diren preposizioak erabiliz. Horretarako hitzen arteko erlazioak zeintzuk diren zehazten du eta hauek dira erabiltzen dituen erlazioak: *agent*, *object*, *instrument*, *beneficiary*, *purpose*, *attribute*, *reason*, *state*, *location*, *theme*, *time*, *manner*. Preposizioei balio batzuk ematen dizkie, eta, horien eta testuinguruaren arabera, hitzen arteko erlazioak desanbiguatzen saiatzen da.

Harabagiu *et al.*-ek (1999) *ExtendedWordNet* proiektua azaltzen dute. Bertan diotenez, *WordNet*eko glosetan dauden hitzak sintaktikoki etiketatzen dituzte, glosen forma logikoa ematen dute, eta glosetan dauden izenen, aditzen, adjektiboen eta adverbioen adierak zehazten dituzte eskuz. Honek *synset*en arteko loturak handitzen ditu, eta kontzeptu batek dituen erlazioak handitzen ditu.

Tesi-lan honen VI. kapituluan EHn topatutako hainbat patroi lexikalen berri eman dugu lehenik (ikus VI.2. atala), eta ondoren, euskarazko postposizioei interpretazioak emanez (ikus VI.5. eta VI.6. atalak), eta beste hizkuntzetako testu paraleloak —beste hizkuntzetako hiztegi elebakarretako pareko definizioak— erabiliz interpretazio horiek desanbiguatzen saiatu gara (ikus VI.7. atala).

Baina gure helburua euskal postposizioak eta beste hizkuntzetako preposizioak lotzea, eta desanbiguazioa interpretazioen ebakiduren arabera egitea denez, arazo batekin egin dugu topo: postposizioen eta preposizioen interpretazioen zerrenda ez dago itxirik, eta ebakidura erabilgarria izan dadin, interpretazio-zerrenda bateratua behar dugu.

Preposizioen interpretazioei dagokienez, Dorren eta Habashen lanak (Dorr, 1993) eta (Dorr eta Habash, 2002), eta Canesson-ena eta St. Dizier-ena (2002) izan ditugu kontuan. Hauek Jackendoff-engandik (1983; 1990) jasotako egitura lexiko-kontzeptualak (*Lexical Conceptual Structure*, LCS) erabiltzen dituzte preposizioen interpretazioak zehazteko garaian —nahiz lanetik lanera baden diferentziarik—. Egitura lexiko-kontzeptual hauetan preposizioei buruzko informazio ugari ematen da —*primitiboa, mota, eremua, rol tematikoa*—. Guk ahalik eta preposizio edo postposizio gehien etiketatzea bilatzen dugu, eta, euskararentzat hasierako lan bat denez, interpretazio-etiketa nahikoa sinplea izatea bilatu dugu, Jensen-ek eta Nilsson-ek (2003) egiten duten modura. Izan ere, preposizioei buruzko hamaika azterketa ezberdin badira (ACL-SIGSEM, 2003), baina askotan oso deskribapen konplexuak izateaz gain, preposizio multzo txikientzat egindako deskribapenak dira. Jensen-ek eta Nilsson-ek (2003) ontologia bat osatu nahi dute, eta danieraren zenbait preposizio aztertzen dituzte haiek izan ditzaketen balizko *rol tematikoekin* etiketatuz. Hauek, guk egiten dugun moduan, kontzeptuen arteko erlazioak zehaztu nahi dituzte, eta erabiltzen dituzten *rolen* zerrendaren oinarria Fillmore-k (1968), Stockwell *et al.*-ek (1973), Somers-ek (1987) eta Madsen *et al.*-ek (2001) erabiltzen dutena da. Hauetan agertzen den zerrenda ondokoa da: *agent, cause, caused-by, patient, part-of, comprise, by means of, source, purpose, location, temporality, material, characterize*.

Gure helburua euskarazko postposizio bakoitzari interpretazio multzoa ematea da, eta lan hori egin du —zuzen-zuzenean izan ez bada ere— Aldezabal-ek (2004) bere tesi lanean. Aldezabalek (2004) euskarazko 100 aditzen argumentu-egiturak aztertu ditu, eta, ondorioz, euskarazko postposizio-atzizki zenbaiti interpretazio semantikoak esleitu dizkie. Guk, euskarazko postposizio-sintagmen interpretazioak desanbiguatzeko beste hizkuntzak erabiliko ditugunez, Aldezabalen (2004) interpretazioak hartu ditugu oinarri gisa eta Dorr-en eta Habash-en lanekin (Dorr, 1993; Dorr eta Habash, 2002; Habash, 2002; Habash eta Dorr, 2003) lotu ditugu.

Dorr-en eta Habash-en azterketetan ingelesezko eta espainierazko preposizio multzo handiari buruzko informazioa dago, eta gure helburua ahal den preposizio eta postposizio gehien interpretatzea denez, gero eta datu gehiago, hainbat eta hobeto.²⁴

Euskarazko postposizio konposatuei buruzko informazioa ez da postposizio-atzizkiei buruzkoa bezain zabala, eta hauek multzokatzeko lana gure esku egon da, nahiz eta Pérez Gaztelu *et al.*-en (2004) lanean hainbat informazio topatu dugun.

Aurrekariekin amaituta, hiztegiaren aurreprozesua eta azterketa estatistikoa azaltzen duen kapituluari ekingo diogu.

²⁴ Erabilitako preposizioen zerrenda Maryland-eko Unibertsitateko CLIP laboratorioran landutakoa izan da, eta bertan egindako egonaldian ezagutzeko aukera izan nuen.

Ama, eskerrik asko

gogo	A1	iz.	Giza buru ahalmenen egoitza edo iturburutzat hartzen den zera.
eta	A1	junt.	Maila edo mota berdineko perpausak edo osagarriak lotzen ditu, gehitze bat, harreman bat edo, adierazten duela.
gorputz	A1	iz.	Izaki bizidunen materiazko osagarria.
ZILBOR HESTE	RA0	iz.	Karena eta umekia lotzen dituen heste modukoa, jaiotzean ebakitzen zaiona.

Gogo eta gorputzaren
zilbor hesteak:
bi kate.

Bi kate,
biak ebaki beharrezkoak:
bat gorputzaren bizitzeko,
bestea gogoaren askatzeko.

Adizu, ama badakizu
sortze berean
zuri gorputzez lotzen ninduen zilbor-hestea
sendagileak nola eten zuen.
Lehenengo eten beharra izan zen:
bizitzaren bizitzeko lehen legea.

Hautzaroan titia eman zenidan,
mutil-zaroan eskoletara bidali
bizitzarako armak hartzearen.

Dena eskertzen dizut
duen balore guztian,
nik ahal dutan neurrian
bainan, gaztaroan amatxo maite!
Ohar zaitezte benetan maite nauzun ama
izan nahi baduzu
eta nik zu maitatzea,
ni naizena
nik izatea nahi dutana izatera
utzi behar nauzula,
hau baita bide bakarra
biok alkar sanoki maitatzeko,
biok alkar osoki eta betikoz maitatzeko,
zuk zure nortasunaz,
nik nereaz,
zuk zure nortasunaz,
nik nereaz.

Ama!
Eten dezagun
lehen gorputzarena bezela
orain, gogoaren zilbor-hestea.
Ama!
Gogoaren zilbor-hestea.

(J. Artze)

III. Hiztegiaren aurreprozesua, Murritzapen Gramatikak eta azterketa estatistikoa

III.1. Sarrera

Kapitulu honetan aipatuko ditugu, alde batetik, iturri-hiztegia esku artean hartzen dugunetik bertatik informazioa erauzten hasi arte egin dugun prozesu osoa —aurreprozesua, alegia—; bestetik, erauzketa egiteko garaian erabili dugun gramatika zein den eta zein ezaugarri dituen; eta, azkenik, hiztegian agertzen diren hainbat elementuren azterketa estatistikoa.

Aurreprozesuari dagokionez, kontuan izan behar da hasiera batean *Euskal Hiztegiak* (EH) paperezko formatua zuela eta pertsonen erabilerarako pentsatutako produktua zela. Bazen makinan irakurtzeko moduko bertsiorik, baina bertan zegoen informazioa egituratu gabe zegoen, erabilitako formatuak —*Microsoft Word*— ez duelako halakorik egiteko aukerarik uzten. Beraz, eta makinentzat erabilgarri izan dadin, egin beharreko lehenengo lana, makinen bitartez hiztegia landu ahal izateko formatu egituratu batera aldatzea izan da (Arriola, 2000). Aukeratutako formatua TEI (*Text Encoding Initiative*) (Sperberg-McQueen eta Burnard, 1994) ekimenak gomendatutakoaren arabera kodetutakoa izan da. TEI ekimenaren gidalerroak erabiliz, hiztegi osoaren egitura oso garbi izango dugu, eta makinek unean-unean behar diren eremuak berehala ezagutzeko aukera dute. Guretzat oso interesgarria da aukera hau; izan ere, ikusiko dugun bezala, sarrera-hitz bakoitzak hainbat informazio izango du, baina guk erabat beharrezko ditugun eremuak besterik ez ditugu hartuko.

Ikusiko dugunez, eremu hauek *sarrera-hitza*, *kategoria*, *adiera-zenbakia*²⁵ eta *definizioa* dira.

Hiztegi osoa kategorien arabera banatu dugu, eta hauen arabera multzoak egin ditugu: *izenak*, *aditzak*, *adjektiboak*, *izenak* eta *adjektiboak* direnak, *azpisarrerak*, eta *bestelakoak*. Sarrera-hitz bakoitzaren definizioaren analisi morfologikoa egin dugu, eta baita sarrera-hitzen beraren analisi morfologikoa ere.

Sarrera-hitz zenbaiten azpian *azpisarrerak* ditugu. Eta, nahiz eta hauen artean agertzen diren elementuak aukeratutako kategorietakoak izan, bereziak direnez beste multzo batean jarri ditugu. Izan ere, askotan ez dute kategoriarik jasotzen, eta modu berezian aztertu behar izan ditugu.

Aurreprozesu osoa azaldu ondoren, informazioa erauzteko *Murritzapen Gramatika* (Karlsson *et al.*, 1995; Tapanainen, 1996) erabili dugunez, honi buruzko hainbat ezaugarri eman ditugu. Gramatika honek testuingurua hartzen duenez kontuan, oso egokia da definizioetan genusak, sinonimoak eta erlatore bereziak ezagutzeko; helburu den elementua aurretik eta atzetik dituen elementuen arabera markatuko dugu.

Hiztegian zein hitz ditugun, nolako definizio-patroiak ditugun, eta hiztegiko elementuen kopuruak zeintzuk diren jakin ahal izateko azterketa estatistiko zenbait oso garrantzitsuak dira. Atal honetan agertuko dira hiztegiak dituen hitz-formen kopurua —kategorien arabera banatuta—, hiztegian agertzen diren hitzen maiztasuna eta *token*-segida nagusiak zeintzuk diren.

Kapitulu honen hezurdura honokoa duzue: III.2. atalean *Euskal Hiztegiaren* egitura orokorra ikusiko dugu; III.3. atalean hiztegia makinek irakurtzeko modukoa bihurtuko dugu; III.4. atalean lanerako erabili dugun bertsioa zein den aurkeztuko dugu; III.5.ean definizioen analisi morfologikoaren berri emango dugu; III.6.ean sarrera-hitz eratorrien analisiaren berri emango dugu; III.7.ean erlazioak erauzteko erabili dugun gramatikaren aurkezpena egingo dugu; III.8.ean hiztegiaren analisi estatistikoa egiterakoan agertzen diren elementuen maiztasunak paratu ditugu; eta, amaitzeko, III.9. atalean, kapitulu honen ondorioak bildu ditugu.

²⁵ Adiera-zenbakiaren eremu berean sartuko dugu homografo-zenbakia, nahiz eta paperezko bertsioan bi hauek ez elkarrekin azaldu.

III.2. EHren egitura orokorra

EH (Sarasola, 1996) euskara batuaren biltegi gaurkotua da. Euskarazko hiztegi elebakarra da, orokorra, tradizio literarioan oinarritutakoa, eta gaur egun lexikoa zertan den jakiteko erreferentzia nagusia. Hiztegi honen eskuz errebisatutako TEI bertsioak 33.124 sarrera ditu; adierak, berriz, 53.961 dira —hemen adierak eta ñabardurak daude sartuta—. Azpisarrerak, berriz, 3.209 dira. Kopuruen azterketa zehatzagoa III.8. atalean egingo dugu.

Hona hemen sarrera-hitz baten adibidea paperezko hiztegian agertzen den bezala:

irekidura. *iz.* (1977; *idekidura* 1985). Batez ere ipar.
Pareta, horma... batean gertatzen den gunek irekia. *ik.*
irekigune. *Berriazko irekidura bat teilen artean zuen*
gelak. Etxeko irekidura guztiak ongi hetsiak izanagatik.

Sarrerak honako eremuak ditu (Arriola *et al.*, 1996)

- Sarrera: letra lodiz agertzen da: *irekidura*.
- Kategoria: eremu honetan kategoria gramatikalak nahiz ezaugarri morfologikoak agertzen dira. Adibidean, *iz.* izen kategoriaren laburdura dugu.
- Data: sarreraren agerpen-data hiztegi edo testu literario batean. Hiztegi batean edo aipamen gisa agertzen denean, datak *izartxo* (*) darama; testu batean agertutako lehen aldia adierazten badu, aldiz, ez. Hautatutako adibidean, testu batean agertutako lehen aldia adierazten du: 1977.
- Aldaera(k) eta data: sarrera-hitzak dituen aldaerak ageri dira zenbaitetan —agertzen direnetan letra etzanez—, eta hauen agerpen-data. Adibidean, *idekidura*, 1985: *idekidura* aldaera lehen aldiz 1985ean agertu dela adierazten du.
- Laburdura(k): erabilerari dagozkionak, tematikoak, hizkuntz mailakoak, euskalki-argibideak, eta abar. Adibideko batez ere ipar. laburdurak Ipar Euskal Herrian erabiltzen den hitza dela adierazten du.
- Definizioa: adiera bakarreko sarreretan definizio-testuak osatzen du eremu hau. Adibidean, Pareta, horma... batean gertatzen den gunek irekia. Adiera bat baino gehiago dagoenean, aldiz, bereizgarri desberdinak erabili ohi dira definizioak banatzeko, edo zenbait ñabardura islatzeko. Hiztegi honetan

hainbat ikur erabili ohi dituzte adiera multzoak, adierak eta ñabardurak bereizi ahal izateko.

- Erlazioa(k): eremu honetan *sinonimia* edo *antonimia* bezalako adiera-erlazioak adierazteko kode lexikografiko zenbait daude, hala nola, *ik.*, eta *ant.* Adibidean, *ik.* irekigune dugu.

Oro har, *ikus* honek adieraziko digu sarrera-hitza eta zehatz-mehatz adierazten zaiguna erlazonatuta daudela. Guk, halako marka dugunean, pentsatuko dugu *sinonimoak* direla,²⁶ ia gehienetan hala delako, nahiz eta gerta daitekeen *ik.* ikur horrek sinonimo peto ez den hitz bati egitea erreferentzia. Halako kasuak ditugu *nabar* eta *golde* sarrerak.

2 nabar. *iz.* (...) || Golde mota batzuetan, lurra ebakitzen duen hortza, lurra iraultzen duen golde muturraren aurrean kokatzen dena; beste batzuetan, lurra iraultzen duen hortza. *ik.* **golde.** (...)

golde. *iz.* (*1562, 1571) **1.** Lurra lantzeko lanabes horzduna, idi, zama abere edo traktore batek herrestan eramaten duena eta laborariak atzetik gidatzen duena. *ik.* **2 nabar.** (...)

- Adibidea(k): testu etzana erabili da hauek bereizteko, eta definizioaren ondoren agertu ohi dira: Berinazko irekidura bat teilen artean zuen gelak. (...)
- Azpisarrera(k): maiuskulaz daude, eta sarreratzat har daitezke; hala ere, kontuan izan behar da beti beste sarrera baten barruan daudela. Katēgoria zehatzik paratzen ez zaionean pentsatu dugu sarrera-hitzaren katēgoria bera izango duela.

Goiko adibidean ez dugu azpisarrerarik, baina hona *arbola* sarreraren azpisarrera paperezko bertsioan agertzen den moduan bere agerpen-data eta definizioarekin.

—ARBOLA IPURDI. (1966). Epaitondoa.

Eremu hauek guztiak paperezko bertsioan inplizitu daude, eta begiari deigarri zaizkion metodoen bitartez —letra mota ezberdinak, kasu— ezberdintzen dira. Kontua da, paperean inplizitu dagoen hori, makinari esplizitu egin behar zaiola unean-unean jakin dezan ikur bakoitza zein eremuri dagokion. Horretarako, aurretik

²⁶ *Petit Robert* (Robert, 1990) hiztegian ere erlazonatutako hitzak modu honetan markatzen dituzte.

aipatu ditugun TEI gidalerroak erabili ditugu, eta, labur bada ere, horiek azalduko ditugu.

III.3. TEI ekimena

TEI (*Text Encoding Initiative*) testuak kodetzeko modu berezia gomendatzen duen ekimena da (Sperberg-McQueen eta Burnard, 1994). Ekimen honek erabiltzen duen markatze-sistema SGML (*Standard Generalised Markup Language*) da, eta honoko abantailak ditu: argitasuna, sinpletasuna, zehaztasuna, eta nazioarteko estandar gisa onartua izatea.

Ekimen honen gidalerroak testu-corpusak markatzeko erabiltzen dira, eta *Association for Computational Linguistics*, *Association for Literary and Linguistic Computing* eta *Association for Computers and the Humanities* elkarteek bultzatu dute. Helburua, inplementazio estandarrak lortzea da, makinak irakur ditzakeen testuen arteko elkartrukatzea ahalbidetzeko. TEIren ekarpena da arau zehatz batzuk ezartzea testuak markatzeko estandar hau nola erabili erakusteko.

Hona irekidura sarrera TEIz kodetuta:

```
<entry>
  <form><orth>irekidura27</orth></form>
  <GramGrp><pos>iz.</pos></GramGrp>
  <usg type="time">1977</usg>
  <form type="variant"><orth>idekidura</orth><usg
    type="time">1985</usg></form>
  <usg type="geo">Batez ere Ipar.</usg>
  <def>Pareta, horma... batean gertatzen den gune
irekia.</def>
  <xr type="syn"><lbl>ik.</lbl><ref>irekigune</ref></xr>
  <eg><q>Berinzako irekidura bat teilen artean zuen gelak.
Etxeko irekidura guztiak ongi hetsiak
izanagatik.</q></eg>
  <sense n="n1">
    <eg><q>Ahoko irekidura.</q></eg>
  </sense>
</entry>
```

Elementu edo eremu baten hasiera etiketatzeko, angeludun parentesi baten barruan jarritako marka erabiltzen da:

²⁷ Testu diren elementuak beltzez jarri ditugu etiketa direnetatik errazago bereizteko, baina benetako fitxategietan ez dira hala agertuko.

<entry>: sarrera eremua hor hasten da

Amaiera markatzeko, berriz, hasierako marka bera jartzen da, baina aurretik *barra* duela:

</entry>: sarrera eremua hor amaitzen da

Hona irekidura sarreraren azalpena:

- Sarrera-hitza (*orth*) irekidura da.
- Katetoria (*pos*) izena da.
- Erabilita agertzen den lehen aldiko data (*usg type*=“*time*”) 1977 da.
- Aldaera gisa (*form* “*type*=“*variant*”) idekidura agertzen da 1985 urtean (*usg type*=“*time*”)
- Erabilerari dagokionez (*usg type*=“*geo*”), adierazten zaigu batez ere Iparraldean erabiltzen dela.
- Definizioztat (*def*) hartuko dugun deskribapena honokoa dugu: Pareta, horma... batean gertatzen den gune irekia.
- irekigune hitzarekin zerikusia (*ref*) duela adierazten zaigu.
- Ondoren adibidea agertzen da (*eg*): Berinazko irekidura bat teilen artean zuen gelak. Etxeko irekidura guztiak ongi hetsiak izanagatik.
- Azkenik adierazten da baduela ñabardura bat (*sense n*=“*nl*”), eta horren adibidea (*eg*) dugu: ahoko irekidura. Kasu honetan, ñabardura honek ez du definiziorik, eta guk ezingo dugu bertatik informaziorik erauzi.

Hasiera batean TEIra bihurtze-prozesu hau automatikoki egin zen (Arriola *et al.*, 1997), eta definizioaren eremu garrantzitsuenak —*sarrera*, *kategoria*, *definizioa*, *adibidea*, eta *azpisarrera*— ondo ezagutzen eta markatzen ziren %97ko doitasunarekin. Hala ere, eta hiztegiaren markaketa zehatzagoa lortzeko, egindako lan automatikoa oinarri hartuta, eskuzko zuzenketa egin da (Arregi *et al.*, 2002).

Hiztegiko elementu guztiak TEIren gidalerroei esker markatuta izan ondoren, behar ditugun eremuak hautatu eta lanerako bertsioa sortu dugu.

III.4. Hiztegiaren bertsio tabulatua

Atal hau bitan banatu dugu: batetik sarrera-hitz batek dituen eremuetatik zeintzuk aztertuko ditugun ikusiko dugu, eta, bestetik, lanerako erabilitako fitxategiei zein egokitzapen egin dizkiegun ikusiko dugu.

III.4.1. Beharrezko zaizkigun eremuak

Egin nahi dugun lanerako ez ditugu behar hiztegiko sarrera batean agertzen diren eremu guztiak. Lanerako interesgarri zaizkigun eremuak lau dira: *sarrera-hitza*, *adiera-ikurra*, *kategoria*, eta *definizioa*.²⁸

irekidura A0 iz. Pareta, horma... batean gertatzen den
gune irekia.

Oraintxe komentatu ditugun lau eremu horiek ez dira maila berekoak. Bi *eremu irekiak* dira, eta, beste biak *eremu itxiak*; hau da, sarrera eta definizio eremuetan ia edozer gauza ager daiteke, baina kategoria eta adiera-ikurra eremuetan ager daitezkeen elementuen kopurua itxia da.

Kategoriari dagokionez, 4. taulan ditugu hiztegian topatu ditugunak.

Laburdura	Adierazia	Kopurua	Laburdura	Adierazia	Kopurua
ad.	aditza	6.287	izord.	izenordaina	61
adlag.	adizlaguna	2.924	junt.	juntagailua	70
banatz.	banatzailea	1	lok.	lokailua	15
determ.	determinatzailea	6	mugatz.	mugatzailea	2
erak.	erakuslea	20	ord.	ordinala	34
interj.	interjekzioa	42	posp.	postposizioa	8
iz.	izena	22.603	pred.	predikatua	20
izlag.	izenlaguna	468	zenbtz.	zenbatzailea	151
izond.	izenondoa	4.135			

4. taula. Hiztegian ditugun kategorien laburdurak, haien adieraziak eta kopuruak.²⁹

Gure lanerako kontuan izango ditugunak: izenak, aditzak eta adjektiboak —izenlagunak eta izenondoak— dira.

Sarrera-hitz baten kategoria-eremuan, zenbaitetan, kategoria bat baino gehiago agertzen da. 5. taulan dago konbinatuta agertzen diren kategoria horien zerrenda.

²⁸ Modu honetan emango ditugu tesi-txostenean zehar agertuko diren hiztegiko sarreraren adibideak.
sarrera-hitza adiera-ikurra kategoria definizioa

²⁹ Kategoriarik ez duten hitzak 520 dira.

Laburdura	Kopurua	Laburdura	Kopurua
adlag.#iz.	39	iz.#izlag.	195
adlag.#iz.#izond.	14	iz.#izond.	869
adlag.#izlag.	4	izord.#adlag.#junt.	4
adlag.#izond.	19	ord.#iz.	13
adlag.#lok.	2	pred.#iz.	5
adlag.#part.	1	pred.#izond.	7
adlag.#pred.	9	zenbtz.#izlag.	1
adlag.#zenbtz.	1		

5. taula. Hiztegian agertzen diren kategoria-konbinazioak eta kopurua.

Kategoria-bikote eta –hirukote hauek ikusita, *izen* direnak eta *adjektibo* ez direnak *izenen* artean bildu ditugu; *adjektibo* direnak eta *izen* ez direnak, *adjektiboen* artean; eta *izen* eta *adjektibo* direnekin multzo berezia osatu dugu. 5. taulan agertzen direnetatik, *izen* edo *adjektibo* marka ez dutenak —adlag.#lok., adibidez— ez ditugu aztertuko, eta guztiak *bestelako* gisa sailkatu ditugu.

Adiera eremuan, berriz, honokoak topa ditzakegu: *letrak* eta *zenbakiak*.³⁰ Letrei dagokienez, honokoak izan ditzakegu: *H* (homografo-adierazlea), *M* (adiera multzoa), *A* (adiera-adierazlea), *N* (ñabardura-adierazlea), *R* (azpisarrera-adierazlea), eta *X* (azpisarreraren adieraren azpiadieren adierazlea). Hauetako hiruk (*H*, *A*, eta *N*) ondoren zenbaki bat dute, eta hori da homografo-, adiera-, ñabardura-zenbakia. Multzoa adierazten duen *Maren* ondoren eta azpisarreraren adieren azpiadiera diren *Xen* ondoren dagoen letra minuskulak adieraziko du zein multzori dagokion eta azpisarrera horren adieraren zein azpiadiera den.

abendu	H1A1	iz.	Urtearen hamabigarren eta azken hila, hogeita hamaika egun dituen.
abendu	H1A2	iz.	Eliz urtean, Eguberri aurreko lau asteek osatzen duten denbora.
abendu	H2A0	iz.	Mirua.

Hemen dugu *abendu* sarreraren adibidea. Hitz honek bi homografo (*H*) ditu, *abendu-1* eta *abendu-2*. Lehenengoak bi adiera (*A1* eta *A2*) ditu, *hila* eta *denbora*; eta bigarrenak adiera bakarra (*A0*), *mirua* (*hegazia*). Ikus daitekeenez, homografo batek adiera bat baino gehiago izan ditzake, baina baita adiera multzo bat baino gehiago ere.

³⁰ Gure bertsiotabulatuan agertuko diren ikurrak dira hauek. Paperezko bertsiotan beste modu batean agertzen dira.

zain	H1Ma	iz.	Sustraia, erroa, bereziki azalekoa.
zain	H1MbA1	iz.	Odola dabilen hodia.
zain	H1MbA1N1	iz.	Odola ehunetatik bihotzera itzularazten duen hodia.
zain	H1MbA2	iz.	Zuntz sorta nakar kolorea, giharra hezurari itsasteko egitekoa duena.
zain	H1MbA2N1	iz.	Nerbioa.
zain	H1MbA3	iz.	Zerbaiten azalean gertatzen den beste koloreko bateko zerrenda mehea.
zain	H2A0	iz.	Zaindaria.

zain hitzak bi homografo (*H*) ditu. Bigarren homografoa (*H2*) zaindaria (*santua*, *Ama Birjina* edo *pertsona*)³¹ hitzaren *sinonimoa* da. Lehenengo homografoak (*H1*) bi adiera multzo (*M*) ditu: lehenak (*Ma*) sustrai eta erro ditu sinonimotzat, eta bigarrenak (*Mb*) hiru adiera (*A*) ditu: lehen adiera (*A1*) odola dabilen hodia da, eta honen ñabardura (*A1N1*) izango da odola ehunetatik bihotzera itzularazten duena; bigarren adiera (*A2*) zuntz sorta dugu, eta honek ere ñabardura bat (*A2N1*) du, nerbioa; hirugarren adiera (*A3*) zerrenda mehea da.

zain hitzaren adibidean ikus dezakegunez, adiera multzoek hainbat adiera izan ditzakete, eta adierek, era berean, hainbat ñabardura.

Hala ere, kasu gutxi batzutan, adiera multzo batek izan ditzake ñabardura (*N*) diferenteak. Oro har, adiera multzo baten ondoren adiera (*A*) agertzea da ohikoena.

bero	Ma	iz.	Materiaren egoera, tenperatura gora batean gauzatzen dena; egoera horretan dagoen zerbait ukitzeak eragiten duen sentipena.
bero	MaN1	iz.	Beste energia moten eraldaketatik datorren energia, bere gehikuntzak tenperatura igotzea dakarrena, eta aldaketa ezaugarriak eragiten dituen gorputzetan.
bero	MaN2	iz.	Arreske dagoen ugaztun emearen egoera.
bero	Mb	izond.	Gorputzarena baino tenperatura garaiagoa duena; bero sentipena eragiten duena.
bero	MbN1	izond.	Gartsua, gauzak gogoz hartzen edo egiten dituen.
bero	MbN2	izond.	
bero	MbN3	izond.	Pertsonez mintzatuz, erraz sutzen dena.

³¹ Ez dakigunez zaingariren zein adiera den zain horren sinonimoa, eta zaingarik bi adiera dituenaz, lehena *santua* edo *Ama Birjinari* dagokiona, eta bigarrena *pertsonari* dagokiona, pentsatuko dugu bi adierei egiten diela erreferentzia.

beroren kasuan, ikus dezakegunez, bi adiera-multzo (*M*) ditu, bata (*Ma*) *izen* kategoria duenari dagokiona (egoera, sentipena), eta bigarrena (*Mb*) *adjektibo* kategoria duenari dagokiona. Lehenak (*Ma*) bi ñabardura (*N1* eta *N2*) ditu, energia moten eraldaketatik datorren energia eta arreske dagoen ugaztun emearen egoera; bigarrenak (*Mb*), hiru ñabardura (*N1*, *N2* eta *N3*) ditu, gartsua, gauzak gogoz hartzen edo egiten dituen (*N1*) eta pertsonen mintzatzuz, erraz sutzen dena (*N3*). *N2*k ez du definiziorik, eta gure azterketan definizioak aztertzen ditugunez, ez da gure fitxategian agertuko.

Adieren kasuan, ondoren zer bait agertzekotan ohikoena izango da ñabardura (*N*) agertzea, zain hitzaren kasuan ikusi dugun bezala, baina, kasu urri batzutan, azpisarreraren ñabardura-markak (*X*) agertzen dira —normalean azpisarreraren azpiadieren adierazleak izan ohi dira azpiñabardura hauek—.

ituntze	A1Xa	iz.	Aholkua.
ituntze	A1Xb	iz.	Elkar aditzea.

Adiera bakarreko hitzetan *A0* marka gehitzen zaie, jakiteko adiera bakarrekoak direla. Adiera bat baino gehiago dutenetan *A1*, *A2*, eta abar agertuko dira.

ziza	A0	iz.	Perretxikoa; bereziki, <i>Tricholoma</i> , <i>Lyophyllum</i> eta <i>Rhodapaxillus</i> jeneroetakoa.
zurruitero	A0	izond.	Edalea, zurrut zalea.

Kasu honetan, bai ziza *izena* eta bai zurruitero *adjektiboa* adiera bakarreko hitzak dira.

Azpisarrerak (*R*) ere badira hiztegian, eta hauek hainbat adiera izan ditzakete. Adiera hauek, *X* batekin markatzen dira eta honek adierazten du azpisarrera horren adiera dugula.

AUKERAN	RA0Xa	iz.	Une edo era egokian, gogara, erara.
AUKERAN	RA0Xb	iz.	Hautatzeko eran, ugari.
AUKERAN	RA0Xc	iz.	Aukera badaiteke; (ohi edo espero denarekin) alderatuz.

aukeran azpisarrerak hiru adiera (*X*) ditu: (*Xa*) une edo era egokian, gogara, erara; (*Xb*) hautatzeko eran, ugari; eta (*Xc*) aukera badaitekeren pareko dena.

Azpisarreraren kategoria ez da zehatz-mehatz hiztegian zehazten eta sarrera-hitzaren kategoria heredatzen dutela pentsatu dugu. Baina heredatutako kategoria hori ez da beti kategoria zuzena izango —aurreko kasuan, adibidez, aukeran azpisarrerak *iz.* kategoria heredatu du, baina ez da zuzena—. Beste kasu askotan, azpisarrera hitz anitzeko aditza izango da —hauek izen kategoriako hitzen barruan agertzen dira maiz—, eta beste askotan adizlagun bat. Dagoeneko komentatu dugun bezala, hauekin multzo berezia egingo dugu, aparte aztertu eta definizio-patroiaren bitartez zein kategoriatakoak diren ezagutzen saiatuko gara.

III.4.2. Fitxategien egokitzapena

Lanerako zein eremu erabiliko ditugun azaldu ondoren, fitxategiari beste aldaketa bat egin behar zaio gero aplikatuko zaizkion tresna guztiak modu egokian lan egin dezaten. Lanerako fitxategietan zehatz-mehatz agertuko da sarrera-hitza zein den, adiera-ikurra zein duen eta zein kategoria duen. Datu hauen ondoren definizioa izango dugu:

```
@@lema            irekidura
@@Adiera_string   A0
@@Kategoria       iz.
Pareta, horma... batean gertatzen den gune irekia.
```

Analizatzaile morfologikoak analisiak egiteko garaian arazorik izan ez dezan, hiztegi-sarrera bakoitzean ditugun *lema*, *adiera-ikurra* —adibidean, *Adiera_string*—, eta *kategoria* eremuei aurretik ikur berezi batzuk jarriko dizkiegu: @@. Honela, @@ duten lerroei ez zaizkie gero definizio-eremuari aplikatu beharreko analisi guztiak aplikatuko. Hala ere, tesi-txosten honetan sarreraren adibideak orain arte bezala —@@rik gabe, alegia— ematen jarraituko dugu:

```
irekidura   A0      iz. Pareta, horma... batean gertatzen den
gune irekia.
```

Ikur berezi hauek jartzen ditugun prozesu berean sarrera-hitz guztiak kategoriaren arabera bereiziko ditugu. Sarrera-hitz guztiak duten kategoriaren arabera banatzea garrantzitsua da IV. kapituluan ikusiko dugun bezala kategoria bakoitzean hitzak definitzeko modu ezberdinak erabiltzen direlako. Alde batetik *izenak* jarriko ditugu, bestetik *aditzak* eta bestetik *adjektiboak*. Hiru kategoria hauek izango ditugu aztergai. Baina dagoeneko azaldu dugun bezala, horiezaz gain beste bi multzo sortuko ditugu:

- Hiztegian *izen* eta *adjektibo* gisa agertzen direnak. Hauen definizio mota aztertu nahi dugu, eta erabaki —ahal dela, behintzat— zein kategoriatako definizioa eman dion lexikografoak.
- *Azpisarrera* direnak: hauek askotan —aurretik ikusi dugun bezala— kategorია okerra heredatzen dute, eta euren kategorია zuzena zein den definizioen bitartez ezagutzen saiatuko gara —izen, adjektibo edo aditz direnentzat, behintzat—.

Bestelako sarrerei dagokienez —*adizlagunak*, *banatzaileak*, *determinatzaileak*, *erakusleak*, *interjekzioak*, *izenordainak*, *juntagailuak*, *lokailuak*, *mugatzaileak*, *ordinalak*, *postposizioak*, *predikatuak*, eta *zenbatzaileak*— multzo berean bilduko ditugu —*bestelakoak*—, baina oraingoz ez dugu hauekin lanik egingo. Hauen definizioak nahikoa bereziak dira, eta askotan metalengoaia erabiltzen da.

Behin hiztegian definizioa duten sarrera, adiera, eta azpisarrera guztiak modu berezi honetan antolatu ondoren, eta multzotan bereizi ondoren, definizioen analisi morfologikoa egiteari ekingo diogu.

III.5. Definizioen analisi morfologikoa

Behin aurreko guztia egin ondoren, EHn dauden erlazio lexiko-semantikoak erauzi ahal izateko, definizioetan agertzen diren unitateen tokenizazioa,³² analisi morfosintaktikoa eta etiketatzea egin behar dira.

Analisi hau egin ahal izateko, Ixa Taldeak garatutako *Morpheus* analizatzaile morfologikoa (Urkia, 1997; Alegria *et al.*, 1996) erabili dugu. Analizatzaile morfologiko honen oinarrian *Euskararen Datu-Base Lexikala* (EDBL)³³ dago (Aduriz *et al.*, 1998; Aldezabal *et al.*, 2001).

Hona irekidura sarreraren definizioaren analisi morfologikoa, baina adibidea ikusi aurretik hona analizatzaile morfologikoak ematen duen emaitzaren azalpen txiki bat. irekidura sarrera-hitzaren definizioa Pareta, horma... batean gertatzen den gure irekia da. Analizatzaileak definizio hori hartu eta analizatu egiten du. Paretari dagokionez, hiru analisi ematen ditu; hiruak *izen arruntak* dira

³² *Tokenizatzea* definizioetan agertzen diren elementu guztiak markatzea eta bereiztea da. Definizioetan ez ditugu hitz-formak bakarrik izango, puntuazio-markak, parentesiak eta halakoak ere badaude, eta horiek guztiak bereizi behar dira.

³³ Egunez egun handitzen ari den datu-basea da eta tesi honen garapenean 3.3.6 berstsioa erabili dugu.

(IZE ARR), amaierako *a berezkoa* dute (AORG), eta *hasiera maiuskula* da (HAS_MAI). Baina lehenak *kasu gramatikal absolutiboa* (DEK ABS) du eta *mugagabea* (MG) dago; bigarrenak ere *kasu gramatikal absolutiboa* du (DEK ABS) eta *singularra eta mugatua* da (NUMS MUGM); hirugarrena *ez dago deklinatuta* (DEK ZERO).

```

/<@lema irekidura>/<ID>/
/<@Adiera_string A0>/<ID>/
/<@Kategoria iz.>/<ID>/
"<Pareta>"
    "pareta" IZE ARR DEK ABS MG AORG HAS_MAI
    "pareta" IZE ARR DEK ABS NUMS MUGM AORG HAS_MAI
    "pareta" IZE ARR ZERO AORG HAS_MAI
"<,>"
    PUNT_KOMA
"<horma>"
    "horma" IZE ARR DEK ABS MG AORG
    "horma" IZE ARR DEK ABS NUMS MUGM AORG
    "horma" IZE ARR ZERO AORG
    "hormatu" ADI SIN AMM ADOIN NOTDEK
"<...>"
    PUNT_HIRU
"<batean>"
    "bat" DET DZH NUMS DEK NUMS MUGM DEK INE
    "bat" IZE ARR RARE+ DEK NUMS MUGM DEK INE
    "bate" IZE ARR DEK NUMS MUGM DEK INE
    "batean" ADB ALGARR
"<gertatzen>"
    "gertatu" ADI SIN AMM ADIZE DEK INE
    "gertatu" ADI SIN AMM ADOIN ASP EZBU NOTDEK
"<den>"
    "izan" ADL A1 NOR NR_HU ERL MEN ERLT
    "izan" ADL A1 NOR NR_HU ERL MEN MOS
    "izan" ADL A1 NOR NR_HU ERL MEN ZHG
    "izan" ADT A1 NOR NR_HU ERL MEN ERLT
    "izan" ADT A1 NOR NR_HU ERL MEN MOS
    "izan" ADT A1 NOR NR_HU ERL MEN ZHG
"<gune>"
    "gune" IZE ARR DEK ABS MG
    "gune" IZE ARR ZERO
"<irekia>"
    "ireki" ADI SIN AMM PART DEK ABS NUMS MUGM
"<$.>"

```

Ikusi dugun analisia anbigua da, hau da, hitz batek izan ditzakeen balizko interpretazio guztiak daude. Adibidez, batean elementuak hainbat lema izan ditzake: bat, bate, eta batean. bat *izena* eta *determinatzailea* dugu, eta analizatutako hitza hauen forma inesiboa da; bate *izena* dugu, eta analizatutako hitza *izen* honen forma inesiboa da; eta batean, bere horretan, *adberbioa* da. Analisiak desanbiguatuta

baleude, determinatzailearen irakurketa bakarrik izango genuke, baina ondoren ikusiko dugun arrazoiengatik analisi anbiguoaren gainean lan egiten dugu:

- Azterkizuna definizioak direnez, kontuan izan behar dugu ez direla esaldi osoak izaten —izen eta adjektiboen definizioak batik bat—. Esaldi osoak sortzeko ahalegina egin genuen da partikula definizio bukaeraren aurretik txertatuz, eta ondoren desanbiguatuz, baina erauzketa-emaitez aurretik lortzen genituenak baino kaskarragoak ziren.
- Desanbiguazio-errore batek guk markatu nahi dugun elementua ez harrapatzea ekar dezake.
- Ez desanbiguatzeak ez dakar arazo handirik; izan ere, guk sortutako erregelek badakite zer nahi duten eta elementu horren bila joango dira. Guk *izen* kategoria duen zerbait markatzeko eskatuz gero, berdin da markatu beharreko lemak bi kategoria izatea —izen eta adjektibo, kasu—, guk agindu dioguna besterik ez baitu markatuko —kasu honetan izena—. Honen azalpen zehatzagoa erabiltzen dugun gramatika azaltzen dugunean (III.7. atala) emango dugu.

Beraz, analisi morfologiko anbigua erabiliko dugu gure lanerako, eta hori dugu *Morpheus*en emaitza. Baina, nahiz eta *Morpheus* edozein testutan aplikatzeko prestatuta egon, testu berezietekin arazo zenbait izan ditzake. Hiztegiko definizioetan ezagutzen ez dituen hainbat laburdura lexikografiko daude, eta horiek akatsik sor ez dezaten, laburdura hauei trataera berezia eman diegu. Laburdura horiek testu-hitz bihurtu ditugu programa baten bitartez. Honela, definizioan agertzen direnean *berezki*, *hed.*, *at. m.* modukoak guk programaren bitartez haien adieraziak idatziko ditugu: *bereziki*, *hedaduraz*, *atomo masa*.³⁴

6. taulan definizioetan agertzen diren laburduren zerrenda eta haien adierazia jarri dugu.

³⁴ Laburdurak bere horretan utziz gero, analizatzaile morfologikoak, oraingo egoeran, bi arazo larri izango lituzke: alde batetik laburdura ez ezagutzea, eta, bestetik, eta hau larriago litzateke, *puntu* bat dagoenez, definizio amaiera dela markatzea.

laburdura	adierazia	laburdura	adierazia
ad.	aditza	G.	Gutxi
ad.-aren	aditzaren	hed.	hedaduraz
at. m.	atomo masa	Ik.	Ikus
at. z.	atomo zenbakia	Ipar.	Iparraldean
berezk.	bereziki	iz.-ak	izenak
er.	erabilia	km.-ren	kilometroren
etab.	eta abar	km.-rena	kilometrorena
etab.-en	eta abarren	mulk.	multzokaria
etab.-etako	eta abarretako	pl.	pluralean
etab.-etan,	eta abarretan,	Sing.	Singularrean
etab.-ez	eta abarrez		

6. taula. Hiztegiko definizioetan agertzen diren laburdurak eta haien adieraziak.

Hona normalizazio-prozesu hau adibide batekin azaldua:

izozgailu A0 iz. Ura, eta berezk. janariak, izozteko tresna.

Aurreko adibidean agertzen den *berezk.* laburdura lematizatu nahi izanez gero, dugun analizatzailearen gaur egungo bertsioak ez du jakingo zein duen lema zehatza, eta hitz ezezagun gisa, balizko lema jasoko ditu. Ondoren puntua dagoenez, hurrengo hitza —*janariak*— definizio-hasiera dela markatuko du. Aldiz, definizioan *bereziki* agertuz gero, analizatzaileak ez du aurreko arazorik topatuko, eta ondo egingo du analisia.

Beraz, definizioetako hitzen analisiak egiteko garaian *Morpheus* da erabili dugun analizatzaile morfologikoa. Baina definizioak ez dira analizatu beharreko elementu bakarrak; sarrera-hitzen analisia ere egin beharko dugu tesi honen V. kapitulurako. Horretarako ere *Morpheus* analizatzailea erabili dugu ondoko atalean azalduko dugun bezala.

III.6. Sarrera-hitzen analisia

Sarrera-hitz eratorriak zeintzuk diren, eta eratorpen prozesu horretan zein elementu agertzen diren ezagutu nahi dugunez, sarrera-hitz guztiak analizatu ditugu. Analizatzaile morfologikoak oinarria eta atzizkia bereizten ditu. Horretarako beharrezko duen informazioa EDBLtik hartzen du, eta hitz baten morfologia ezagutu dezan, behar duen informazioa EDBLn egon behar du.

Baina gerta daiteke atzizki lexikal jakin bat datu-basean landuta ez egotea, eta atzizki hori daraman hitza ez du eratorri gisa ezagutuko. Demagun arrazakeria analizatu nahi dugula *Morpheus* tresnarekin *arraz* + *keria* emaitza lortzeko.

Baina *-keria* atzizkia tesirako erabili dugun datu-basearen bertsioan atzizki lexikal gisa landuta ez dagoenez, eta *arrazakeria* hitza bera datu-basean landuta dagoenez, emaitza ez da egokia eratorpena lantzeko, hitz hori eratorria denik ez dakigulako.³⁵

```
/<arrazakeria>/
```

```
("arrazakeria"  IZE ARR + DEK ABS MG)
("arrazakeria"  IZE ARR + DEK ABS NUMS MUGM)
("arrazakeria"  IZE ARR)
```

Hitza datu-basean ez badago, berriz, hitz ezezagun gisa analizatuko du. Ezezagunen analisi horretan, laburdurekin ikusi dugun bezala, interpretazioa asmatzen saiatzen da modu logiko batean, baina atzizki lexikalei dagokienez, datu-basean daudenak baino gehiago erabiltzen ditu ezagutza egiteko. Demagun testu batean *gorrikeria* moduko zerbait agertzen zaigula. Hitz hau ez dago EDBLn landuta eta analizatzaile morfologikoa nolabaiteko analisisia ematen saiatzen da; hainbat posibilitate ematen ditu ea asmatzen duen. Posibilitate horien artean hitz eratorriarena ere agertzen da.

```
/<gorrikeria>/
```

```
EZEZAG "gorrikeria "  ADJ IZO
EZEZAG "gorrikeria "  ADJ IZO + DEK ABS MG
EZEZAG "gorrikeria "  ADJ IZO + DEK ABS NUMS MUGM
EZEZAG "gorri+keria "  ADJ IZO + ATZ IZE ARR
EZEZAG "gorri+keria "  IZE ARR + ATZ IZE ARR
EZEZAG "gorri+keria "  ADJ IZO + ATZ IZE ARR + DEK ABS MG
EZEZAG "gorri+keria "  ADJ IZO + ATZ IZE ARR + DEK ABS NUMS MUGM
EZEZAG "gorri+keria "  IZE ARR + ATZ IZE ARR + DEK ABS MG
EZEZAG "gorri+keria "  IZE ARR + ATZ IZE ARR + DEK ABS NUMS MUGM
```

Hau ikusita, eratorpena lantzeko EDBLko informazioa erabili beharrean, hitz ezezagunekin egiten den analisi hau erabiltzea pentsatu genuen. Baina, analisi mota honetan edozer gauza ager daiteke, eta egile moduko hitz bat *e + gile* gisa analizatuta agertuko zaigu. Ikusi dugunez, ezezagun gisa tratatzen diren hitzei aplikatzen zaizkien atzizki lexikalak ez dira hainbeste, eta ondoren azalduko dugun tarteko bidea hartzea erabaki dugu.

Lehenik, hiztegiko sarrera-hitz guztiak hartu ditugu, eta amaiera usuenak zein diren jakiten saiatu gara maiztasunak atereaz. Hona adibide bat: adjektiboen artean hitzen amaieran dauden letra-seikote usuenak hauek dira:

³⁵ Hiru analisisietan *arrazakeria* dugu lema gisa.

124	agarri	19	sgarri
93	tzaile	17	agaitz
28	igarri	14	igaitz
27	ngarri	13	logiko
24	rgarri	12	egarri

Horiek banan-banan ikusi, eta dagoeneko badugu atzizki izan daitekeen elementua: –*tzaile* 93 aldiz agertzen da adjektiboen artean. Ondoren letra-boskoteak ikusiz gero, hona emaitza:

274 garri
93 zaile
69 gaitz

Hemen aurreko multzoan dagoeneko aurreikus zitekeen beste bat agertu da (-*garri*), eta, nahiz eta atzizki lexikal ez izan, hitz-amaiera gisa askotan agertzen den beste elementu bat ere bai (-*gaitz*).

Ondoren, sarrera-hitzen artean usuen agertzen diren esanahidun amaierak hartu eta *Morpheus* tresnari txertatu dizkiogu. Hala, guk agintzen diogun atzizkiaren pareko zerbait topatzen badu sarrera-hitz batean, oinarria eta eratorria bereiziko ditu, baina oinarriak EDBLn egon beharko du; bertan agertzen ez bada, hitz hori ez dugu kontuan hartuko.

gerrate A0 iz. Gerra jakin bat; gerra denbora.

gerrate hitzak badu oinarrian *gerra izena*, eta *denbora* adierazten duen –*te* atzizki lexikala.

Hala ere, zenbait hitz ez eratorri agertuko zaizkigu eratorri gisa; halakoak definizioa aztertzerakoan baztertu egingo ditugu. Hona adibidea:

pate A0 iz. Sukaldaritzan, haragi edo gibel xehatuaz, zerri edo hegaztienaz batik bat, egindako oreia.

pate hitza ez da izango *pa izena* —*muxua* adierazten duena— gehi –*te* atzizkia, nahiz eta analizatzaile morfologikoak hala topatu. Eratorpena azaltzen dugunean (V. kapitulua) ikusiko dugu bat hautatu eta bestea saihesteko erabiltzen dugun metodoa zein den.

Beraz, dagoeneko hiztegi osoa analizatuta dugunez, erlazioak erauzteko zein tresna erabili dugun azalduko dugu.

III.7. Erlazioak erauzteko tresna: *Murriztapen Gramatika*

Vanderwende-k (1995) bere lanean agertzen duenez, definizio bateko hitzen arteko erlazioak erauzteko garaian garrantzi handikoa da definizio horien analisi sintaktikoa eginga izatea. Halako analisirik izanez gero, lotuta dauden hitzak zeintzuk diren jakin ahal da, eta erlazioak lotuta dauden hitz horien artekoak izango dira. Dagoeneko esan dugun bezala, euskarazko analizatzaile sintaktikoa garapenean dago, eta definizioak esaldi osoak ez direnez, arazoak sortzen dira haiengan aplikatzerakoan.

Beraz, definizioetan dauden hitzen arteko erlazioak erauzi ahal izateko sintaxi analizatzaile xumea garatu dugu. Gehien bat, izen sintagmen ezaugarriak izan ditugu kontuan, eta, horien arabera analizatzailea sortu.

Analizatzaile hori sortzeko *Murriztapen Gramatika* (*Constraint Grammar*) erabili dugu (Karlsson *et al.*, 1995; Tapanainen, 1996).

Gramatika honekin helburu dugun hitza bilatzen dugu, inguruan izan ditzakeen hitzen ezaugarriak kontuan hartuta. Demagun *indioilar* izenaren definizioa dugula:

```
indioilar  A0      iz.  Indioilo arra.
```

Izenen kasuan, definizioa izen-sintagma bat da, eta, demagun, sintagma horren buru sintaktiko eta semantikoa³⁶ topatu nahi dugula. Kasu honetan, *indioilo* hitza interesatzen zaigu, eta horren ingurua aztertuko dugu. Hona inguruaren analisi morfoloikoa:

```
/<@lema      indioilar>/<ID>/
/<@@Adiera_string  A0>/<ID>/
/<@@Kategoria  iz.>/<ID>/
"<Indioilo>"
    "indioilo"  IZE ARR DEK ABS MG  HAS_MAI DEF_HASI NOTGELGEN
    "indioilo"  IZE ARR  ZERO HAS_MAI DEF_HASI NOTGELGEN
"<arra>"
    "ar"  ADJ IZO DEK ABS NUMS MUGM
    "ar"  IZE ARR DEK ABS NUMS MUGM  NOTGELGEN
    "arra"  IZE ARR DEK ABS MG  AORG NOTGELGEN
    "arra"  IZE ARR DEK ABS NUMS MUGM  AORG NOTGELGEN
    "arra"  IZE ARR  ZERO AORG NOTGELGEN
"<$."
    PUNT_PUNT
```

³⁶ Ikusiko dugun bezala, zenbaitetan ezberdinak dira buru sintaktikoa eta semantikoa.

Aurreko adibidea eskuartean izanda, gunea topatu nahi dugu (IZE), eta gunea horri marka bat jarriko diogu (&GENUS-IZE003).³⁷ Marka hori jartzeko agindua MAP da. Baina IZE horri marka hori jarriko diogu baldin eta (IF) deklinabide markarik ez badu (0 ZERO), ondoren *adjektibo absolutibo mugatua* baldin badu (1 ADJ + ABS + MUGATUA), eta honen ondoren *definizio bukaera* baldin badu (2 DEF_BUKA).³⁸ Analisi ambigua erabili dugula azaldu dugunean esan dugun bezala, ez du axola ar adjektibo eta izen gisa markatuta egotea. Hona aurreko guztia islatzen duen erregela:

```
MAP (&GENUS-IZE003) IZE
  IF (NOT 0 ADJ OR DET) (0 ZERO) (1 ADJ + ABS + MUGATUA) (2 DEF_BUKA) ;
```

Aurrekoez gain, erroreak sor ez daitezen, badira beste baldintza batzuk, hala nola, markatu nahi dugun hitzak adjektibo edo determinatzaile irakurketak ez izatea (NOT 0 ADJ OR DET). Murritzapen zehatz hauek beharrezkoak dira. Gerta daiteke bi izen elkarren ondoan agertzea, eta orduan bigarren izena izango da izen-sintagma horren buru sintaktiko eta semantikoa, besakaren definizioan topatu dugun bezala.

```
besaka      H1A0    iz.  Zuhaitz  adarra.
```

Hona definizio horren analisi morfologikoa:

```
/<@lema      besaka>/<ID>/
/<@Adiera_string  H1A0>/<ID>/
/<@Kategoria  iz.>/<ID>/
"<Zuhaitz>"
  "zuhaitz"  IZE ARR DEK ABS MG  HAS_MAI DEF_HASI NOTGELGEN
  "zuhaitz"  IZE ARR  ZERO HAS_MAI DEF_HASI NOTGELGEN
  "zuhaiztu" ADI SIN AMM ADOIN  NOTDEK HAS_MAI DEF_HASI
"<adarra>"
  "adar"    IZE ARR DEK ABS NUMS MUGM  NOTGELGEN
"<$."
  PUNT_PUNT
```

Kasu honetan aurreko erregelak ez du ezer topatzen lehenengo izenaren eskuinean ez dugulako adjektiborik. Kasu honetan gunea sintaktiko eta semantikoa markatuko duen erregela honokoa da:

```
MAP (&GENUS-IZE000) IZE
  IF (NOT -1 PUNTUAZIOA) (NOT 0 ADJ OR DEF_HASI)
    (0 ABS + SINGULARRA + MUGATUA) (1 PUNTUAZIOA) ;
```

³⁷ Erregela bakoitzak marka bat du —zenbaki batekin—, erregela zenbatetan aplikatzen den ezagutzeko, edo arazoren bat sortzen duenean arazoa non dagoen berehala detektatzeko.

³⁸ Bilatzen ari garen elementuarekiko posizioak zenbakiekin markatzen dira: 0a helburu den hitzak duen posizio bera da; ezkerretara lehen posizioa -1 da, ondorengoa -2, eta abar; eta eskuinetara lehen posizioa 1 da, ondorengoa 2, eta abar.

Markatu (MAP) &GENUS-IZE000 markarekin izen (IZE) bat, baldin eta (IF) ezkerrean lehenengo posizioan *puntuazio-markarik* ez badago (NOT -1 PUNTUAZIOA), markatu nahi dugun *izena absolutiboa singularra* eta *mugatua* bada (0 ABS + SINGULARRA + MUGATUA), eta eskuinean *puntuazio-marka* badu (1 PUNTUAZIOA).

indioilarren definizioaren kasuan ar izena dugu, baina aurretik ez du puntuazio-markarik. Gainera, beste murriztapen bat jarri diogu erregela honi, hala nola, markatu behar duen *izenak* ez izatea *adjektibo* irakurketa edo definizio hasiera ez dadila izan (NOT 0 ADJ OR DEF_HASI).

Murriztapen Gramatika erabiliz sortu ditugun erregela guztiak I. eranskinetan paratu ditugu.

III.8. Hiztegiaren azterketa estatistikoa

Atal honetan EHren datuak emango ditugu, bai sarrera kopurua, bai hitz-formen maiztasuna, eta, definizioen egitura ezagutzeko asmoz, baita definizioetan agertzen diren *token*-konbinazioen maiztasuna ere.

Hiztegiko datu gordin hauek erakutsiko digute zein definizio-patroi erabili diren kategoria bakoitzarekin. Definizio-patroi horiek aztertuta, definizioeko gunek ezagutzeko erregelak egin ahal izango ditugu, eta baita patroik lexikalak ezagutzeko erregelak ere.

III.2. atalean esan badugu ere, 7. taulan EHren TEI bertsioak dituen sarrera-, adiera- eta azpisarrera-kopurua paratu ditugu.

sarrerak	33.124
adierak	53.961
azpisarrerak	3.209

7. taula. Hiztegiko sarrerak, adierak eta azpisarrerak: kopuruak.

Zenbaitetan gerta daiteke bai ñabardura batek edo bai azpisarrera batek definizio-eremurik ez izatea, eta horiek ez ditugu lan honetan kontuan izango.³⁹ Beraz, kontuan izan behar dugu zenbat diren definizio-eremua duten adierak edo azpisarrerak, eta horiek, 8. taulan agertzen den bezala, 35.549 dira.

³⁹ Denera 13.955 balizko adiera daude definiziorik gabe. Horietatik 2.359 sarrera hilak dira; 6.460 ñabardurak; eta 1.474 azpisarrerak. Halako askok definizioa izan beharrean, erabilera-adibideak dituzte.

izenak	21.511
adjektiboak	4.310
izen-adjektibo	1.020
aditzak	6.132
azpisarrerak	2.576
Guztira	35.549

8. taula. Hiztegiaren definizioa duten adierak kategoriak.

9. taulan, berriz, aztergai ditugun definizio horietan agertzen diren hitz-forma usuenen kopurua paratu dugu. Kategoria bakoitzeko definizioak bere aldetik aztertuko ditugunez, maiztasun-kopurua kategoriak eman dugu. Kopuru osoa emateaz gain, definizioetan agertzen diren hitz guztiekiko ehunekoa paratu dugu.

Izenak			Aditzak			Adjektiboak			Izenak-adjektiboak			Azpisarrerak		
dena	105604	%		21957	%		15530	%		4571	%		8817	%
edo	3215	3,0	edo	1227	5,6	duena	1185	7,6	duena	326	7,1	eta	182	2,1
eta	2983	2,8	egin	496	2,3	dena	560	3,6	dagokiona	120	2,6	edo	168	1,9
den	1965	1,9	bihurtu	471	2,1	edo	490	3,2	dena	118	2,6	egin	140	1,6
duen	1209	1,1	behartu	429	2,0	ez	438	2,8	biztanlea	106	2,3	den	109	1,2
dena	1112	1,1	eman	310	1,4	dagokiona	231	1,5	edo	99	2,2	izan	104	1,2
duena	837	0,8	hartu	280	1,3	dituena	177	1,1	eta	63	1,4	dena	85	1,0
bat	723	0,7	bat	275	1,3	mintzatuz	154	1,0	ez	55	1,2	bat	66	0,7
baten	707	0,7	izan	204	0,9	eta	151	1,0	mintzatuz	45	1,0	gabe	63	0,7
egiten	680	0,6	eta	186	0,8	betea	151	1,0	egiten	39	0,9	ez	63	0,7
erabiltzen	645	0,6	jarri	169	0,8	egiten	105	0,7	dituena	35	0,8	duena	54	0,6
bereziki	636	0,6	zerbait	129	0,6	asko	105	0,7	aldekkoa	34	0,7	landarea	45	0,5
nolakotasuna	564	0,5	galdu	120	0,5	dagoena	95	0,6	bereziki	30	0,7	dituena	44	0,5
dituena	536	0,5	kendu	98	0,4	joera	94	0,6	duen	29	0,6	ere	41	0,5
diren	498	0,5	Zerbait	92	0,4	Ezin	65	0,4	pertsona	27	0,6	erabiltzen	40	0,5
denaren	480	0,5	joan	90	0,4	Erraz	62	0,4	bat	27	0,6	hartu	39	0,4
dituen	460	0,4	bidez	88	0,4	merezi	53	0,3	den	23	0,5	egiten	38	0,4
pertsona	435	0,4	ez	87	0,4	gabea	53	0,3	sortua	21	0,5	egon	36	0,4
multzoa	431	0,4	baten	84	0,4	ezin	51	0,3	herri	20	0,4	familiako	34	0,4
bakoitza	398	0,4	eraman	83	0,4	daitekeena	51	0,3	Zerbait	20	0,4	eman	34	0,4
ematen	366	0,3	bere	83	0,4	ematen	48	0,3	esaten	19	0,4	duen	34	0,4
ez	361	0,3	gertatu	82	0,4	ia	46	0,3	dagoena	18	0,4	dela	28	0,3
batean	337	0,3	bete	80	0,4	handia	46	0,3	da	18	0,4	ematen	26	0,3
batez	302	0,3	bereziki	79	0,4	erraz	41	0,3	kidea	17	0,4	batean	25	0,3
egiteko	285	0,3	ezarri	77	0,4	den	41	0,3	zen	15	0,3	baten	24	0,3
duten	274	0,3	den	75	0,3	adierazten	40	0,3	ematen	15	0,3	lore	23	0,3
txikia	260	0,2	mintzatuz	74	0,3	baino	37	0,2	batez	14	0,3	esapidea	23	0,3
tresna	258	0,2	beste	71	0,3	zale	34	0,2	dituen	13	0,3	eginez	23	0,3
handia	244	0,2	utzi	68	0,3	gabe	34	0,2	mintzatzen	12	0,3	alde	23	0,3
dagoen	227	0,2	batean	67	0,3	eragiten	34	0,2	Antzinako	12	0,3	dituen	21	0,2
bi	227	0,2	jo	62	0,3	dagokiena	34	0,2	hartzen	11	0,2	oso	20	0,2
egoera	225	0,2	batez	61	0,3	bat	33	0,2	ere	11	0,2	bere	20	0,2
gauza	217	0,2	zerbaiten	57	0,3	agertzen	32	0,2	bizi	11	0,2	Erabat	20	0,2

9. taula. Definizioetan dauden hitz usuenen kopurua.

Hiztegiaren agertzen diren hitz-forma usuenen zerrendak erakusten du zeintzuk izan daitezkeen definizioetan topatu ahal izango ditugun definizio-patroietan agertuko diren hitz-forma usuenak. Adibidez, izenen artean erlatibozko adizkia askotan agertzen da (den, duen), eta honen arrazoia da izen baten definizioaren gunearen

aurretik askotan erlatibozko perpausa agertzen dela. Adjektiboetan, aldiz, usuenak izango dira elipsidun formak (*duena*, *dena*) —ia *izenetan* duten kopuru bera da, baina askozaz definizio gutxiagoren artean: *izenetan*, *duena* %0,8 eta *dena* %1,1; adjektiboetan, *duena* %7,6 eta *dena* %3,6—. Aditzetan, aldiz, goi mailan agertzen zaizkigu partizipioak. Partizipio hauek oso esanahi orokorrekoak dira, eta, genusaren posizioan agertzen badira, agian erlatore berezi gisa aztertu beharko ditugu (ikus IV.3.3. atala). Izen eta adjektibo diren sarreretan nabariak dira elipsidun formak. Agian honek adierazten du gehien bat adjektiboen artean erabiltzen diren definizio-patroiak erabiliko direla (ikus IV.4.11. atala). Multzo berean —*izenak-adjektiboak*— oso nabaria da *biztanlea* hitzaren maiztasuna; gerora ikusiko dugun bezala, *-ar* eta *-tar* atzizkia daramaten hitzak definitzeko erabiltzen den patroia zatia da, eta halakoek izen eta adjektibo kategoria jasotzen dute askotan. Azpisarreretan, badirudi goian agertzen diren batzuk aditz-partizipioak direla —*egin*, *izan*, *hartu*, *egon*—; honek adieraziko du azpisarreretan aditz asko izango dugula. Horrezaz gain, *landarea* hitz-forma ere maiz agertzen denez azpisarreraren definizioetan, *landare-izen* dezente egon daitezkeela pentsatuko dugu.

Aurreko neurketa egin ondoren, lehen maiztasuna neurtu dugu, lehenen bat maiztasun txikiagoa duten forma ezberdinekin agertzen den ikusteko; baina aurreko neurriarekiko ezberdintasun handirik ekarri ez digunez, ez dugu horretan gehiago sakonduko.

Eta neurketekin jarraituz, definizioetan dauden elementuen segidak aztertu ditugu agertzen diren molde usuenak zeintzuk diren ezagutzeko. Definizioetako elementu horiei *token* deitu ohi zaie, eta horien azterketa egingo dugu. Definizioetan hainbat *token* —hitz-formak, puntuazio-ikurrak— agertzen dira, eta, hemen aztertuko duguna da *token* horiek elkarren ondoan zenbatetan agertzen diren. Lehenik *token* horiek bere horretan hartuko ditugu, kontuan izan gabe kasu gramatikalak eta postposizio-atzizkiak, baina ondorengo azterketa batean hauek ere kontuan izango ditugu.

Azterketa hau kategoria guztietako definizioak kontuan hartuta egin dugu. 10. taulan paratu ditugu *token*-laukote, -hirukote eta -bikote usuenak.

laukote	zenb.	hirukote	zenb.	bikote	zenb.
ematen zaien izena .	55	denaren nolakotasuna .	336	duena .	1721
. Gai bakuna ,	47	erabiltzen dena .	247	dena .	1307
hartu edo eman .	43	duen pertsona .	183	dituena .	549
. - r en	39	ez duena .	177	duena ,	467
den aldi bakoitza .	33	; bereziki ,	116	nolakotasuna .	434
. Norbait edo zerbait	32	, batez ere	111	denaren nolakotasuna	414
. Pertsonen mintzatuz ,	28	ez dena .	110	egin .	386
oso aintzat hartua .	27	egiten duena .	102	behartu .	380
. Zerbait edo norbait	27	, eta bereziki	81	pertsona .	361
egiteko erabiltzen dena .	25	joera duena .	78	bakoitza .	350
, jateko ona .	24	ematen dituen .	68	, bereziki	343
den aldietako bakoitza .	23	dituen pertsona .	66	multzoa .	333
ematen zaion izena .	23	ematen zaien izena	63	. Zerbait	328
egiten dituen eskulangilea .	22	hartu edo eman	63	mintzatuz ,	323
gisa erabiltzen dena .	21	den pertsona .	60	bihurtu .	318
joan edo eraman .	21	familiako landarea ,	59	dena ,	301
dena , ia ezin	21	- r en	58	, eta	295
galdu edo kendu .	21	ez duena ,	58	erabiltzen dena	272
egiten dituen pertsona .	19	zaien izena .	56	ez duena	260
hartu edo eman ,	18	duen zera .	53	erabiltzen den	255
Gai bakuna , metal	18	; atomo zenbakia	51	; bereziki	233
. Belar landarea ,	18	; atomo masa	51	egiten den	231
, - r en	17	egiten dena .	51	duen pertsona	211
egiten duen pertsona .	17	gertatzen dena .	49	izan .	210
. Norbaiti buruz ,	17	. Gai bakuna	47	eman .	207

10. taula. Hiztegiaren definizioetan agertzen diren *token*-laukote, -hirukote eta -bikote usuenak.

Azterketa honetatik dagoeneko zenbait erregela-eskema lor ditzakegu, hala nola “ematen zaien izena .” laukotea. Badirudi gehien errepikatzen diren segidak ondoren puntuazio-marka dutenak direla, alegia, gero ikusiko dugun bezala, erlatore bereziak direnak (ikus IV.4. atala).

Hirukoteen artean, goian dauden artean elipsidunak dira (ikus IV.4.11. atala), eta *genus orokor* deituko ditugunak (ikus IV.3.3. atala) ere agertzen dira —nolakotasuna, pertsona, izena, hartu, eman—.

Bikoteetan, argi da ondoren puntuazio-marka duten elipsidun adizkiak agertzen direla lehen leku guztietan, eta IV.4.11. atalean ikusiko dugun bezala, elipsi hau erlatore berezi baten isla da.

Elementu osoak kontuan hartzeaz gain, analisi morfologikoa eginda dugunez, kasu gramatikalak eta postposizio-atzizkiak ere kontuan har ditzakegu. Halakoei *postposizio-token* deituko diegu. Postposizio horren azpian kasu gramatikalak eta postposizioak biltzen dira, eta *token*aren azpian, hitzak eta definizioan egon daitezkeen bestelako ikurrak —puntuazio-markak, parentesiak, gidoiak, eta abar—.

Multzo honetan dagoeneko 10. taulan agertu diren zenbait daude, eta horiek ez ditugu 11. taulan errepikatuko. 11. taulan *postposizio-token* segidak kontuan hartuta sortu diren segida berri usuenak jarri ditugu, dagokien azalpenarekin (*balizko transkripzioa*).

postposizio-token	balizko transkripzioa
ABS izan ERLT !ELI GEN nolakotasun ABS .	-a denaren nolakotasuna.
GEN ABS nolakotasun ABS.	-en nolakotasuna.
ABS izan ERLT !ELI GEN !ELI	-a dena
INE izan ERLT !ELI ABS .	-n dena.
ABS ukan ERLT !ELI ABS .	-a duena
ABS edun ERLT !ELI ABS .	-a duena
INE ukan ERLT !ELI ABS .	-n duena.
INE edun ERLT !ELI ABS .	-n duena.
DAT egon ERLT !ELI ABS.	-i dagokiona.
DAT egon ERLT !ELI ABS ;	-i dagokiona;
GEN !ELI ABS multzo ABS .	-en multzoa.
GEL !ELI ABS bakoitz ABS .	-ko bakoitza.

11. taula. *Postposizio-token* segida usuenak kontuan hartuta ditugun patroik berriak.

Aurreko multzoan ditugu aurretik kasu markaren bat maiz duten patroiak. Kasu marka horiek zeintzuk diren kontuan hartzea garrantzitsua da ondorengo ataletan definizio-eskemak aztertzen ditugunerako.

III.9. Ondorioak

Kapitulu honetako lanaren helburua iturri dugun hiztegia ezagutzea izan da, bai azalean, eta bai sakonean. Hiztegiaren egitura orokorra zein den aztertu dugu. Modu honetan, paperezko bertsioan sarrera-hitza bakoitzean hainbat ikur lexikografikoren bitartez markatuta dagoena zer den jakin ahal izan dugu: *homografo-zenbakia*, *sarrera*, *kategoria*, *data*, *aldaera*, *erabilera-laburdura*, *definizioa*, *erlazio-ikurra*, *adibidea* eta *azpisarrera*. Hau ezinbestekoa da paperezko hiztegi gisa erabiltzeko pentsatuta dagoen produktua makinek erabiltzeko produktu bihurtzeko. Lan horretan eremu bakoitza markatzeko TEI ekimenaren gidalerroak erabili ditugu. Gidalerro hauen abantaila nagusiak argitasuna, sinpletasuna, zehaztasuna, eta nazioarteko estandartzat onartuta egotea dira.

Tesi-lan honetarako, ordea, ez ditugu eremu horiek guztiak behar, eta guretzat erabilgarri diren elementuak baino ez ditugu hartu: *sarrera-hitza*, *kategoria*, *adiera-ikurra*, *definizioa*.

Eremu guztiak behar ez izateaz gain, hiztegian agertzen diren kategoria guztiak ere ez ditugu landu. Azterkizun izan ditugunak *izenak*, *adjektiboak* eta *aditzak* dira. Bakoitza bere aldetik aztertu dugunez, multzo ezberdinetan paratu ditugu. Zenbaitetan sarrera-hitz batek adiera berean kategoria-marka bat baino gehiago ditu —adjektibo eta izen kategoria, adibidez— eta halakoekin multzo berezia osatu dugu lehen begi kolpean ezin izan baitugu multzo batean edo bestean sailkatu. Azpisarrerak ere aztertu ditugu, eta, nahiz eta sarrera-hitzak bezala aztertu ditugun, sarrera-hitz bati lotuta daudenez, eta zenbait ezaugarri berezi izaten dituztenez, hauekin ere beste multzo bat osatu dugu. Beraz, bost multzo izan ditugu azterkizun: *izenak*, *adjektiboak*, *aditzak*, *izenak eta adjektiboak* eta *azpisarrerak*.

Definizioan dagoen informazio inplizitua makinari esplizitu egiteko, definizio guztien analisi morfologikoa egin dugu. Dagoeneko definizioak ez dira hitz-forma multzoak, baizik eta lema-postposizio multzoak. Analisi morfologiko hori gauzatzeko Ixa Taldeak garatutako *Morpheus* (Urkia, 1997) erabili dugu.

Lan honetan definizioko hitzen analisi anbigua erabili dugu. Definizioak esaldi osoak ez direnez, desanbiguazioa gauzatzeko garaian akatsak sortzen dira, eta akats hauek gure emaitzak kaskartzen dituzte.

Analisi anbiguen gainean lan egitea ez da eragozpen handia izan definizioetan dugun informazioa kudeatzeko eta nahi dugun informazioa ateratzeko testuingurua kontuan hartzen duen *Murritzapen Gramatika* erabili dugulako. *Murritzapen Gramatika* erabiliz sortutako erregelei esker, guk elementu zehatz bat bilatu nahi dugunez, ez da eragozpen handia hitz-forma baten analisi guztiak izatea.

Baina, Vanderwendek (1995) dioten bezala, definizioen sintaxia eta azaleko sintaxia izatea garrantzitsua da, erlazionatutako hitzak zeintzuk diren jakiteko. Azaleko sintaxia izanez gero, erlazionatutako hitzak zeintzuk liratekeen jakina litzateke, eta gure lana, horien arteko erlazioa zein den zehaztea litzateke; baina azaleko sintaxia ezin izan dugunez erabili, definizio-guneak ezagutzeko erregela sintaktiko batzuk idatzi ditugu. Idatzitako sintaxia, gehien bat, izen sintagmen osaera ezagutzeko erregelak izan dira.

Definizioez gain azterkizun ditugu sarrera-hitz eratorriak. Lan honetarako ere *Morpheus* analizatzaile morfologikoa erabili dugu. Analizatzailea atzizki lexikal berriekin aberastu dugu ahalik eta informazio gehien lortzeko. Analizatzailean zein

atzizki berri txertatu jakiteko, hiztegiko sarrera guztien letra-amaiera usuenak zeintzuk diren aztertu ditugu, horietatik atzizki zeintzuk diren erabaki, eta analizatzailean txertatu. Letra-amaiera usuenen artean atzizki lexikalak ez ezik, konposizio-elementuak ere agertu dira eta hauek ere analizatzailean txertatu ditugu (ikus V.5. atala).

Hitzen arteko erlazioak zehaztea denez gure helburua, hiztegiko definizioetan usuen agertzen diren hitz-formen eta lemen azterketa egin dugu. Hitz-forma edo lema bat maiz agertzen bada, garrantzi berezia duela pentsatu dugu, eta ondoko kapituluetan ikusiko dugunez, hau izan da erlatore berezien bilaketan erabilitako oinarri nagusia. Horrezaz gain, *token*-segidak eta *postposizio-token* segidak ere aztertu ditugu, eta analisi honek erlatore berezien eta patroien lexikalen bilaketan lagundu gaitu.

Beraz, atal honetan hiztegiaren hezurdura nolakoa den azaldu dugu. Ondoko lana izango da, hezurdura hori lotzen den moduaren arabera, makinari adieraztea erlazio baten edo bestearen isla den.

Bizitzan jaiotza / heriotza
maiz dituguuntu xamar...

sehaska	A0	iz.	Haurrentzako ohe txikia, gehienetan kulunka daitekeena.
kanta	A0	iz.	Ahapaldietan banatutako neurtitz herrikoa, doinu baten gainean kantatzen dena.

Amaren bularra
haurraren janari
haurraren negarra
beti hunkigarri,
baratzaren erdian
goseak nago ni
zure bular lastanen
antsiz ta egarriz

Zure bular politik
zerrien jostailu
edertasunak kontuz
gorderik behar du,
zaude ixilik haurra
ez egin negarrik
munduak ez du eta
malkoen beharrik

Handitzen zeranean
ikusiko duzu
ixilik egoteak
zenbat balio dun,
jarrai zure bidetik
hobe da horrela
gauza denen gainetik
haizea bezala.

Amaren bularra
haurraren janari
aterpe xamur hortan
lo hartu nahi dut nik

(X. Lete)

IV. Definizioen guneak erauzten: sinonimoak, genusak, eta erlatore bereziak

IV.1. Sarrera

Hiztegi elebakarretatik informazioa erauzterako garaian lehenik sortzen diren loturak sarrera-hitza eta definizioan dagoen hitz baten artekoak dira (Castellón, 1992; Artola, 1993). Definizioak eta sarrera-hitzak lotuta daude; definizioek sarrera-hitzen baliokide izan behar dute. Baina baliokidetasun horretan, hitz guztiek ez dute pisu bera. Kapitulu honen helburua definizio gune diren elementuak ezagutzea da, eta horretarako horien inguruan agertzen diren elementuak eta sintaxia aztertuko ditugu, baina ez definizio osoko sintaxia. Definizio osoaren analisia egitea beste kapitulu baten helburua izango da (ikus VI. atala). Zenbait definiziotan gerta daiteke bi guneri buruzko informazioa izatea. Hau ez da arazo izango, sarrera-hitz batek adiera beraren barruan eta definizio-eremu bakarrean izan baititzake genus bat baino gehiago; baita sinonimo bat baino gehiago ere.

Hiztegian dauden definizio-ereduak kontuan hartuz gero, sinonimoak, genusak, edo erlatore bereziak izango ditugu —banan-banan dagokien atalean ikusiko ditugu—. Hiruek erlazio lexiko-semantiko ezberdinak islatuko dituzte definizioan markatzen dugun hitzaren eta sarrera-hitzaren artean: sinonimoek, *sinonimia* erlazioa; genusek, *hiperonimia* / *hiponimia* erlazioa; eta erlatore bereziek beste hainbat erlazio islatuko dituzte erlazionatutako hitzaren eta sarrera-hitzaren artean.

Erlazionatuta agertzen diren hitzen kategoriaren arabera, erlazioak paradigmaticoak edo sintagmaticoak dira. Paradigmaticoak kategoria bereko hitzen artekoak dira, eta sintagmaticoak kategoria ezberdineko hitzen artekoak. Erlazio horiek guztiak erauzteko hiztegi elebakarrek dituzten hainbat ezaugarritan oinarrituko gara.

Sinonimiari dagokionez, bi modutan agertu ohi da hiztegietan: batetik eremu markatu gisa —definizioa ez den eremu batean, alegia— edo definizio-eremuan bertan.

Eremu markatu gisa agertuz gero, erauzketa-lana asko errazten da. Lexikografoak berak markatzen du, etiketa jakin bat erabiliz, zein den sarrera-hitzaren sinonimoa. Gure kasuan *ik.* marka duen eremua dugu. Sinonimiari buruz hitz egiten denean, kontu handia izan behar da. Erabateko sinonimia existitzen ez dela aski onartua dagoenez (Robert, 1990; Miller, 1998a; Sparck-Jones, 1964; 1986), hemen, kontestu jakin batzuetan aldagarri diren hitzei buruz ariko gara. Hiztegian *ik.* eremuaren barruan halakoak agertzen dira, baina baita besterik gabe nolabaiteko erlazioa duten hitzak.

Eremu markatu horrezaz gain, zenbaitetan definizioan sarrera-hitzaren deskribapen bat izan beharrean, hitz bakar bat izango dugu, edo komaz bereizitako hitz multzoa. Halakoak sinonimotzat hartuko ditugu. Kasu hauetan azaleko erlazioa eta sakonekoa bera dira: *sinonimia*, alegia. Erlazio hau kategoria bereko hitzen artekoa da.

Bestetan —hiztegi elebakar batean ohikoena— sarrera-hitzaren deskribapena izango dugu definizio-eremuan. Halakoetan, bi definizio mota ditugu:

- Gune sintaktikoa eta semantikoa hitz bera duena.
- Gune sintaktikoa eta semantikoa diren hitzak ezberdinak dituen.

Gune sintaktikoa eta semantikoa hitz bera denean, gune hori sarrera-hitzaren genusa da, eta bien arteko erlazioa *hiperonimia* / *hiponimia*. Erlazio hau —*sinonimikoa* bezala— kategoria bereko hitzen artean aurkitzen dugu.

anbulantzia	A0	iz.	Eriak eramateko prestaturiko ibilgailua .
hegazkin	A0	iz.	Airean ibil daitekeen ibilgailua , airea baino astunagoa, hegoduna eta motorduna.
tranbia	A0	iz.	Hiri eremuetan errailen gainean zirkulatzen duen ibilgailua , argindarrez higitzen dena.

anbulantzia, hegazkin, eta tranbia hitzen definizio-eremuan agertzen zaigun izen-sintagmaren buru sintaktikoa eta semantikoa ibilgailu da. Baina ibilgailu horiek ez dituzte ezaugarri berak izango, eta ezaugarri horiek zehazten dituzte gunearen inguruan agertzen diren modifikatzaileek: eriak eramateko, airean ibil daitekeena, eta errailen gainean zirkulatzen duena. Modifikatzaile horiek dira, hain zuzen ere, definizio horren *differentia*. Hitzak definitzeko era hau —*genus + differentia specifica*— Aristotelesek aipatu zuen. *Differentiari* buruzko azterketa VI. atalean egingo dugu; honetan, definizioaren guneari buruzko azterketan zentratuko gara.

Zehaztu behar dugu, hala ere, *hiperonimia* / *hiponimia* erlazioa izenen eta aditzen definizioetan topatuko dugula.⁴⁰ Adjektiboen artean ez dugu *hiperonimia* / *hiponimia* erlazorik. Adjektiboen artean, hain zuzen ere, gehien erabiltzen den deskribapen mota gune sintaktikoa eta semantikoa ezberdina direnekoa da. Halakoetan, gune sintaktikoari erlatore berezi deituko diogu, eta erlazonatutako hitza —oro har, gune semantikoa— zein den bilatzen lagunduko gaitu. Halako erlatoreek hainbat erlazio sortuko dituzte, bai paradigmaticoak, bai sintagmaticoak.⁴¹

ihar	A0	izond.	Hezea ez dena, bizitasuna galtzeraino lehortu dena.
------	----	--------	---

Aurreko adibidean bi erlazio ditugu, bata paradigmaticoak eta bestea sintagmaticoak. Paradigmaticoak ihar eta hezeren artekoak dugu: *antonimia*. Eta sintagmaticoak ihar eta lehorturen artekoak: lehortu (bizitasuna galtzeraino) iharra denaren *ezaugarria* dugu.⁴²

Erlatore bereziek sortuko dituzten erlazioak kasuz kasu aztertuko ditugu IV.4. atalean eta A. eranskinean.

⁴⁰ Aditzen artean *troponimia* erlazioa dugu. IV.3.2. atalean sakonduko dugu.

⁴¹ Beste kategorietan ere erlatore bereziak agertzen dira.

⁴² *Ezaugarria* bizitasuna galtzeraino lehortua bada ere, gure lanean lehenik ezagutuko dugu lehortua dela eta ondoren bizitasuna galtzeraino.

Batzuetan zailtasunak sor daitezke hitz bat erlatore edo genus den erabakitzeke garaian. Hasiera batean gune sintaktikoa eta semantikoa hitz bera dela badirudi ere, gunean dagoen hitz hori esanahiz oso ahula baldin bada, erlatore gisa landu dugu. Multzo honetan ditugu pertsona, tresna, egin, izan, eman, hartu, eta abar moduko hitzak. Genusen erregelak aplikatu ondoren, gune-posizioan agertzen diren esanahi ahuleko hitz hauek dira usuen agertzen direnak. Beraz, definizioetako gune-posizioan usuen agertzen direnak aztertu ditugu, eta esanahi ahula duten horiek erlatore berezi gisa landu ditugu (ikus 13. taula, 14. taula, eta 15. taula).

erretzaile	A0N1	iz. Zerbait erretzen duen pertsona .
grafologo	A0	iz. Grafologian diharduen pertsona .
hargin	A0	iz. Eraikuntzan harria lantzeko ogibidea duen pertsona .

Hiru kasu hauetan garrantzitsuena ez da pertsonak direla,⁴³ baizik eta batak erre egiten duela, besteak grafologiarekin duela zerikusia eta besteak harria lantzen duela.

Aditzen artean agertzen direnean, askotan hitz anitzeko unitate lexikal baten bigarren atala dira genus ahul horiek: hitz egin, laket izan, baimena eman, eta abar. Aditzetan agertzen zaizkigun hitz anitzeko unitate lexikal horiek erlazionatzeko, erlazio bereziak beharko ditugu: *hitz anitzeko sinonimia* eta *hitz anitzeko hiperonimia*. Erlatore-marka aditzak du eta erlazionatutakoaren marka aurreko izenak, baina definizioan duguna hitz anitzeko aditz bat denez, erlazioa bi aditzen artekoa da, eta, beraz, paradigmatikoa.

errekontziliatu	A0	ad. Adiskidetu; bakezkoak egin .
erroldatau	A0	ad. Erroldan izena eman .

errekontziliaturen kasuan bakezkoak egin da *hitz anitzeko sinonimoa*; eta erroldaturen hiperonimoa izena eman.

Hitz anitzeko elementu horiek azterketa hau egin aurretik detektatzeko gai izango bagina, ez genuke halako arazorik izango. Gaur egun hitz anitzeko horiek automatikoki ezagutzeko lanak burutzen ari dira Ixa Taldean (Alegria *et al.*, 2004).

⁴³ pertsona hori sarrera-hitzaren tasun semantikoa —edo hautapen-murriztapena— da.

Erlazio berezien zenbait tratamendu zehazten ari garenez, bada bi izenen arteko erlazio bitxia. Zenbaitetan izen bat beste izen baten *berezko objektua* dela adieraziko dugu.

erle	A1	iz. Intsektu himenopteroa, taldean bizi dena, eztia eta argizaria egiten dituen.
------	----	--

Kasu honetan esango dugu ezti eta argizari, erleren *berezko objektuak* direla. Hau aski berezia da; oro har, *berezko objektu* erlazioa aditz eta izen baten artekoa izango baitugu. Kontua da definizioan aditz ahul horietako bat dugula (egin aditza kasu honetan), eta hori da sarrera-hitzaren (erle) eta definizioan ditugun erlazonatutako hitzen (ezti eta argizari) arteko zubi-lana egiten duena. Halako erlazioak eratorpena azaltzerakoan agertuko zaizkigu (ikus V. atala).

Erlatore berezi batzuekin lortuko dugun erlazioa ez da aurretik zehaztutako bat izango, baizik eta erlatoreak berak adieraziko diguna. Eskema jakina dute definizio zenbaitek adjektibo eta izen baten arteko erlazioa agertzeko garaian, baina erlazio hori zein den erlatorearen arabera da. Halakoetan erlatore horretatik erauzi beharko dugu erlazio mota zein den. Halakoak adjektibo eta izenen arteko erlazioak izan ohi dira.

absolutista	A0	izond. Absolutismoaren aldek oa.
beltz	A1	izond. Batere kolorerik edo argirik ez duena, ikatzaren kolore koa.
kurbatu	A0	izond. Kakotua, kurba form akoa.

absolutista **eta** absolutismoaren arteko erlazioa aldekok zehazten du. beltz **eta** ikatzaren artekoa kolorek, **eta** kurbatu **eta** kurbaren artekoa formak (ikus IV.4.13. atala).

Beraz, atal honetan definizioaren gunea aztertu dugu, eta gune horren eta inguruan dituen elementuen arabera, sarrera-hitzaren eta definizioko elementu baten arteko erlazioa zehaztu dugu. Lana izenei, adjektiboei, izen-adjektiboei, aditzei eta azpisarrerei aplikatu diegu. Eskema horiek automatikoki ezagutzeko garaian, testuingurua kontuan hartzen duen *Murriztapen Gramatika* erabili dugu (ikus III.7. atala).

Kapituluko lehen hiru ataletan sinonimoak (IV.2. atala), genusak (IV.3. atala) eta erlatore bereziak (IV.4. atala) aztertu ditugu. Izen-adjektibo kategoriak dituzten elementuen azterketa IV.4.11. atalean egin dugu, eta IV.6. atalean azpisarreren

azterketa azaldu dugu. IV.7. atalean erregelak definitu eta ebaluatzeko metodologia erakutsi dugu eta baita ebaluazioari buruzko datuak eman ere. Amaitzeko, IV.8. atalean ondorioak eta etorkizuneko lanak aipatu ditugu.

Atal honekin lotuta A. eranskina eta I. eranskina daude. A. eranskinak IV.4. atala osatzen du eta erlatore berezien informazio zabala dago; I. eranskinean kategoria bakoitzean sinonimoak, genusak eta erlatore bereziak erauzteko egindako erregela guztiak daude (ikus I.1., I.2., I.3., I.4., I.5. eranskinak).

IV.2. Sinonimoak

Atal honetan, alde batetik sinonimoak agertzen diren definizio-patroien egitura azaldu dugu, eta, bestetik, horiek islatzen duten sakoneko erlazioa zein den. Sinonimiaren kasuan, dagoeneko esan dugun moduan, azaleko erlazioa eta sakonekoa erlazio bera da.

IV.2.1. Definizioen hezurdura

Dagoeneko esan dugun bezala, EHn ez dago erlazio *sinonimikoa* zuzenean agertzen duen eremurik, baina bada beste eremu bat, *ik.* eremua, sinonimoa adieraz dezakeena. Honezaz gain, kontuan izan behar dugu sarrera-hitzaren kategoria bereko hitz bakar batez —edo batzuez, komaz bereizita— osatutako definizioak lotura *sinonimiko* inplizitua islatzen duela gehienetan.

Sinonimo inplizitu horien egitura honelakoa litzateke *izenetan*, *aditzetan* eta *adjektiboetan*:

(5) Izenaren_definizioa = izena[, izena[, izena]].⁴⁴

(5) Asentsio H1A0 iz. Igokundea.

(5) agoñu A0 iz. Kemena, adorea.

(5) ahakar A0 iz. Liskarra, aharra, errieta.

Hemen ditugun hiru izen hauek *sinonimoen* bitartez definitzen dira hiztegian. Asentsioren *sinonimoa* Igokunde dugu; agoñuren *sinonimoak* kemen eta adorea dira; eta, azkenik, ahakarren *sinonimoak* liskar, aharra eta errieta dira.

⁴⁴ Parentesi karratu hauek *aukera* adierazten dute, hau da, definizio-eskema honetan adierazten dugu ager daitezkeela *sinonimo* bat, edo bi, edo hiru.

(6) Aditzaren_definizioa = aditz-partizipioa[, aditz-partizipioa[, aditz-partizipioa]].

(6) abadetu A2 ad. Apaiztu.

(6) jabetu A3 ad. Babestu, zaindu.

(6) arrobotu A0 ad. Ebatsi, ostu, lapurtu.

Aurreko hiru *aditz* hauek *sinonimoen* bitartez definituta daude. abadeturen *sinonimoa* da apaiztu; jabeturen *sinonimoak* babestu eta zaindu dira; eta arroboturen *sinonimoak*, ebatsi, ostu eta lapurtu.

(7) Adjektiboaren_definizioa = adjektiboa[, adjektiboa[, adjektiboa]].

(7) aldabera A0 izond. Aldakorra.

(7) sabelzuri A0 izond. Itxuratia, elezuria.

(7) agradagarri A0 izond. Atseгина, atsegingarria, gustagarria.

Hemen hiru *adjektibo* ditugu *sinonimoen* bitartez definituta. aldaberaren *sinonimoa* da aldakor; sabelzurik bi *sinonimo* ditu, itxurati eta elezuri; eta agradagarrik, aldiz, hiru: atsegin, atsegingarri eta gustagarri.

Beraz, eta laburbilduz, sarrera-hitz bat definitzerakoan kategoria bereko hitz bakarra erabiliz gero definizio-eremuan, hitz hori *sinonimotzat* hartuko dugu. Izen eta adjektiboen kasuan hitz hau *mugatu singularrean* agertzen da, eta, aditzen kasuan, partizipio mugagabea dugu.

Baina dena ez da hain hertsia; *sinonimoak* azaldu ohi dira deskribapenekin batera, bai hauen aurretik, bai hauen atzetik.⁴⁵

(8) Izenaren_definizioa = izena, izenaren_deskribapena.

(9) Izenaren_definizioa = izenaren_deskribapena, izena.

(10) Aditzaren_definizioa = aditz-partizipioa, aditzaren_deskribapena.

(11) Aditzaren_definizioa = aditzaren_deskribapena, aditz-partizipioa.

(12) Adjektiboaren_definizioa = adjektiboa, adjektiboaren_deskribapena.

(13) Adjektiboaren_definizioa = adjektiboaren_deskribapena, adjektiboa.

Hona aurreko kasuen zenbait adibide:

(8) almirante H2A0 iz. Amirala, itsas armadako buruzagi gorena.

(9) ekarpen A3 iz. Askoren arteko egiteko bati inork ekartzen diona, ekarria.

(10) balaztatu A0 ad. Galgatu, ibilgailu edo makina bat geldiarazi edo haren lastertasuna gutxitu.

(11) adorarazi A0 ad. Adoratzera behartu, gurrarazi.

⁴⁵ Definizio hauen kasuan lehen komaren aurrean dagoena —deskripzioaren aurretik dagoena, alegia—, edo azken komaren atzetik dagoena —deskripzioaren atzetik dagoena, alegia—, izaten da *sinonimoa*. Zenbaitetan, definizioaren ondoren ez da *sinonimo* bakarria agertzen, baizik eta bi.

- (12) maitati A0 izond. Maitekorra, maitasuna sentitzen edo adierazten duena.
- (13) baketsu A1 izond. Bakea maite duena, liskarra gogoko ez duena, bakezkoa.

Definizioei erreparatuz gero ikus daiteke almiranteren *sinonimo* dela amiral eta ondoren deskribapena agertzen dela, eta ekarpenen *sinonimo* dela ekarri eta aurretik deskribapena agertzen dela. *Aditzen* definizioen adibide ditugu balaztatu eta adorarazi; balaztaturen *sinonimo* da galgatu eta ondoren deskribapena du, eta adoraraziaren *sinonimo* da gurrarazi eta deskribapena du aurretik. Azkenik, adjektiboen adibideei erreparatuta, maitatiaren *sinonimo* da maitekor eta ondoren du deskribapena, eta baketsuren *sinonimo* da bakezko eta deskribapena aurretik du.

Beraz, definizio-eremuan *koma* baten ondoren hitz *absolutibo mugatu singularra* baldin badugu —izen eta adjektiboen kasuan— edo *partizipio mugagabea* —aditzen kasuan—, eta ondoren *definizio-bukaera*, hitz hori sarrera-hitzaren *sinonimoa* da (ikus (9), (11) eta (13)ko adibideak).

Definizio-hasieretan, berriz, legea ez da hain zorrotza. Izenen kasuan argi izan behar dugu definizio-eremu hasieran *komaren* aurretik *izen mugatu singularra* agertzeak ez duela esan nahi sarrera-hitzaren *sinonimo* denik. Ustezko *sinonimo* honen hurrengo hitza *absolutibo mugatu singularrean* dagoen beste izena baldin bada, eta hitz-zerrenda baten parte baldin bada, ezin esango dugu lehen *izen* hori sarrera-izenaren *sinonimo* denik. Honelakoetan ia ziur enumerazio bat izango dugu.

(14) Izenaren_definizioa = izen_abs_mugatua, izen_abs_mugatua, [izen_abs_mugatua, [izen_abs_mugatua,]] izenaren_deskribapena.

- (14) tilet A0 iz. **Puntua, koma, azentua** edo kideko zernahi idazkera marka.
- (14) modelo A0 iz. **Eskultura, eraikuntza, tresna...** baten eredua, gehienetan tamaina txikiagoan egiten dena.

Bi adibide hauetan ezin esango dugu puntu eta koma, tilet sarreraren *sinonimo* direnik; ezta ere modeloren *sinonimo* direnik eskultura, eraikuntza eta tresna.

Hala ere, zenbaitetan gertatzen da bi *sinonimo* izatea eta ondoren deskribapena azaltzea. Hauek oraingoz ez ditugu automatikoki ezagutuko, alde batetik modu honetakoak asko ez direlako, eta, bestetik, hauek ezagutzearren sortuko genukeen errorea handia litzatekeelako.

(15) Izenaren_definizioa = sinonimoa, sinonimoa, izenaren_definizioa.

(15) tronu A0 iz. **Jargoia, jauralkia**, errege aulkia.⁴⁶

(15) samats A0 iz. **Simaurra, ongarria**, bereziki landare hondakinez egiten dena.

Aurreko adibideetan ez bezala, hauetan tronuren *sinonimo* dira jargoi eta jauralki, eta samatsen *sinonimo* simaur eta ongarri.

Aditzen eta adjektiboen definizioetan ez da mota honetako enumeraziorik agertzen.

IV.2.2. Semantika

Hona *sinonimiaren* definizio zabala: zentzu bera edo bertsua, eta forma desberdinak dituzten hitzak dira *sinonimoak*. Honek bide ematen die hiztegigileei, zenbait testuingurutan beste baten ordeztar jar daitezkeen hitzen zerrenda luzeak emateko. Hala azaltzen du Miller-ek eta sinonimia kontzeptua hala erabiltzen da *WordNet* datu-basean. Miller-ek (1998a) esaten duenez datu-base honetan *sinonimotzat* hartzen diren hitzak zenbait testuingurutan aldagarri dira. Ez da beharrezko izango agertzen diren testuinguru guztietan aldatu ahal izatea.

WordNet does not entail interchangeability in all contexts; by that criterion, natural languages have few synonyms. The more modest claim is that WordNet synonyms can be interchanged in some contexts. To be careful, therefore, one should speak of synonymy relative to a context, but in order to facilitate the discussion this qualification will usually be presupposed, not asserted. (24. or.)

Hau Sparck-Jones-ek (1964; 1986) zioenaren bidetik doa. Honek zioenez bi hitz *sinonimo* izango dira hirugarren bat bien ordeztar erabil badaiteke, eta adibide hau jartzen zuen: *pellet* eta *injection* berdintsu dira zeren, kontestu egoki bat bilatuz gero, biak aldatu ahalko baitira *shot* hitzarekin.

Sinonimia, kontzeptuak azaltzeko, hiztegietan asko erabiltzen den teknika da. Ikusteko teknika honek zein garrantzia duen EHn, laginak hartu ditugu —100 sarrera kategoria bakoitzeko— eta hona emaitzak: izenen laginean 49 dira *sinonimoa* duten definizioak, aditzen artean 44, eta adjektiboen artean 38.

⁴⁶ Kasu honetan, gainera, *errege aulki* deskribapen bat ote den ez dago batere garbi, izen konposatu gisa har baitaiteke.

	lagineko definizioak	sinonimodun definizioen kopurua
izenak	100	49
adjektiboak	100	38
aditzak	100	44
denera	300	131

12. taula. Sinonimoen bitartez emandako definizio-kopurua.

Emaitza hauei erreparatuz gero, nabaria da *sinonimoen* erabilera hitzak definitzeko garaian; ia definizioen erdietan agertzen baita *sinonimoren* bat.

IV.3. Genusak

Atal hau ere bi zatitan banatu dugu. Lehenengo zatian genusdun definizio-patroi zenbait ikusiko ditugu. Aztertutako definizio-patroi guztien erregelak I. eranskinean paratu ditugu. Bigarren zatian *hiperonimia* / *hiponimia* zer den ikusiko dugu. Genusak sakoneko erlazio bakarraren isla dira.

IV.3.1. Definizioen hezurdura

IV.2.1. atalean ikusi dugun bezala, sarrera-hitzaren deskribapenak *sinonimoekin* batera ager daitezke, baina definizio batean ez dugu zertan beti *sinonimo* bat izan. Zenbaitetan deskribapen linguistikoa besterik ez da agertuko.

Gainera, hitz baten definizio tipikoaren deskribapena egiten denean, hau da eskema nagusia: definizioaren gunea —*genusa*— definitutako hitzaren kategoria bereko hitz bat izaten da, diferentzia partea —*differentia*— osatzen duen modifikatzaile edo osagai baten ondoan etorri ohi dena. Genusa da definizioaren gune sintaktiko eta semantikoa. Adibidez, hona genusdun definizio-patroi bat:

(16) Izenaren_definizioa = [...] + izenlaguna + **izena** + puntuazio-marka.

Honelako definizio batean, *izenlagunaren* eta *puntuazio-markaren* artean —ohikoena puntua da— dagoen *izena* genusa izan ohi da.

(16) urmargo A0 iz. Uretan deseginiko koloreez paper gainean
ontzen den **margolana**.

(16) zuberera A0 iz. Zuberoako **euskalkia**.

Definizio hauetan urmargoren genusa dugu margolan eta zubererarena euskalki.

Hala ere, eta genusen eskema ohikoena (16)koa baldin bada ere, baditu beste hainbat eskema; genusak definizioko azken hitzaren lekua ez ezik, beste lekurik izan

dezake. Definizioko azken hitza ez denean, ohikoena da ondoren *adjektiboa(k)* agertzea (17); edo genusa den hitza *mugagabea* agertzea, aurretik *determinatzailea* duelako (18); edo ondoren *adjektiboa* bakarrik izan beharrean, *maila-adberbio* bat eta *adjektiboa* izatea (19); edo genusak ondoren *juntagailu* bat izatea, honen atzetik beste genusa duelako (20); edo aurrekoen nahasketa bat gertatzea (21). Hona, ohikoak diren zenbait kasu ekarri ditugu, eta egindako erregela guztiak I. eranskinean paratu ditugu.

(17) Izenaren_definizioa = (...) + izenlaguna + **izena** + adjektiboa(k) + puntuazio-marka.

(18) Izenaren_definizioa = (...) + determinatzailea + **izen_mugagabea** + puntuazio-marka.

(19) Izenaren_definizioa = (...) + izenlaguna + **izena** + adberbioa + adjektiboa + puntuazio-marka.

(20) Izenaren_definizioa = (...) + izenlaguna + **izena** + juntagailua + **izena** + puntuazio-marka.

(21) Izenaren_definizioa = (...) + izenlaguna + **izena** + juntagailua + **izena** + adjektiboa + puntuazio-marka.

(17) almirante H2A0 iz. Amirala, itsas armadako **buruzagi** gorena.

(17) gerezi A1 iz. Gereziendoak ematen duen **fruitu** txiki mardul eta hezurduna, gehienetan azal gorri iluna duena.

(18) iratze A2 iz. Pteridophyta taldeko edozein **landare**; bereziki Filicopsida klaseko edozein **landare**.

(19) kolibri A0 iz. Ameriketako **txori** guztiz txikia, moko luzea eta lumak distiratsuak dituena.

(20) kukurruku A1 iz. Oilarraren **oihua** edo **kantua**.

(21) legartza A0 iz. **Toki** edo **eremu** legartsua.

Adibide hauetan almiranteren genusa buruzagi da, gerezirena fruitu, iratzerena landare, kolibrirena txori, kukurrukurena oihi edo kantu, eta legartzarena toki edo eremu.⁴⁷

Nabaria da genusa dela definizio-eremuan dugun sintagmaren buru sintaktikoa. Buru horrek inguruan hainbat modifikatzaile izan ditzake, baina —lanaren puntu honetan— gure helburua buru hori ezagutzea da. Erregela hauek guztiak egin behar izatea saihestu ahal izango guke sintagmak eta hauen guneak automatikoki ezagutzeko gai izango bagina, eta horretarako desanbiguazio sintaktikoa behar beharrezkoa da. Sarreran aipatu dugu lan honetan ez dugula desanbiguazio

⁴⁷ Definizio hauetan lortzen dugun bakarra ez da genusaren informazioa; almiranteren kasuan badakigu sinonimoa amiral dugula. gereziren kasuan azala izango da erlazonatutako hitza. kolibriaren kasuan erlazonatutako hitzak moko eta luma izango dira. Eta kukurrukuren kasuan oilar.

sintaktikorik erabiliko emaitza hobeak lortzen ditugulako analisi anbiguoekin, baina erabaki honek erregela gehiago egitera behartzen gaitu.

Aditzen kasuan gunea beti izango dugu *partizipioa*, eta definizioan ager daitezkeen aldaerak askozaz txikiagoak dira; honexegatik, guneak markatzeko bi erregela besterik ez ditugu erabiltzen.

(22) Aditzaren_definizioa = [...] **aditz-partizipioa** + puntuazio-marka.

(23) Aditzaren_definizioa = [...] **aditz-partizipioa** + juntagailua + **aditz-partizipioa** + puntuazio-marka.

(22) asmatu A1N2 ad. Tresna berri bat **sortu**.

(23) beheratu A0 ad. Beherago **joan** edo **eraman**; (...).⁴⁸

asmatu sarreraren kasuan, honen genusa *sortu* da, *puntuazio-markaren* aurretik dagoen *aditz-partizipioa*, alegia —kasu honetan *puntuazio-marka* hori *definizio-bukaera* ere bada—; beheraturen kasuan, bi genus ditugu *juntagailu* baten bitartez lotuta: *joan* eta *eraman*.

Adjektiboak definitzeko *sinonimia* erabiltzen da, eta deskribapenak ditugunetan, erlatore bereziak erabiltzen dira hauek deskribatzeko. Erlatore bereziei IV.4. atala eskaini diegu.

Adjektiboen kasuan ez zaigu genusa agertuko —eta agertzen denetan pentsa dezakegu *izen* baten definizioa dugula esku artean—. Genus moduko hauek harrapatzeko izenetan dugun erregularik orokorrena besterik ez dugu aplikatuko (16) —*definizio-amaieran* agertzen denekoa, alegia—; izan ere, ez da ohikoena, eta kategoria ebazterako garaian mugan daudenekin besterik ez da gertatzen. Hauetan suposatu ahal izango dugu, nahiz eta zehatz-mehatz esan kategoria adjektiboa dela, izen bat dugula begi bistan —gutxienez ezin ukatuzkoa da, izen bati dagokion definizio mota duela—. ⁴⁹

geografo A0 izond. Geografiaren irakaskuntzan edo ikerketan aditua den **pertsona**.

espioitza A0 izond. Espioi **lana**.

Bi adibide hauei kategoria gaizki esleitu zaie.

⁴⁸ Eskema hau asko errepikatzen da *da* / *du* formak dituzten aditzen definizioetan.

⁴⁹ Halakoak detektatzea oso interesgarria izan daiteke, halako asko *erratak* direlako.

IV.3.2. Semantika

Genusa den hitza sarrera-hitzen *hiperonimoa* dugu, eta sarrera-hitza, aldiz, genusa den hitzen *hiponimoa* da. Beraz, erlazioa *hiperonimia* / *hiponimia* da.

aireontzi A0 iz. Hegalda daitekeen zernahi **ibilgailu**.

Hiperonimia da termino zehatzago bat orokorrago batean harrapatzen duen erlazioa. aireontziaren *hiperonimoa* ibilgailu da; eta, ibilgailuren *hiponimoetako* bat, aldiz, aireontzi; izan ere, ibilgailuren beste *hiponimo* bat jeep da, ondoko adibidean ikusten dugun bezala.

jeep A0 iz. **Ibilgailu** sendoa, txikia eta arina, edozein lekutatik ibiltzeko gai dena.

Miller-ek (1998a) aipatzen duenez, *izenak* antolatzeko garaian hau da erlaziorik garrantzitsuenak. Subordinazio-erlazioa da, eta kontzeptu batzuk besteen barruan agertzen dira.

the semantic relation that is most important in organizing nouns is a relation between lexicalized concepts. It is the relation of subordination (or class inclusion or subsumption), which in this context we will call *hyponymy*. For example, the noun *robin* is a hyponym (subordinate) of the noun *bird*, or, conversely, *bird* is a hypernym (superordinate) of *robin*. It is this semantic relation that organizes nouns into a lexical hierarchy. (24 or.)

Iragankortasuna da erlazio honek duen ezaugarri garrantzitsuenak. Hau da, baldin eta A *Bren hiperonimoa* bada, eta B eta Cren, orduan A ere Cren *hiperonimoa* da.

ibilgailu A0 iz. Lurrean, uretan edo airean higitzen den eta pertsonak edo gauzak garraiatzeko balio duen edozein **tresna**.

Alegia, definizioen arabera baldin badakigu ibilgailua tresna bat dela, eta aireontzia ibilgailua dela, orduan, ezagutu ahal izango dugu aireontzia tresna dela.

Aditzen kasuan *hiperonimia* / *hiponimia* erlazioa erabili beharrean, *troponimia* / *hiperonimia* erlazioa dagoela esan ohi da.

ziatu A0 ad. Atzerantz *arraun egin*, atzerantz *bogatu*.

Kasu honetan, ziatu sarrerak bi *hiperonimo* izango ditu, egin (arraun) eta bogatu. Eta, ziatu, berriz, bogaturen *troponimoa* da.

Miller-ek (1998a) hau azaltzeko esaten du goitik beherako erlazio hau ez dela berdina izenetan eta aditzetan. Izenetan test hau erabiltzen dute

An X is a kind of Y (79. or.)

Lyons-ek (1977) ere esaten duenez, aditzak ezin daitezke patroï honetan sartu. Beraz, beste test bat bilatu behar da. Talmy-k (1985) dioenez, *modua* izango da gakoa. Fellbaum-ek eta Miller-ek (1990) kontu honi erreparatuta *troponimia* hitzaren alde egiten dute.

In WordNet the many different kinds of elaborations that distinguish a "verb hyponym" from its superordinate have been merged into a manner relation that Fellbaum and Miller (1990) have dubbed *troponymy*. The troponymy relation between two verbs can be expressed by the formula *To V₁ is to V₂ in some particular manner*. (79. or.)

Hiperonimia / hiponimia bezalaxe, erlazio hau ere iragankorra da.

Bai *izenetan* eta bai *aditzetan* hierarkia bat osatuko dugu, eta hierarkia horretan ditugun elementuek goragokoen ezaugarriak heredatu ahal izango dituzte.

kolibri	A0	iz. Ameriketako txori guztiz txikia, moko luzea eta lumak distiratsuak dituen.
txori	H1A0	iz. Tamaina txikiko hegaztia .
hegazti	A0	iz. Animalia ornodun odol-beroa, gorputza lumaz estalia duena, mokoduna, hegoak eta hankak dituen, eta oro har hegan egiteko egokitua dagoena.

kolibria txoria baldin bada, eta, txoria hegaztia baldin bada, hegazti honek dituen ezaugarri guztiak heredatuko ditu kolibriak, hala nola, animalia ornoduna dela, gorputza lumaz estalia duela —lumei buruzko informazio zerbait badugu bere definizioan, distiratsuak ditu—, mokoduna dela —bere definiziotik badakigu luzea dela moko hori—, hegoak eta hankak dituen, eta oro har hegan egiteko egokitua dagoela.

Genusak identifikatzeko egindako lan hau —*hiperonimiak* duen iragankortasun ezaugarriari esker— dagoeneko erabili da itzulpen automatikoa egiteko garaian (Díaz de Ilarraza *et al.*, 2000a). Artikulu honetan izen-sintagmak ingelesetik edo espainieratik euskarara itzultzeko ahalegina agertzen da, eta, zehatz-mehatz, izen baten modifikatzaile den genitibo bat itzultzeko ahalegina. Beste hizkuntza bateko genitibodun forma euskarara itzuli nahi izanez gero, bi aukera ditugu: *lekuzko genitiboa* edo *edutezko genitiboa* erabiltzea. *Genitiboa* daraman hitza *biziduna* baldin bada, *edutezkoa* hartuko du. Beraz, jakin beharko dugu zein hitz diren *bizidun*, eta horretarako genusen informazioa erabiliko dugu. Hitz bat *biziduna* baldin bada,

haren *hiponimo* guztiak *bizidunak* izango dira. Beraz, genusaren informazio hau itzulpen automatikorako baliagarria dela konprobatu da.

Hala ere, genus diren hitzen artean ezberdintasun handiak topa ditzakegu; ondoko atalean, hain zuzen ere, genus mota berezia azaldu dugu.

IV.3.3. Genus orokorrak

Badira esanahi aldetik oso orokorrak diren genus batzuk. Halakoek ez dute apenas informaziorik ematen eta Bruce *et al.*-ek (1992) *disturbed heads* deitzen dituzte. Lan horretan agertzen den bezala, horiek erlatore berezi bezala tratatzea komeni da. Zerbait *multzo* bat dela esateak ez du asko lagunduko, eta benetan informazio berria eta erabilgarria emango diguna izango da zerbait *zeren multzo* den jakiteak.

helize Al iz. Ardatz baten inguruan biratzen diren eta
 indarra eragiten duten palen **multzoa**.

Definizio honetatik *helize* hori *multzo* bat dela jakiteak ez gaitu lagunduko ezagutza gehiago izateko orduan, eta benetan lagunduko gaituena *helizeak* palak dituela jakiteak izango da. Horregatik *multzo* hitza erlatore berezien artean sartu dugu, eta IV.4. atalean azaldu dugu.

Baina genus huts hauek izan ohi dira definizioetan gehien agertzen direnak —hau definizioz hala da; izan ere, hitz bat definitzeko esanahi orokorragoa duen beste hitz bat erabili ohi da—, eta lehen-lehenik saiatuko gara jakiten zein diren definizioetan gune posizioan agertzen diren hitz usuenak. Hitz hauek detektatzeko, genusak ezagutzeko sortu ditugun erregela guztiak aplikatu ditugu. Hala, definizioen gune sintaktikoan agertzen diren hitzen maiztasuna lortuko dugu. Zerrenda hori izan eta gero, usuen agertzen direnen esanahia aztertu dugu, eta esanahi ahulena dutenak *genus orokortzat* hartu ditugu eta erlatore berezi gisa tratatu.

13. taulan izenen definizioetan gune sintaktikoaren posizioan usuen agertzen diren hitzak paratu ditugu.

Genusa	Kopurua	Genusa	Kopurua	Genusa	Kopurua
nolakotasuna	555	tokia	116	lana	67
multzoa	426	zera	115	ontzia	66
pertsona	385	landarea	110	eza	65
tresna	219	denbora	105	etxea	64
egoera	197	lekua	102	bitartea	63
zattia	180	aldea	91	kidea	57
izena	165	zati	87	jakintza	55
gauza	141	lanbidea	87	kopurua	53
mota	132	era	79	burua	53
egitea	128	aldia	75	neurria	52
bidea	118	Gai	68		

13. taula. *Izenen artean topatutako genus usuenen zerrenda.*

14. taulan aditzen definizioetan gune sintaktikoaren posizioa duten hitz usuenen zerrenda paratu dugu.

Genusa	Kopurua	Genusa	Kopurua	Genusa	Kopurua
egin	472	jarri	161	bete	73
bihurtu	468	galdu	110	ezarri	71
behartu	426	kendu	93	utzi	63
eman	297	joan	86	jo	59
hartu	269	eraman	75	estali	51
izan	196	gertatu	74		

14. taula. *Aditzen artean topatutako genus usuenen zerrenda.*

15. taulan izen eta adjektibo diren hitzen definizioen artean gune sintaktikoaren posizioan usuen agertzen diren hitzak paratu ditugu.

Genusa	Kopurua
biztanlea	106
pertsona	22
aldekoa	19
kidea	15
izena	9
zera	8

15. taula. *Izen eta adjektibo direnen artean topatutako genus usuenen zerrenda.*

Hauetan oso genus gutxi agertu dira; izan ere, adjektibo bezala definitutako hitzek ez dute genusik.

Aurretik esan bezala, aztertu beharko da ea hauetako batzuek erlature berezien lana betetzen duten, eta ea definizio horietatik informazio gehiago atera dezakegun. Hala, erlature bereziei helduko diegu.

IV.4. Erlature bereziak

III.8. atalean ikusi dugu *token*-segida bat, edo hitz bat, zenbatetan agertzen den hiztegian. Horietako zenbait erlature berezizat joko ditugu. Zerbait erlaturetzat jo

edo ez erabakitzeke garaian aztertu behar da ea erlazio lexiko-semantikoren baten isla den; alegia, ea erlazioan jartzen dituen inguruko hitzen bat eta sarrera-hitza.

Adibidez, elkarrekin agertzen diren bikoteen azterketan ikusi dugu honoko *token-segida* 295 aldiz agertzen dela:

, eta

Baina “, eta” segida askotan agertzeak ez du esan nahi erlatore berezia denik, baizik eta bi elementu horiek askotan agertzen direla elkarrekin definizioen barruan. Gainera zerbait erlatore berezi izan edo ez erabakitzerakoan garrantzia izango du ondoren *puntuazio-marka* agertzeak. Puntuazio-markaren aurreko lekua izan ohi da genus diren hitzen posizioa —deskribapenaren amaiera, alegia—, eta erlatore berezi hauek genus horien lekua —fisikoa— beteko dute.

Erlatorea, beraz, genusaren posizioan agertzen den elementua da, eta inguruko hitz bat —erlazonatutako hitza deituko duguna— eta sarrera-hitza jartzen ditu harremanean.

Hona hiztegian maiz agertzen diren bikote batzuk:

bikote-segida	kopurua
duena .	1.721
dena .	1.307
dituena .	549

16. taula. Bikote-segida usuak.

duenaren, denaren eta dituenaren ondoren *puntua* dugu. Honek *definizio-amaiera* adierazten du. Hiru elementu horiek bere barruan *elipsia* dute, eta genusa beharko lukeen elementua elidituta dago. Beraz, *elipsia* duten elementuak erlatore berezizat hartuko ditugu, eta IV.4.11. atalean aztertuko dugu nolako erlazioa islatzen duten.

Azterketa estatistiko honetan agertzen diren elementu horiez gain, aurreko atalean esan dugun bezala, genus orokorrak direnak ere erlatore berezi gisa aztertuko ditugu.

Orain arte kapitulu honetan ikusi ditugun erlazio lexiko-semantikoak —*sinonimia* eta *hiperonimia*—, erlazio paradigmaticoak direla esan ohi da, kategoria bereko hitzen artean gertatzen direlako. Erlatoreak aztertzen hasita, erlazio paradigmaikoez gain, konturatuko gara hauek markatzen dituzten zenbait erlazio kategoria ezberdinen arteko elementuen artekoak direla, eta erlazio hauei, sarreran esan dugun bezala, erlazio sintagmatiko deitu ohi zaie (Artola, 1993).

Ondoren, erlatoreak sailkatuko ditugu sortzen dituzten erlazioen arabera, eta horien ezaugarriak atalez atal ikusiko ditugu. Atal bakoitzean ezaugarri zenbait, adibideren bat eta laburpen taula jarri dugu. Adibide gehiago A. eranskinean paratu ditugu, eta erlatoze bereziak eta erlazonatutako hitzak ezagutzeko egindako gramatikak I. eranskinean daude, kategoriaz kategoria banatuta.

Hemen paratutako laburpen-aula horietan dugun informazioa honokoa da: lehenengo zutabea erlatozeari gramatikak jarriko dion marka —*etiketa*— agertzen da, eta bigarren zutabea —*erlazioa*— sortzen den erlazioaren izena. *sar-kat* zutabea sarrera-hitza zein kategoriatakoa den agertuko da, eta *erlz-kat* zutabea erlazonatutako hitzaren kategoria. *def-eskema*k definizioetan agertzen den patroia edo eskema orokorrenaren berri emango digu. *erlt-kop* eta *erlz-kop* zutabeek hiztegiko definizioen artean erlatoze bat edo erlazonatutako hitz bat zenbatetan detektatu dugun adieraziko dute.

Adibidez, *-i eman izena* erlatozearen kasuan, hona definizio-eskema eta adibidea:

(24) Izenaren definizioa = izena (datiboa) + ematen z(itz)ai[e/o]n izena.

(24) elizate. A0 iz. Bizkaiko infanzonado deritzan alderdiko
herriei ematen zaien izena.

Definizio-eskemak dioenez, datiboa daraman *izena* agertu behar du, eta ondoren *ematen zaion* —edo *zaien*, *zitzaion*, *zitzaien*— *izena*.

Eta hona etiketa eta erlazioak agertzen dituen taula:

etiketa	erlazioa	sar-kat	erlz-kat	def-eskema	erlt-kop	erlz-kop
-I EMAN IZENA	hiperonimia	IZE	IZE	-i ematen zaion izena	65	60

17. taula. *-i eman izena* erlatozearen laburpen-aula.

Mota honetako erlatozea 65 aldiz markatu dugu —*erlt-kop*—, eta erlazonatutako hitza 60 aldiz —*erlz-kop*—.

IV.4.1. Ia-Sinonimia

Erlazio hau *sinonimiarengandik* aski hurbil dago (ikus IV.2. atala). *Ia-sinonimoen* artean, bi mota ditugu:

- Batzuetan *sinonimia* adierazten duen eskema dugu definizioan, baina bertan agertzen diren elementuak eta sarrera-hitza kategoria ezberdinekoak dira. Hitz

horien kategoria ia ziur muga-mugakoa izango da, eta, zenbaitetan izen, zenbaitetan adjektibo izango dira.

- Besteetan bi hitz izango ditugu definizioan, *graduatzaila* eta *ia-sinonimo* deituko duguna.

Azken honi buruz, Miller-ek (1998b) dio erlazio semantiko gisa har daitekeela, baina, adjektiboen kasuan, askotan zaila dela zehazten zein den *adjektibo* ez-markatua:

Gradation, therefore, must also be considered as a semantic relation organizing lexical memory for adjectives (Bierwish 1989). For some attributes gradation can be expressed by ordered strings of adjectives, all of which point out to the same attribute noun in WordNet. (...) Gradation is usually accomplished in other ways, much of it by means of adverbs of degree, such as *very*, *decidedly*, *extremely*, *rather*, *quite*, *somewhat* (Cliff, 1959). But most grading is done by morphological rules for the comparative and superlative degree, which can be extended if *less* and *least* are used to complement *more* and *most*. Since this conceptually important relation of gradation does not play a central role in the organization of adjectives, it has not been coded in WordNet. (53. or.)

Testuan azaltzen dutenez, azkenean erabaki zuten *WordNet* datu-basean halakorik ez jartzea. Guk esango dugu *ia-sinonimo* direla, eta *familiartasuna* dagoela bi *adjektibo* horien artean.

(25) Izenaren_definizioa = adjektibo_mugatua.

(26) Adjektiboaren_definizioa = adberbioa + adjektibo_mugatua.

(25) aroi A0 iz. **Alferra, nagia.**

(26) barrabas A0 izond. Oso **bihurria.**

etiketa	erlazioa	sar-kat	erlz-kat	def-eskema	erlt-kop	erlz-kop
ERLT-ADJ-SIN	ia-sinonimia	IZE	ADJ	<i>adj.</i>	82	—
IA-SIN	ia-sinonimia	ADJ	ADJ	adb. + <i>adj.</i>	32	32

18. taula. *Ia-sinonimia* erlazioa markatzen duten erlatoreak.⁵⁰

IV.4.2. Antonimia

Bi hitz *antonimo* direla esango dugu baldin eta kontrakoa adierazten badute. Adjektiboen artean agertzen den erlazioa da, eta aipatu izan denez, adjektibo deskribatzaileen artean inportanteenetakoa (Miller, 1998b):

⁵⁰ *erlt-adj-sin* kasuan erlatorea eta erlazonatutako hitza makak hitz berean jartzen ditugu.

the basic semantic relation among descriptive adjectives is
antonymy. (48. or.)

Gure kasuan, sarrera-adjektiboa eta definizioan agertzen dena *antonimoak* izango dira.

(27) Adjektiboaren_definizioa = adjektibo_mugatua + “ez dena”.

(27) erlatibo A0 izond. Beste zerbaiten arabera hartzen dena,
absolutua ez dena.

etiketa	erlazioa	sar-kat	erlz-kat	def-eskema	erlt-kop	erlz-kop
ANT	Antonimia	ADJ	ADJ	adj. ez dena	60	—

19. taula. *Antonimia* erlazioa markatzen duten erlatoreak.⁵¹

Lehen aipatu dugu, hiztegiaren ezaugarriak aipatzerakoan (III.2. atalean), hiztegiko definizioetan badela eremu bat *antonimoa* esplizitu egiten duena. *ant.* eremua dugu, eta hortik informazioa inolako erregularik gabe aterako dugu.

Adibidez, arin sarreraren lehenengo adieran —Pisu edo astuntasun gutxikoa, erraz jasotzen dena— badugu eremu hau, eta esaten zaigu bere *antonimoa* astun dela.

arin. izond. (1571) **1.** Pisu edo astuntasun gutxikoa, erraz jasotzen dena. **Ant. astun.** (...)

IV.4.3. Hiperonimia / hiponimia

Erlazio hau ere azaldu dugu IV.3.2. atalean, eta honekin lotuta dauden erlatoreak ikusiko ditugu atal honetan.

(28) Izenaren_definizioa = izena-markarik gabe + ‘modukoa’.

(28) besare. A0 iz. Bost, zazpi edo bederatzi hortz V bat eratuz dituen *golde modukoa*.

etiketa	erlazioa	sar-kat	erlz-kat	def-eskema	erlt-kop	erlz-kop
MODUKOA	hiperonimia	IZE	IZE	-en modukoa	91	92
-I EMAN IZENA	hiperonimia	IZE	IZE	-i ematen zaion izena	65	60
IZENA	hiperonimia	IZE	IZE	-en izena	103	53
BAT / BAKOITZA	hiperonimia	IZE	IZE	-etako bat / bakoitza	423	265
MOTA	hiperonimia	IZE	IZE	ize + mota	135	100

20. taula. *Hiperonimia* erlazioa markatzen duten erlatoreak.⁵²

⁵¹ *antonomiaren* kasuan erlatorea eta erlazonatutako hitza makak hitz berean jartzen ditugu.

⁵² Normalean, erlatore-kopurua erlazonatutako hitzen kopurua baino handiagoa izan ohi da, erlatorea harrapatzea errazagoa delako. Adibidez:

IV.4.4. Ezaugarria

Halako erlazioa dugunean pentsatuko dugu sarrera-hitzak izango duela erlazionatutako hitzak adierazten duen *ezaugarriren* bat. Oro har, eta Miller-ek (1998a) ere hala dio, ezaugarrien balioak *adjektiboek* emango dituzte:

Values of attributes are expressed by adjectives. (40. or.)

Erlazioa izen eta adjektibo baten artekoa izan ohi da, baina, zenbaitetan ikusiko dugunez, bi izenen artean ere gertatuko da.

WordNeten erlazio hau zenbaitetan jaso dute, baina ez modu orokorrean (Miller, 1998a).

It might be possible to incorporate into WordNet an association between {robin} and {red} that would constitute a kind of selectional restriction for the adjective *red*. Such associations have not been incorporated, since they are semantically complex. What is straightforward, however, is the connection between the attribute and the adjectives that express values of that attribute: between the noun *size* and the adjectives *large* and *small*; between the noun *color* and the adjectives *red*, *yellow*, *green* and so on. Many of those associations have been incorporated into WordNet. (40. or.)

WordNeten topatuko duguna zenbait izenek izan ditzaketen *ezaugarriak* dira. Guk hiztegitik ateratako *ezaugarri* erlazio honetan, gehiago izango dugu izen batek *berez duen ezaugarria* edo *egoera baten ezaugarria*.

(29) Adjektiboaren_definizioa = izena-instrumentala + 'betea'.

(29) goibeltsu A0 izond. *Goibeltasunez bete*a; hodeitsu.

agata A0 iz. *Silize* kristaldu **mota**, bere edertasunagatik harri bitxitzat erabiltzen dena.

agata sarrera-hitzaren definizioan erregelen bitartez gai gara *mota* erlatorea ezagutzeko, baina ez *silize* erlazionatutako hitz gisa, tartean kristaldu agertzen delako eta erregetan ez dugulako partizipioaren presentzia aurreikusi.

Baina, *modukoa* erlatorearen kasuan erlazionatutako hitzen kopurua handiagoa da, erlatore batek, zenbaitetan, erlazionatutako bi hitz markatzen dituelako. Hona adibidea:

borda A0 iz. Basetxe arrunta; bereziki mendian edo basoan, abereak edo belarra gordetzeko *etxola* edo *etxe modukoa*.

borda sarrera-hitzaren definizioan *modukoa* erlatorea dugu, eta erlazionatutako hitzak *etxola* eta *etxe* dira.

etiketa	erlazioa	sar- kat	erlz- kat	def-eskema	erlt- kop	erlz- kop
NOLAKOTASUNA ⁵³	ezaugarria / edukitzaile	IZE	ADJ	adj. denaren nolakotasuna	548	504
	ezaugarria / edukitzaile	IZE	IZE	ize. denaren nolakotasuna		
NOLAKOTASUNA EZ	ezaugarria- gabezia	IZE	IZE	ize. ez denaren nolakotasuna	9	9
-Z BETEA	ezaugarria / edukia	ADJ	IZE	-z betea	141	141
	ezaugarria / edukia	IZE	IZE	-z betea	7	7
ADJ-KOA	ezaugarria	ADJ	IZE	ize. + adj.-koa	105	98
-I EGOKI	ezaugarria	ADJ	IZE	-i dagokiona	237	238
	ezaugarria	IZE	IZE	-i dagokiona	20	20

21. taula. *Ezaugarri* erlazioa markatzen duten erlatoreak.

IV.4.5. Jasale

Honelakoetan pentsatuko dugu erlazonatu gisa markatutako hitzak *jasango* duela sarrera-hitzak adierazten duen *egoera*. Beste modu batean, sarrera-hitzak *egoera* bat adierazten du, eta, *egoera* hori izango da erlazonatutako hitzak *izango* duena edo *jasango* duena.

(30) Izenaren_definizioa = izena-*genitiboa* + ‘egoera’.

(30) itsasgora A0 iz. *Itsasoaren egoera*, batez besteko maila baino gorago dagoenean.

etiketa	erlazioa	sar- kat	erlz- kat	def-eskema	erlt- kop	erlz- kop
EGOERA	jasale	IZE	IZE	ize.-en egoera	192	89
	jasale	IZE	ADJ	adj.-en egoera		

22. taula. *Jasale* erlazioa markatzen duten erlatoreak.

IV.4.6. Berezko ekintza

Definizioan erlazonatutako hitz gisa markatuko dugun aditza da sarrera-hitzaren *berezko ekintza*.

(31) Izenaren_definizioa = aditza-*t(z)en* + ‘duen pertsona’.

(31) albaitari A0 iz. Abereen eta, oro har, animalien osasuna zaintzen duen pertsona.

23. taulan erlatoreen kopurua ematerakoan (*erlt-kop*), kasu batzuetan bi zenbaki paratu ditugu. Halakoek bi erlatore marka dituzte. Adibidez, *-ko erabiliren* kasuan,

⁵³ *nolakotasuna* eta *-z betea* erlatoreak bata bestearen alderantzizko erlazioak dira. *nolakotasuna* erlatorearen kasuan, *ezaugarria* sarrera-hitza da, eta *ezaugarri* horren edukitzailea erlazonatutako hitza; *-z betea* erlatorearen kasuan, berriz, *ezaugarria* eta erlazonatutako hitza da, eta sarrera-hitza *ezaugarri* hori duena.

erlatore gisa markatzen ditugu *erabili* eta *aditz laguntzailea*. Biak markatzea erabaki dugu, ondoren *elipsiaren* erregelak marka ez ditzaten. Hemendik aurrera erlatoreen kopurua adierazten duen zutabeen bi zenbaki horiek agertzen direnean bi erlatore markatzen ditugula adieraziko du: lehenengo zenbakiak adierazten du zenbat erlatore dauden, eta parentesi artean dagoen zenbakiak zenbat marka jarri ditugun. *–ko erabili* kasuan, 179 erlatore daude, eta 258 marka jarri ditugu.

etiketa	erlazioa	sar-kat	erlz-kat	def-eskema	erlt-kop	erlz-kop
PERTSONA	berezko ekintza	IZE	ADI	-tzen duen pertsona	374	278
GAUZA	berezko ekintza	IZE	ADI	-tzen d(u)en gauza	124	53
ZERA	berezko ekintza	IZE	ADI	-tzen d(u)en zera	114	69
TRESNA	berezko ekintza-helburu	IZE	ADI	-tzeko tresna	217	119
	ezaugarria	IZE	IZE	ize + tresna		5
-KO ERABILI	berezko ekintza-helburu	IZE	ADI	-tzeko erabiltzen d(u)ena	179 (258)	179
ERABILI	erabilera	IZE	IZE/ADJ	ize/adj gisa erabiltzen dena	116 (232)	30
JOERA	berezko ekintza-helburu	ADJ	ADI	-tzeko joera duena	88	88
	berezko ekintza-helburu	IZE	ADI	-tzeko joera duena	6	6

23. taula. *Berezko ekintza* erlazioa markatzen duten erlatoreak.

IV.4.7. Meronimia

Meronimia hitzaren azpian, *parte-oso*a erlazioa dago. Honen alderantzizkoa *holonimia* da (Miller, 1998a):

The part-whole relation between nouns is generally considered to be a semantic relation, called *meronymy* (from the Greek *meros* 'part'; Cruse 1986), comparable to synonymy, antonymy, and hyponymy. This relation also has an inverse: if S_m is a meronym of S_h , then S_h is said to be a *holonym* of S_m . (37. or.)

Normalean erabiltzen den testa hau da: *–en parte*a da. Beraz S_m S_h ren *parte*a dela esatea onargarria baldin bada, S_m S_h ren *meronimo*a dela esango dugu; eta, bestalde, onargarria baldin bada esatea S_h S_m *duela parte* gisa, orduan esango dugu S_h S_m ren *holonimo*a dela.

Kontuan izan behar dugu erlazio *meronimikoa* transitiboa dela, hau da, erlazioa hereda daitekeela. Adibidez, moko eta hego, txoriaren *meronimo*ak baldin baditugu, eta txoriaren *hiponimo*a txantxangorri baldin bada, orduan pentsatu ahalko dugu txantxangorriaren *meronimo* direla moko eta hego.

Batzuetan herentzia kontu hau gertatzen ez denean, definizioan zehatz-mehatz topatuko dugu azalduta. Adibide gisa *orgaren* definizioa hartu dugu.

orga A1 iz. Bi edo lau *gurpileko* **ibilgailua**, zamak eramateko erabiltzen dena eta zaldiek, idiek... higiarazten dutena.

Definizioan ikusten da *orgaren* *hiperonimoa* *ibilgailu dela eta meronimo gisa gurpil duela*.⁵⁴

Hona orain *leraren* definizioa

lera H1A0 iz. *Gurpilik gabeko* **orga**, bi albo oholen bidez irristatzen dena, elur gainean ibiltzeko, edo belarra-eta garraiatzeko toki malkartsuetan erabiltzen dena.

Definiziotik aterako dugu *leraren* *hiperonimoa* *orga dela*, baina hark duen *ezaugarri* bat ez duenez —gurpilak izatea, alegia—, esplizitu agertzen da hiztegian. Halakoak ondo ezagutuz gero, benetan gertatzen ez diren erlazioak saihesteko balioko digute.

Meronomiaz ari garela, argitu behar dugu *meronomia* mota bakarra ez dela *parte-oso meronomia*, baizik eta badirela gehiago (Miller, 1998a):

Such observations raise questions about how many different IS-A-PART-OF relations there are. Winston, Chaffin, and Hermann (1987) differentiate six types of meronyms: component-object (*branch/tree*), member-collection (*tree/forest*), portion-mass (*slice/cake*), stuff-object (*aluminum/airplane*), feature-activity (*paying/shopping*), and place-area (*Princeton/New Jersey*). Chaffin, Hermann, and Winston (1988) add a seventh: phase-process (*adolescence/growing up*). But Iris, Litowitz, and Evens (1988) distinguish only four: functional part (*wheel/bicycle*), segment (*slice/loaf*), member (*sheep/flock*), and subset (*meat/food*). Meronymy is obviously a complex set of semantic relations, and there is considerable disagreement about how to distinguish among them. (39. or.)

Beraz, ikusten dugunez, aski zaila dirudi ados jartzea zenbat *meronimo* klase ditugun zehazteko. *WordNet* datu-basean erabaki dute hiru *meronomia* mota besterik ez dituztela kodetuko: *partea* (aulki / hanka), *kidea* (ardi / artalde), *zerezkoa* (harresi / harri).

Guk, hasiera batean, denak multzo berean sartu ditugu.

⁵⁴ Adibide honekin *meronomia* erlazioaren transitibotasuna erakutsi nahi izan dugu, nahiz eta gure azterketaren puntu honetan, *ibilgailua* eta *gurpil meronimoak* direnik ezagutzeko gai ez garen.

(32) Izenaren_definizioa = izenlaguna-en + ‘multzoa’.

(32) hezurdura A0 iz. Gizon edo abere baten **hezurren** multzoa, gorputzean ezarriak dauden eran.

(33) Izenaren_definizioa = izenlaguna-en / -ko + ‘zattia’.

Aurretik *lekuzko genitiboa* baldin badu, hau izango da *holonimoa*, baldin eta honen aurretik *edutezko genitiborik* agertzen ez bada.

(33) mezerdi A0 iz. Sagara, **meza^{ren}** erdiko **zattia**.

etiketa	erlazioa	sar-kat	erlz-kat	def-eskema	erlt-kop	erlz-kop
MULTZOA	meronimia	IZE	IZE	-en multzoa	446	403
TALDEA	meronimia	IZE	IZE	-en taldea	58	26
ATALA	holonimia	IZE	IZE	-en atala	15	17
ZATIA	holonimia	IZE	IZE	-en zattia	243	178
KIDEA	holonimia	IZE	IZE	-ko kidea	55	40

24. taula. *Meronimia* / *holonimia* erlazioa markatzen duten erlatoreak.

IV.4.8. Berezko objektua

Berezko objektu erabiliko dugu zerbaiti berez dagokion *objektuari* erreferentzia egiteko. Zenbaitetan *helburua* ere izan daiteke. Esaten dugunean ogigileak ogiak egiten dituela, pentsatuko dugu ogigilearen *berezko objektua* ogia dela, eta, era berean, bere *helburua* ere izango da. Honekin oso lotuta agertzen da hitz anitzeko batek adieraziko duen *berezko ekintza* erlazioa, izan ere ogigile horren *berezko ekintza* ogia egin dugu.

(34) Adjektiboaren_definizioa = izena-absolutibo mugatua + ‘egiten d(it)uena’.

(34) bihurri H1A1N2 izond. Haurrez eta hitz eginez, **gaiztakeriak** edo **bihurrikeriak** **egiten dituen**a, esanekoa ez dena.

Aurretik esan dugun bezala, erlatoreen kopurua ematerakoan bi zenbaki agertzen direnean, erlatoretzat bi hitz markatzen ditugula adierazten du: lehenengo zenbakiak zenbat erlatore dauden adierazten du, eta parentesi artean dagoen zenbakiak zenbat marka jarri ditugun.

etiketa	erlazioa	sar-kat	erlz-kat	def-eskema	erlt-kop	erlz-kop
EGIN	berezko objektu	IZE	IZE	ize + egiten d(it)uena	151 (302)	129
	berezko objektu	ADJ	IZE	ize + egiten d(it)uena	91 (182)	80
EMAN	berezko objektu	IZE	IZE	ize + ematen d(it)uena	70 (140)	45
	berezko objektu	ADJ	IZE	ize + ematen d(it)uena	15 (30)	15
EGITE	berezko objektu	IZE	IZE	ize. + egitea	205	22
	berezko ekintza (HAUL)	IZE	ADI	adb. + egitea		9
	berezko egile	IZE	IZE	-en + egitea		3
	ezaugarria	IZE	ADJ	adj-i dagokion egitea		22
	ezaugarria	IZE	IZE	ize-i dagokion egitea		6
	hiperonimia	IZE	IZE	ize edo egitea		15

25. taula. Berezko objektu erlazioa markatzen duten erlatoak.

IV.4.9. Gabezia, eza

Sarrera-hitzak, erlazonatutako hitzak –definizioan ageri dena, alegia– adierazten duenaren *gabezia* edo *falta* ageri du.

(35) Izenaren_definizioa= izena-rik + ‘eza’

(35) askatasun A1 iz. *Eragozpenik eza*, traba gabezia.

askatasunen definizioa eragozpenik eza, traba gabezia baldin bada, askatasuna duenak bi *ezaugarri* hauek ez ditu izango: eragozpena eta traba.

etiketa	erlazioa	sar-kat	erlz-kat	def-eskema	erlt-kop	erlz-kop
GABEZIA	gabezia	IZE	IZE	ize + gabezia	26	26
	gabezia	ADJ	ADJ	ize + eza	65	65

26. taula. Gabezia adierazten duen erlatoak.

IV.4.10. Betekizun / betegai

Aditzen artean bete erlatoa agertzen da, eta aurretik zer duen arabera zehaztuko dugu erlazioa: *betegaia*, *betekizuna*, edo *modua*.

(36) Aditzaren_definizioa=izena-z + ‘bete’.

(36) belartu A0 ad. *Belarrez bete*.

belarturen definizioan, belar da *betegaia*.

etiketa	erlazioa	sar-kat	erlz-kat	def-eskema	erlt-kop	erlz-kop
BETE	betegai	ADI	IZE	-z bete	71	62
	modua	ADI	ADB	adb + bete		9
	betekizun	ADI	IZE	ize (abs) + bete		13

27. taula. Betekizun / betegai erlazioak markatzen dituen erlatoak.

IV.4.11. Aditzen genus orokorrak

Aditzen artean dauden genus usuenen azterketa egiterakoan hauek agertu zaizkigu (ikus IV.3.3. atala): egin, bihurtu, behartu, eman, hartu, izan, jarri, galdu, kendu, joan, eraman, gertatu, bete.

Hauek guztiak nahiko esanahi ahuleko aditzak dira, eta, askotan, beste izen edo adjektibo batekin batera hitz anitzeko aditz bat osatzen dute. Gu *hitz anitzeko sinonimoak*, eta *hitz anitzeko genusak* bereizten saiatu gara. Erlatorea, partizipio ahul hori da; eta, hitz anitzeko terminoaren lehenengo osagaia, erlazonatutako hitza.

Aditz ahul gisa tratatu ditugu: egin, bihurtu, behartu, eman, hartu, eman edo hartu, izan, jarri, galdu, kendu, galdu edo kendu, joan, eramane, joan edo eramane.

Hitz anitzeko sinonimoen adibideak:

(37) Aditzaren_definizioa = [,] ize/adj/adb [def-hasi]⁵⁵ + aditz ahula.

(37) abaildu A1 ad. Guztiz nekatu, *leher egin*.

(37) ilaundu A1 ad. *Ilaun bihurtu*, erraustu; suntsitu.

Hitz anitzeko hiperonimoen adibideak:

(38) Aditzaren_definizioa = modifikatzailea + ize/adj/adb + aditz ahula + [, (...)].

(38) azaldu A1 ad. Zerbaiti buruz *hitz egin*, hura ulertarazteko edo ezagutarazteko.

(38) labakitu A0 ad. Lur eremu bat *labaki bihurtu*.

Bi erlatore dituztenen adibideak:

ihartu A0 ad. *Hezetasuna kendu edo galdu*, (...)

jarauntsi H1A0 ad. Oinordetu, *oinordekotzan hartu edo eman*.

iharturen kasuan erlatoreak *kendu* eta *galdu* dira. Erlatoreen kontaketa egiteko garaian, *kendu* erlatorea eta *galdu* erlatorea agertzen direla kontatu behar dugu, eta taulan *galdu/kendu* etiketaren barruan ez da erlatore-kopuru zehatza ematen, baizik eta aurrekoekin nahasten da. Hau da, *galdu/kendu* errenkan, marra bat agertzen da erdian eta adierazten du hiztegian *galdu* erlatorea 110 aldiz agertzen dela, eta *kendu* erlatorea 93 aldiz. Erlazonatutako hitza, aldiz, bakarra agertzen denez, erraza da kontatzen; kasu honetan 26 erlazonatutako hitz daude.⁵⁶

⁵⁵ Mota honetako sinonimo bat izan dezagun, erlazio marka daraman hitzak definizioa —edo deskribapeneko— lehenengo elementua izan behar du, edo puntuazio-marka baten ondoren azaldu.

⁵⁶ Halako bi erlatore *da / du* formak dituzten aditzekin agertu ohi dira.

etiketa	erlazioa	sar- kat	erlz- kat	def-eskema	erlt- kop	erlz- kop
EGIN	sin-haul	ADI	ADI	ize/adj/adb + egin	471	300
	genus-haul	ADI	ADI	mod. ⁵⁷ + ize/adj/adb + egin		103
BIHURTU	sin-haul	ADI	ADI	ize/adj + bihurtu	468	306
	genus-haul	ADI	ADI	mod. + ize/adj + bihurtu		86
BEHARTU	sin-haul-helburu	ADI	ADI	-tzera + behartu	426	417
	genus-haul-helburu	ADI	ADI	mod. + -tzera + behartu		9
EMAN	sin-haul	ADI	ADI	ize/adb + eman	296	132
	sin-haul-helburu	ADI	ADI	-tzera + eman		5
	genus-haul	ADI	ADI	mod. + ize/adb + eman		50
	genus-haul-helburu	ADI	ADI	-tzera + eman		4
EMAN / HARTU	sih-haul	ADI	ADI	ize/adb + hartu edo eman	269	40
HARTU	sin-haul	ADI	ADI	ize/adb + hartu		84
	genus-haul	ADI	ADI	mod. + ize/adb + hartu		27
IZAN	sin-haul	ADI	ADI	ize/adj/adb + izan	195	95
	genus-haul	ADI	ADI	mod. + ize/adj/adb + izan		61
JARRI	sin-haul	ADI	ADI	ize/adj/adb + jarri	161	87
GALDU	sin-haul	ADI	ADI	ize/adj/adb + galdu	110	40
GALDU / KENDU	sin-haul	ADI	ADI	ize/adj/adb + galdu edo kendu		26
KENDU	sin-haul	ADI	ADI	ize/adj/adb + kendu	93	22
JOAN	sin-haul	ADI	ADI	-ra joan	86	21
JOAN / ERAMAN	sin-haul	ADI	ADI	-ra joan edo eraman		16
ERAMAN	sin-haul	ADI	ADI	-ra eraman	75	16

28. taula. Hitz anitzeko sinonimo / hitz anitzeko hiperonimo erlazioa markatzen duten erlatoreak.

IV.4.12. Elipsia

Halakoak maiz agertzen dira definizioetan eta multzo handia osatzen dutenez, denak batera ematea erabaki dugu, nahiz eta sortzen diren erlazioak hainbat izan.

Genusaren posizioa betetzen duen elementuak du *elipsia*.

(39) Adjektiboaren_definizioa = aditza-*t(z)en* + 'duena'.

(39) gaiztatzaile A0 izond. *Gaiztatzen duena*.

(39) geldigarri A0 izond. *Geldiarazten duena*.

⁵⁷ Genus direnak aurretik *modifikatzaile* (*mod.*) bat —edo *differentia*— behar dute.

etiketa	erlazioa	sar-kat	erlz-kat	def-eskema	erlt-kop	erlz-kop
ELIPSIA	berezko ekintza	ADJ	ADI	-t(z)en d(u)ena	760	883
	ezaugarria	ADJ	ADJ	adj. d(u)ena	906	169
	ezaugarria	ADJ	IZE	ize. d(u)ena		560
	berezko ekintza	IZE	ADI	-t(z)en d(u)ena	1154	1213
	ezaugarria	IZE	ADJ	adj. d(u)ena	876	116
	ezaugarria	IZE	IZE	ize. d(u)ena		354
	antonimoa	ADJ	ADJ	adj. ez dena	88	18

29. taula. *Elipsidun* erlatoreak.

IV.4.13. Erlatoreak adierazten duena

Halakoetan erlazioa bera zein den erlatoreak adierazten duela pentsatuko dugu.

Erlazioak normalean adjektiboan eta izenen artekoak dira.

etiketa	erlazioa	sar-kat	erlz-kat	def-eskema	erlt-kop	erlz-kop
IZE-KOA	erlatorearen arabera	ADJ	IZE	-en ize+koa	47	47
ADLG	erlatorearen arabera	ADJ	IZE	ize (dek.) + partizipioa	19	19

30. taula. Erlatoreak adierazten duen erlazioa markatzen duten erlatoreak.

Halako erlazioak bereziak direnez, hemen bertan azalpenak emango ditugu:

(40) Adjektiboaren_definizioa = izena-genitiboa + izena-koa.

(40) absolutista A0 izond. *Absolutismoaren aldekoa*.

(40) antisemita A0 izond. *Juduen kontrakoa*.

absolutista adjektiboa absolutismo izenarekin dago erlazionatuta aldeko erlazioaren bitartez, eta antisemita judurekin, kontrako erlazioaren bitartez.

Baina, zenbaitetan, izen horrek duen *ezaugarri*ren bati egingo dio erreferentzia.

(40) beltz A1 izond. Batere kolorerik edo argirik ez duena, *ikatzaren kolorekoa*.

Azkenean, pentsatuko dugu beltza eta ikatza, kolorearen bitartez daudela erlazionatuta.

Erlatorearen aurretik markarik gabeko izena dugunean, beti esango dugu sarrera-adjektiboak izango duela erlazionatuta dagoen izenaren *ezaugarri* bat, eta *ezaugarri* hori zein motakoa den erlatoreak adieraziko du.

(41) Adjektiboaren_definizioa = izena-markarik gabea + izena-koa.

(41) gaztainakara A0 izond. *Gaztaina kolorekoa*.

(41) kurbatu A0 izond. *Kakotua, kurba formakoa*.

gaztainakara denak gaztainak duen *ezaugarri* bat du, kolorea, alegia; eta kurbatu denak kurbak duen *ezaugarri* bat du, forma, alegia.

(42) Adjektiboaren_definizioa = izena-postposizio atzizkia + partizipio-mugatua.

(42) bihurri H1A1 izond. Gaiztotasun**era makurtua**.

(42) hieroglifiko A0 izond. Hieroglifo**z osatua**; iluna.

bihurri **eta** gaiztotasun erlazionatuta daude makurtu erlazioaren bitartez; eta hieroglifiko adjektiboa eta hieroglifo izena osatu erlazioaren bitartez.

IV.5. Izen eta adjektiboak

III.8. atalean ikusi dugunez, izen eta adjektibo kategoria duten hitzak 1.020 dira. Zenbaitetan zaila da hitz bati kategoria bakarra esleitzea, gerta baitaiteke izen eta adjektibo gisa erabiltzea inongo esanahi diferentziarik gabe. Halakoek EHn bi kategoria jasotzen dituzte, eta gure helburua izango da izenentzat eta adjektiboentzat egindako erregelen bidez zehaztea zein kategoriari dagokion definizio-patroia duen.

Erlazioak, beraz, aurretik azaldutako berak dira, eta hemen paratuko dugu erlazio bakoitza zenbaitetan agertzen den multzo honetan, eta zein izan den definizioaren azterketaren arabera esleitu diogun kategoria.

Etiketa	kategoria	kop
SIN-IZE	izena	92
SIN-ADJ	adjektiboa	157
SIN-IZE-ADJ	ize / adj	193
ANT	adjektiboa	6
GENUS	izena	212

31. taula. *Sinonimoen, genusen eta antonimoen* informazioa.

Datuok ikusita, 304 aldiz erabiltzen da izenen moldea, 163 aldiz adjektiboen moldea, eta 193 aldiz zehaztu ezina den moldea, bai izenentzat eta bai adjektiboentzat egokia delako.

Erlatoreak modu berean aztertu ditugu.

Etiketa	kategoria	erltkop	erlzkop	Etiketa	kategoria	erltkop	erlzkop
PERTSONA	izena	23	13	-I EGOKI	adjektiboa	112	114
KIDE	izena	15	13	IZE-KOA	adjektiboa	38	41
ZERA	izena	8	7	ADJ-KOA	adjektiboa	14	14
IZEN	izena	7	4	JOERA	adjektiboa	9	3
BAT / BAKOITZA	izena	7	6	ADLG	adjektiboa	2	2
ERABILI	izena	7	1	IA-SIN	adjektiboa	1	1
-I EMAN IZENA	izena	3	3	-Z BETEA	adjektiboa	1	1
GABEZIA	izena	2	2	ELIPSIA	ize / adj	449	464
GAUZA	izena	1	1	EGIN	ize / adj	54	27
-KO ERABILI	izena	2	1	EMAN	ize / adj	4	2

32. taula. Erlatoreen informazioa.

Erlatore berezien artean, 75 aldiz erabiltzen dira izenen moldeak, 177 aldiz adjektiboenak, eta 507 aldiz ezingo dugu zehaztu zer-nolako kategoriari dagokion definizioa duen, erabilitako moldeak bi kategorietan erabiltzen direlako.

Adibidez greziar sarrerak bi definizio ditu eta bakoitza kategoria bati dagokio:

```
greziar      A0      izlag.#iz. Greziakoa, Greziari dagokiona;
                                   Greziako biztanlea.
```

“;” aurretik dugun definizioa, adjektibo bati dagokiona da: greziar-grezia erlazioa dugu; ondoren dagoena, izen baten definizioa da, eta greziarraren genusa biztanle da.

Ondoko adibidean erlatorea elipsia da, eta definizio-patroi hau izenetan eta adjektiboetan erabiltzen da.

```
haizatzaile A0i      izond.#iz. Haizatzen duena.
```

haizatzaile eta haizaturen arteko erlazioa dugu.

IV.6. Azpisarrerak

III.8. atalean ikusi dugunez, azpisarrerak 2.576 dira EHn. Multzo honetan edozein kategoriatako hitzak topatu ditugu. IV.5. atalean egin dugun bezala, hemen ere ezagunak zaizkigun kategorietako erregelak zenbatetan aplikatzen diren aztertu dugu; hartara, kategoria bakoitzeko zenbat definizio ditugun jakingo dugu, eta *azpisarrera* horren kategoria zein den ezagutuko dugu, dagoeneko III. kapituluan aipatu dugun bezala, askotan kategoria gaizki heredatzen baitu.

Erlatorea	kategoria	kop.
SIN-IZE	izena	235
SIN-ADJ	adjektiboa	222
SIN-IZE-ADJ	ize / adj	29
SIN-ADI	aditza	186
GENUS-IZE	izena	475
GENUS-ADI	aditza	205
ANT-ADJ	adjektiboa	3

33. taula. Azpisarreraren sinonimoen eta *genusen* informazioa.

Izenen moldeak erabiltzen dituzten erlatoreak 710 dira, adjektiboenak dituztenak 225 eta aditzenak 391. Izen edo adjektibo diren erabaki ezin direnak 29 dira.

Erlatoreak modu berean aztertu ditugu.

Erlatorea	kategoria	erlt kop	erlz kop
ERABILI	izena	19	3
MOTA	izena	16	15
-I EMAN IZENA	izena	8	8
IZEN	izena	8	6
ZATI	izena	7	5
PERTSONA	izena	6	2
BAT / BAKOITZA	izena	5	5
TRESNA	izena	4	0
GABEZIA	izena	3	3
EGOERA	izena	3	1
ZERA	izena	2	2
EZA	izena	2	2
NOLAKOTASUN	izena	2	1
GAUZA	izena	1	1
MULTZO	izena	1	1
TALDE	izena	1	1
ERA	izena	1	1
KIDE	izena	1	0
KIDEKO	izena	1	0
MODUKO	izena	1	0
-KO ERABILI	izena	0	9
ELIPSIA	ize / adj	128	123
EGIN	ize / adj	8	7
EMAN	ize / adj	7	7

Erlatorea	kategoria	erlt kop	erlz kop
IZE-KOA	adjektiboa	12	12
ADJ-KOA	adjektiboa	9	7
IA-SIN	adjektiboa	5	5
-I EGOKI	adjektiboa	5	5
ADLG	adjektiboa	2	1
JOERA	adjektiboa	0	1
EGIN-SIN-HAUL	aditza	134	111
EGIN-GENUS-HAUL	aditza		12
IZAN-SIN-HAUL	aditza	90	48
IZAN-GENUS-HAUL	aditza		13
HARTU-SIN-HAUL	aditza	32	21
HARTU-GENUS-HAUL	aditza		1
EMAN-SIN-HAUL	aditza	27	17
EMAN-SIN-HELBURU-HAUL	aditza		2
EMAN-GENUS-HAUL	aditza		3
JOAN-SIN-HAUL	aditza	7	1
JARRI-SIN-HAUL	aditza	5	2
GERTATU-SIN-HAUL	aditza	5	0
GALDU-SIN-HAUL	aditza	3	2
BEHARTU-GENUS-HAUL	aditza	2	2
BIHURTU-SIN-HAUL	aditza	1	1
KENDU-SIN-HAUL	aditza	1	0

34. taula. Erlatoreen informazioa.

Erlatore bereziak dituztenen artean, izenen moldeak dituztenak 92 dira, adjektiboenak 33, izen edo adjektibo diren erabaki ezin direnak 143, eta aditzen moldeak dituztenak 307.

AFARI-MERIENDA RA0 iz. Ilunabarrean egiten den **otordua**.

afari-merienda azpisarrerak izen baten definizioa du, eta *genusa* otordu da.

ALIMALEKO RA0 izlag. Izugarri **handia**.

alimalekok adjektibo baten definizio-patroia du, eta *ia-sinonimoa* handi da.

BIDEAK EGIN RA0 iz. Asko **ibili**.

bideak egin azpisarrerak aditzaren definizio-patroia du eta *hiperonimoa* ibili da.

Nabaria denez, askotan aditzen moldeak erabiltzen dira, eta horietako askok kategoria gisa *izena* dute —bideak egin kasuan bezala—, izen baten azpisarrerak direlako. Beraz, oso garrantzitsua da kategoria beste bat dela zehaztea.

Hauezaz gain, badira inongo markarik ez duten hainbat definizio, baina horietako gehienak adverbioen definizioak dira, eta horiek ez ditugu lan honetan

aztertu. Horien guztien zenbatekoak emaitzak aztertzerakoan ikusiko ditugu (ikus IV.7.5. atala).

IV.7. Metodologia eta ebaluazioa

Lortutako emaitzak aztertzen hasi aurretik labur-labur erabilitako metodologia azalduko dugu.

Lehenik, hiztegitik kategoria bakoitzeko 200 definizio hartu ditugu zoriz, eta hortik saiatu gara erabiltzen diren eskemak eta definizio-patroiak erauzten. Hauek topatu ondoren, lortutako deskribapena erregela bihurtu dugu *Murritzapen Gramatika* erabiliz. Erregela hauen doitasuna neurtzeko, aipatutako laginean aplikatu ditugu eta fintzen joan gara. Nahikoa finduta daudenean, hiztegi osoari aplikatu dizkiogu. Ondoren zoriz hartutako beste 100 definizioko lagina hartu eta egindako erregelen zuzentasuna neurtu dugu, eta egon daitezkeen patroik eta eskema berriak identifikatzen saiatu gara. Aurreko pausoa bezala, erregela berriak egin ondoren, eta laginean aplikatu ondoren, hiztegi osoari aplikatu dizkiogu. Azken urrats bezala, markatu gabe gelditu diren 50 definizio hartu ditugu eta zergatiak analizatu. Hortik erregela berriak atera dira eta horiek berriro ere hiztegi osoari aplikatu dizkiogu. Modu horretan sinonimoen eta genusen eskema orokorrak lortu ditugu. Erlature berezien eskemak, berriz, ez ziren hainbeste ateratzen, eta III.8. atalean azalduko *token*-segida usuenen azterketa baliatu dugu horiek harrapatzeko garaian.

Metodologia aplikatu ondoren, zein emaitzak lortzen ditugun ikustearren, landu dugun multzo bakoitzeko 100 sarrerako laginak hartu ditugu.

Multzo bakoitzeko laginetan markatu ditugu zeintzuk diren ezagutu beharko genituzkeen sinonimoak, genusak, erlatureak eta erlazionatutako hitzak. Hauek *helburu* zutabea jarri ditugu eta zenbat diren zehaztu dugu. Erregelak aplikatu ondoren, kontatuko ditugu ea hauetako zenbat markatu ditugun ondo —*ondo* zutabea—, zenbat markatu ditugun gaizki —*gaizki* zutabea—, denera zenbat elementu markatu ditugun —*markak* zutabea— eta zenbat elementu utzi ditugun markatu gabe —*galdua* zutabea—. Kontu hauek erabilia, bi ehuneko atera ditugu: *estaldura*, hau da, helburu diren horietatik guztietatik zenbat markatzen ditugun ondo; eta *errorea*, hots, markatzen ditugun guztietatik zenbat markatzen ditugun

gaizki. *denera* errenkan aurreko guztien baturak ditugu, eta, ehunekoen kasuan, batez besteko orokorrak.

Datu hauek multzoka eman ditugu, eta dagokien atalean azaldu.

IV.7.1. Izenak

Hona izenen ehun sarrerako lagina aztertu ondoren ditugun emaitzak.

	Helburu	Ondo	Gaizki	Markak	Galduta	Estaldura	Errorea
Sin	61	59	0	59	2	%96,72	%0,00
Genus	49	48	3	51	1	%97,96	%6,12
Erlt	37	35	0	35	2	%94,59	%0,00
Erlz	32	21	1	22	11	%65,63	%3,13
Denera	179	163	4	167	16	%91,06	%2,40
Definizioak	108						

35. taula. Izenen laginetik ateratako emaitzak.

Estaldurak kontuan izaten baditugu, batez bestekoa %91,06 dugu, eta errore-tasa %2,40 bakarrik. Denera 167 marka ditugu definizioetan. Baina dagoeneko txostenean agertu denez, erlatoreak eta erlazonatutako hitzak elkarrekin agertu ohi dira erlazio bakarra adierazteko. Beraz, jakin ahal izateko definizio bakoitzeko zenbat erlazio topatuko ditugun, erlatoreak bakarrik kontatuko ditugu.⁵⁸ Hala, denera 145 erlazio ditugu.

108 definizio horietatik —taulan *guztira* zutabea— 100 dira adierak eta 8 “;” baten ondoren sortutako definizioak; 107k dute nolabaiteko informazioa —taulan *inform. bai* zutabea—, hau da, bakarra geratu zaigu inongo informaziorik gabe —taulan *def. ona, inform. ez* zutabea—. Lagin honetan ez zaigu definizio txarrik —errataren bat duena-edo— agertu. Halakorik agertzen denean, *def. txar.* zutabea kontatuko dugu.

	Guztira	Inform. bai	Def. txar.	Def. ona, inform. ez
Definizioak	108	107	0	1
Adierak	100	99	0	1

36. taula. Izenen lagineko definizioei buruzko datuak.

Erlazioak 145 direnez, eta informazioa duten adierak 99, informazioa duten adierek duten batez besteko erlazio-kopurua 1,46 da.

⁵⁸ Kontatzeko modu hau kategoria guztietan erabiliko dugu eta baita datu orokorrak emateko garaian ere.

Informaziorik gabe geratu zaigun definizioa pixka bat berezia dela esan daiteke, adberbio izan daitekeen hitz batena baita.

gatazka A4 adlag.#iz. Elkarren aurkako interesen,
sentimenen... aurrez aurrekoa.

Definizio honetan pentsatu ahal izango dugu aurrez aurrekoa horretan *elipsia* dagoela, eta, beraz, erlatore bat dugula. Honek ihes egiten digu.

Atalez atal, lagin honetako definizioen artean 61 sinonimo daude, horietako 59 ondo ezagutu ditugu, eta 2 dira ezagutu gabe gelditu zaizkigunak. Genusei dagokienez, berriz, 49 daude eta 48 ezagutzen ditugu ondo, eta 1 galdu dugu. Hauezaz gain, badira beste 3 elementu genus gisa markatu ditugunak eta benetan ez direnak.

Erlatoreei erreparatzen badiegu, 37 daude laginean; horietako 35 ezagutzen ditugu ondo, eta 2k egin digute ihes. Erlazionatutako hitzak, berriz, 32 daude eta 21 markatzen ditugu ondo. Hauezaz gain, bada kasu bat erlazionatutako hitz gisa hartzen duguna eta benetan ez dena.

Lagina analizatu ondoren hiztegiko izen guztien definizioak etiketatu ditugu sinonimoen, genusen, eta erlatore berezien gramatikarekin (ikus I.1. eranskina), eta hona lortutako emaitzak.

	Sin	Genus	Erlz	Erlt
Denera	11.926	11.650	4.768	6.768
Definizioko	1,27	1,17	1,11	1,24

37. taula. Izenen artean ditugun erlazioak.

Definizioko errenkan marka bera duten adiera guztiak kontuan hartuta ateratako ehunekoa paratu dugu. Ehuneko hau, oro har, laginean ateratzen dena baino baxuagoa da —izenen laginean erlazioaren bat duten adieretan dauden batez besteko erlazio-kopurua 1,46 dela esan dugu—, eta honek badu arrazoirik. Laginean ateratako emaitza eratorpenaren erregelak aplikatu aurrekoa da, eta ez da eratorpen-markarik agertzen. Hau, berriz, marka guztiak jarri ondorengoa da, eta eratorpen-erlazio bat baldin badago, ez du erlatorearen erlazioa markatuko, eta, beraz, ez da emaitza hauetan agertuko. Hala, erlatore gutxiago agertzen dira.

IV.7.2. Aditzak

Hona aditzen ehun sarrerako lagina aztertu ondoren ditugun emaitzak.

	Helburu	Ondo	Gaizki	Markak	Galduta	Estaldura	Errorea
Sin	60	58	0	58	2	%96,67	%0,00
Genus	33	33	0	33	0	%100,00	%0,00
Erlt	51	51	0	51	0	%100,00	%0,00
Erlz	51	38	1	39	13	%74,51	%2,56
Denera	195	180	1	181	15	%92,31	%0,55
Definizioak	100						

38. taula. Aditzen laginetik ateratako emaitzak.

Estaldurak kontuan izaten baditugu, batez bestekoa %92,31 dugu, eta errore-tasa %0,55. Denera 142 erlazio markatu ditugu.

100 definizio horietatik 97 dira adierak eta 3 “;” baten ondoren sortutako definizioak; 96k dute nolabaiteko informazioa, hau da, bakarra geratu zaigu inongo informaziorik gabe.

	Guztira	Inform. bai	Def. txar.	Def. ona, inform. ez
Definizioak	100	99	1	0
Adierak	97	96	1	0

39. taula. Aditzen lagineko definizioei buruzko datuak.

142 erlazio markatu ditugunez, eta informaziodun adierak 96 direnez, informazioa duen definizio bakoitzean dauden erlazioen batez bestekoa 1,48 da.

Markatu gabe geratu zaigun definizioa txartzat hartuko dugu,⁵⁹ nahiz eta *aditz* kategoriakoa dela agertu, *adjektibo* bat da, eta gainera beste bi *adjektiboren* bitartez definitzen da.

hertsi A2 ad. Meharra, estua.

Atalez atal, lagin honetako definizioen artean 60 sinonimo daude. Horietako 58 ondo ezagutzen ditugu, eta 2 dira ezagutu gabe gelditu zaizkigunak. Genusei dagokienez, berriz, 33 daude, eta 33ak ezagutzen ditugu ondo. Erlatoreei erreparatzen badiegu, 51 daude laginean eta horietako 51 ezagutzen ditugu ondo. Erlazionatutako hitzak, berriz, 51 daude eta 38 markatzen ditugu ondo. Hauezaz gain bada kasu bat erlazionatutako hitz gisa hartzen duguna eta benetan ez dena.

Lagina analizatu ondoren hiztegiko aditz guztien definizioak etiketatu ditugu eta hona agertzen den erlazio-kopurua:

⁵⁹ Errata itxura du.

	Sin	Genus	Erlz	Erlt
Denera	3.828	2.428	2.044	2.729
Definizioiko	1,35	1,20	1,09	1,15

40. taula. Aditzen artean ditugun erlazioak.

IV.7.3. Adjektiboak

Hona adjektiboen ehun sarrerako lagina aztertu ondoren ditugun emaitzak.

	Helburu	Ondo	Gaizki	Markak	Galduta	Estaldura	Errorea
Sin	51	50	0	50	1	%98,04	%0,00
Genus	1	0	0	0	1	%0,00	%0,00
Erlt	88	78	0	78	10	%88,64	%0,00
Erlz	84	63	2	65	21	%75,00	%3,08
Denera	224	191	2	193	33	%85,27	%1,04
Definizioak	108						

41. taula. Adjektiboen laginetik ateratako emaitzak

Estaldurak kontuan izaten baditugu, batez bestekoa %85,27 dugu, eta errore-tasa %1,04. Denera 128 erlazio markatu ditugu.

108 definizio horietatik 99 dira adierak eta 9 “;” baten ondoren sortutako definizioak; 108 definizio horietatik 99k dute nolabaiteko informazioa, hau da, 9 geratu zaizkigu inongo informaziorik gabe.

	Guztira	Inform. bai	Def. txar.	Def. ona, inform. ez
Definizioak	108	99	3	6
Adierak	99	90	3	6

42. taula. Adjektiboen lagineko definizioei buruzko datuak

Erlazioak 128 direnez, eta erlazioen bat duten adierak 90, informazioa duen adiera bakoitzeko 1,42 erlazio ditugu.

Esan bezala, 9 definizio geratu zaizkigu informaziorik gabe. Batzuk agertzen diren definizio-patroiak ez direlako oso maiz agertzen hiztegian eta landu gabe gelditu direlako.

klasiko	A1N1	izond.	Hizkuntzez mintzatuz, ereduzkotzat hartzen den epekoa.
ertain	A2	izond.	Batez bestekoa.
silababakar	A0	izond.	Silaba bakar batek osatua.

Aurrekoetan dauden erlatore motak ez daude landuta. Hurrengoak aldiz, landuta daude, baina tartean beste elementuak agertzen direnez, ez ditugu harrapatzen. Halakoak gainditzeko azaleko sintaxia markatzen duten tresnak erabiltzea lagungarri

litzateke, baina sarrera nagusian (ikus I. kapitulua) aipatu dugun bezala, oraingoz ez ditugu aplikatu, eta etorkizuneko lan gisa utzi dugu.

porotsu	A0	izond.	Zulo txiki asko dituen.
sendo	A1	izond.	Lodia edo gorputz handikoa delako, indartsua edo irmoa (pertsonez eta mintzatuz, orobat osasuntsua, sasoikoa) dena.
zimurtsu	A0	izond.	Zimur asko duena.

Erlatore hauek landuta daude: dituen, dena, duena; baina kontua da erlazonatutako hitzaren —zulo, indartsu / irmo, zimur— eta erlatorearen artean beste hitz mordoa daudela, eta horregatik ez ditugu ezagutzen.

Eta geratzen zaizkigun hiruak esan dugu definizio txarrak direla...

erromar	H2A0	izond.	Europako sartaldean XI-XII. mendeetan garatu zen arte estiloa, ⁶⁰ arkitekturakoa bereziki.
kezkati	A0N1	izond.	Kezkaz beterik.
trufakeria	A0	izond. ⁶¹	Trufa.

erromar hitzaren definizioa *izen* batena da: *genusa estilo da*,⁶² kezkatiaren definizioa ez dirudi oso ondo osatuta dagoenik; eta trufakeria definitzeko sinonimo bat erabiltzen da, baina gure datu-basean trufa *izen* gisa agertzen da, eta *Euskal Hiztegian* bertan ere *izen* bezala agertzen da.

Atalez atal, lagin honetako definizioen artean 51 sinonimo daude, horietako 50 ondo ezagutzen ditugu, eta bakarra ez dugu ezagutzen.

Genusei dagokienez, kontuan izan behar dugu, definizioz adjektiboen artean ez lukeela izen-genusik agertu behar, baina, kategoria-mugan dauden hitzek zenbaitetan izenei dagozkien definizio motak azaltzen dituztenez, genusak harrapatzeko erregela orokorra erabiltzen dugu. Kasu honetan genusa den elementu bat agertu zaigu eta ez dugu ezagutu.

Erlatoreei erreparatzen badiegu, 88 daude laginean eta horietako 78 ezagutzen ditugu ondo, eta 10ek egin digute ihes. Erlazonatutako hitzak, berriz, 84 daude eta 63 markatzen ditugu ondo. Hauek gain badira erlatoreen artean 2 kasu erlaretzat

⁶⁰ Badirudi hemen *estilokoa* beharko lukeela. Beraz, errata dugu.

⁶¹ Errata da.

⁶² Nahiz eta errata izan, *estilo* izena denez, eta gune sintaktikoaren posizioa betetzen duenez, erregelek genustzat hartzen dute.

hartzen ditugunak eta benetan ez direnak, eta erlazonatutako hitzen artean ere markatu behar ez genituzkeen bi elementu markatzen ditugu.

Lagina analizatu ondoren hiztegiko adjektibo guztien definizioak etiketatu ditugu, eta hona lortu dugun erlazio-kopurua:

	Sin	Erlz	Erlt
Denera	2.553	2.477	2.878
Definizioko	1,33	1,16	1,28

43. taula. Adjektiboen artean ditugun erlazioak.

Izenen laginean dugun estaldura %91,06 da, aditzen laginean %92, 31, eta adjektiboen laginean %85,27. Beraz, adjektiboen artean dugun estaldura beste kategorietan duguna baino baxuagoa da. Honen arrazoia adjektiboen definizioetan erlatore berezien erabilera beste bietan baino handiagoa delako da. Eta, erlatore hauek askotan eratorpen atzizkiekin lotuta daudela ikusi ahal izan dugu. Honek motibatu du, nolabait, V. kapituluaren ikusiko dugun eratorpenaren azterketa, nahiz eta kategoria guztiei aplikatu.

IV.7.4. Izenak eta adjektiboak

Dagoeneko aipatu dugun bezala, talde batean bildu ditugu hiztegiaren bi kategoriak dituzten hitzak. Bertatik ateratako 100 sarrerako lagina aztertu dugu.

	Helburu	Ondo	Gaizki	Markak	Galduta	Estaldura	Errorea
Sin	53	45	0	45	8	%84,91	%0,00
Genus	12	12	2	14	0	%100,00	%14,29
Erlt	88	81	0	81	7	%92,05	%0,00
Erlz	83	70	5	75	13	%84,34	%6,67
Denera	236	208	7	215	28	%88,14	%3,26
Definizioak	124						

44. taula. Izen eta adjektiboen laginetik ateratako emaitzak.

Estaldurak kontuan izaten baditugu, batez bestekoa %88,14 dugu, eta errore-tasa %3,26. Denera 140 erlazio markatu ditugu.

124 definizio horietatik 99 dira adierak eta 25 “;” baten ondoren sortutako definizioak; 99 sarrera horietatik 97k dute nolabaiteko informazioa, hau da, 2 geratu zaizkigu inolako informaziorik gabe.

	Guztira	Inform. bai	Def. txar.	Dudazkoa def.
Definizioak	124	122	1	1
Adierak	99	97	1	1

45. taula. Izen eta adjektiboen lagineko definizioei buruzko datuak.

Erlazioak 140 direnez, eta erlazioen bat duten adierak 97, informazioa duen definizio bakoitzeko 1,44 erlazio ditugu.

Ikusten dugunez, 2 definizio ditugu informaziorik ez dutenen artean, eta esango dugu bi horiek txarrak direla. Lehenengoa garbi da txarra dela, eta, bestean, baztergarria izan daitekeen patroia bat erabiltzen da (*-z esaten da*).

besoetako	A0	izond.#iz.	Norbait bere aitabitxi edo amabitxiari buruz.
txikigarri	A0N1	izond.#iz.	Hitz baten esanahia arintzen duen atzizkiaz, edo horrelako atzizki batez eratutako hitzaz esaten da. ⁶³

Kontuan izan behar dugu hauek definizio mota bereziak erabiltzen dituztela; orain arte ikusi dugun bezala, izenek beren definizio motak erabiltzen dituzte eta adjektiboek berenak. Hemen biak batera agertzen dira eta bi kategoria horietan aplikatutako erregelak aplikatu behar dira. Erregela horiek, zenbaitetan, adieraziko digute zein kategoriatako moldea erabili den definizioan. Gehienetan erabiltzen diren definizio-patroiak bi kategorietan agertzen direnak dira, eta ezin izango dugu zehaztu batekoak edo bestekoak diren —73 definizio—. Izenen moldeak 23 definiziotan erabiltzen dira, eta adjektiboenak direnak 26 definiziotan.

informaziodun definizioak	izenak	adjektiboak	biak
122	23	26	73

46. taula. Izen eta adjektibo diren sarreren definizio-patroien arabeko kategoria.

Moldeak bi kategorietara egokitzeak hainbat errore ekarriko dizkigu, eta hori oso nabaria da erlazioen eta genusen kasuan. Erlazioen kasuan, zenbaitetan kontuan hartu ez ditugun molde konplexuak erabiltzen dira:

azkargarri	A0	izond.#iz.	Azkar tzen , sendot tzen , indart tzen duena.
behartsu	A1	iz.#izond.	Pertsonez mintzatuz, gauzarik beharrekoen ak ere ez dituen a .

azkargarriren kasuan, indartu harrapatzen dugu, baina azkartu eta sendotu ez. behartsuren kasuan, dagoeneko aipatu dugun sintaxiaren arazoa sortu zaigu berriro.

Genusei dagokienez, errorea ohikoa baino handiagoa da (%14,29), baina adjektiboetan aipatu dugun bezala, gutxi dira. 12 daude ondo markatuta, eta ez diren 2 markatzen ditugu.

⁶³ *-z esaten da* eskema oso gutxitan agertzen da hiztegian. Honen ordeko dugu *-z mintzatuz* eskema, eta hau baino askoz egokiagoa, gainera.

Beste erlazioak direla-eta, definizioen artean 53 sinonimo daude, horietako 45 ondo ezagutzen ditugu, eta 8 dira ezagutu gabe gelditu zaizkigunak.

Erlatoreei erreparatzen badiegu, 88 daude laginean eta horietako 81 ezagutzen ditugu ondo, eta 7k egin digute ihes. Erlazionatutako hitzak, berriz, 83 daude eta 70 markatzen ditugu ondo. Erlatoreetan ez dugu errorerik eta erlazionatutako hitzetan 5 markatu ditugu gaizki.

Lagina analizatu ondoren hiztegiko izen eta adjektibo kategoria duten definizioak etiketatu ditugu, eta hona markatu dugun erlazio-kopurua:

	Sin	Genus	Erlz	Erlt
Denera	442	212	720	765
Definizioko	1,14	1,09	1,22	1,23

47. taula. Izen eta adjektibo diren sarrera-hitzetan dauden erlazioak.

IV.7.5. Azpisarrerak

Azpisarreretan agertuko zaizkigu kategoria guztietako elementuak, baita aztertu ez ditugun kategorietakoak ere. Aurreko atalean egin dugun bezala, guk aztertutako hiru kategoriatako erregelak aplikatuko ditugu benetan zein kategoriatako hitzak diren jakin ahal izateko.

Hartutako laginean 92 definizio —89 adiera— ditugu. Hauetatik, 40 dira aztertu ez ditugun kategorietakoak eta horiek ez dute erregelengandik markarik jasotzen, gaizki markatzen duen kasu batean izan ezik. Horiezaz gain, badira definizio txarra duten beste 2 definizio.

Izenen definizio motak dituzten sarrerak 23 dira, adjektiboenak 4, aditzenak 16, eta izen eta adjektibo izan daitezkeenak 4. Ezagutu gabe geratu dira 2 definizio, bat aditz batena eta bestea adjektibo batena.

informaziodun definizioak	izenak	adjektiboak	aditzak	izen / adjektibo	Ezagutu gabe
49	23	4	16	4	2

48. taula. Azpisarreraren definizio-motak.

Harrapatzen ditugun elementuen datuak honokoak dira:

	Helburu	Ondo	Gaizki	Markak	Galduta	Estaldura	Errorea
Sin	21	19	1	20	2	%90,48	%5,00
Genus	29	24	1	25	5	%82,76	%4,00
Erlt	23	23	1	24	0	%100,00	%4,17
Erlz	18	15	1	16	3	%83,33	%6,25
Denera	91	81	4	85	10	%89,01	%4,71
Definizioak	92						

49. taula. Azpisarreraren laginetik ateratako emaitzak.

Emaitzak ez dira aurrekoetan ditugunak bezain onak, baina kontuan izan behar da *azpisarrerak* modu berezian definitzen direla; izatez *sarrera berezitzat* har ditzakegu. Kontuan izanda beste kategorietan egindako lana erabili dugula definizio hauetan dauden erlazioak erauzteko eta azpisarreraren kategoria ezagutu edo konfirmatzeko, emaitza onak dira.

Estalduak kontuan izaten baditugu, batez bestekoa %89,01 dugu, eta errore-tasa %4,71. Denera 69 erlazio markatu ditugu.

92 definizio horietatik 89 dira adierak eta 3 “;” baten ondoren sortutako definizioak. Horietatik 41 dira guk aztertzen ez ditugun kategorietako hitzak, eta 2 definizio txar daude. Beraz, 49 dira azterkizun izango ditugun definizioak, eta horietatik 2 gelditu dira inongo markarik gabe.

	Guztira	Kategoria ez landuak	Def. txar.	Inform. bai	Def. ona, inform. ez
Definizioak	92	41	2	47	2
Adierak	89	39	2	46	2

50. taula. Azpisarreraren lagineko definizioei buruzko datuak.

Erlazioak 69 direnez, eta erlazioren bat duten adierak 46, informazioa duen adiera bakoitzeko 1,5 erlazio ditugu.

Lagina analizatu ondoren hiztegiko azpisarrera guztiak etiketatu ditugu, eta hona ezagutu ditugun erlazioak:

	Sin	Genus	Erlz	Erlt
Denera	673	695	485	589
Definizioko	1,15	1,10	1,14	1,14

51. taula. Azpisarreretan markatu ditugun erlazioak.

IV.7.6. Laburpena

Prozesu hauek guztiak aplikatu ondoren 52. taulan jarri ditugu lortu ditugun emaitzak. Taula honetan kategoria —edo azterkizun izan den multzo— bakoitzarekin lortu ditugun estaldura (*est*), eta errore-tasa (*err.*) daude.

	izenak		adjektiboak		izenak-adjektiboak		aditzak		azpisarrerak	
	est.(%)	err.(%)	est.(%)	err.(%)	est.(%)	err.(%)	est.(%)	err.(%)	est.(%)	err.(%)
sin	96,72	0	98,04	0	84,91	0	96,67	0	90,48	5,00
genus	97,96	6,12	0	0	100	14,29	100	0	82,76	4,00
erlt	94,59	0	88,64	0	92,05	0	100	0	100	4,17
erlz	65,63	3,13	75,00	3,08	84,34	6,67	74,51	2,56	83,33	6,25

52. taula. Kategoria guztietako emaitzen batez-bestekoak.

Aurreko taulako emaitza guztiak kontuan hartuz gero, lortzen dugun batez-besteko estaldura %88,97 da, eta errore-tasa %2,14 (ikus 53. taulako *denera1* errenka). Azpisarrerak pixka bat bereziak direnez, horiek kontuan ez baditugu hartzen, estaldura bera da, %88,97; eta errore-tasa gutxitu egiten da: %1,85 (ikus 53. taulako *denera2* errenka).

	Helburu	Ondo	Gaizki	Markak	Galduta	Estaldura	Errorea
ize	179	163	4	167	16	%91,06	%2,40
adj	224	191	2	193	33	%85,27	%1,04
ize-adj	236	208	7	215	28	%88,14	%3,26
adi	195	180	1	181	15	%92,31	%0,55
azpisar	91	81	4	85	10	%89,01	%4,71
denera1	925	823	18	841	102	%88,97	%2,14
denera2	834	742	14	756	92	%88,97	%1,85

53. taula. Batez besteko orokorrak.

IV.8. Ondorioak eta etorkizuneko lanak

Esan bezala, kapitulu honen helburua definizioko guneak ezagutzea da. Definizioaren gune den posizio horretan, sinonimoa, genusa edo erlatore berezia topatu dugu, agertu ditugun definizio-patroien arabera.

Definizioan hitz bakarra agertzen bada, sarrera-hitzaren sinonimoa dela esan dugu. Definizioan deskribapena dugunean bi aukera daude: gune sintaktikoa eta semantikoa bera izatea edo ezberdina izatea. Bera denean, hitz hori sarrera-hitzaren genusa da, eta bi hitz horien artean dagoen erlazioa *hiperonimikoa* da. Gune sintaktikoa adiera ahuleko hitza baldin bada, berriz, hitz hori erlatore berezia da, eta sarrera-hitzarekin definizioko beste hitz bat jarriko du harremanean. Definizioan

dagoen hitz hori *erlazionatutako hitza* deitu dugu. Sarrera-hitza eta erlazionatutako hitz honen artean hainbat erlazio sortzen dira: *ia-sinonimia*, *antonimia*, *hiperonimia*, *ezaugarria*, *edukia*, *jasale*, *berezko ekintza*, *meronimia*, *berezko objektu*, *berezko egile*, eta *erlatoreek adierazten dituztenak*.

Erlatore berezi zenbaitek zehazten dituzten erlazioak ez dira beti bi hitzen artekoak, eta horien errepresentazioa asko zailtzen bada ere, guk, lehen urrats honetan, ez dugu konplikatu nahi izan eta modu errazean eman dugu erlazioa. Hala, bi izenen artean *berezko objektu* erlazioa dagoela esan dugu, nahiz eta jakin *berezko objektu* erlazio hori gauzatu ahal izateko tartean *aditz* bat egon beharko duela, erlazio hori aditzen eta izenen arteko erlazioa delako. Adibidez, *eskulangileren berezko objektua* *eskulan* dela esan dugu, nahiz eta *berezko objektu* erlazioa definizioan dugun erlatore berezi den *egin* aditz ahularen eta *eskulanen* artekoa den. Halakoak eratorpena aztertzeke garaian berriro ere agertuko zaizkigu (ikus V. kapitulua).

Definizioko gune izan eta esanahiz ahul diren hitzak detektatzeko, gune-posizioan agertzen diren elementu usuenen azterketa egin dugu. Hitz bat definitzeko garaian esanahi orokorragoa duen beste hitza erabili ohi da. Zenbaitetan hitz horren esanahia oso orokorra da, eta guk horiei erlatore berezi deitu diegu. Beraz, gune-posizioan agertzen diren hitz usuenak izango direnez esanahi ahulekoak, posizio horretan gehien agertzen direnak erlatore berezitzat hartu ditugu.

Erlazio konplexuen beste adibidea dira *hitz anitzeko sinonimoak* eta *hitz anitzeko genusak*. Halakoak aditzetan agertzen dira, eta gune posizioan dagoen aditza oso esanahi ahulekoa denean gertatzen da. *mintzaturen* definizioa *hitz* egin baldin bada, gune posizioan *egin* agertuko da, eta *aditz-partizipioa* denez, *genus* itxura du; baina, esan dugun bezala, adiera ahula duten horiei, erlatore deituko diegu, eta aurreko izenarekin batera sarrera-hitza *sinonimo* direla esango dugu. Halakoetan erlatore marka izango du *aditz-partizipioak* eta erlazio marka izenak, baina biak batera erauzi beharko dira erlazio egokia izan dezagun.

Azterketa oso hau izen, adjektibo, eta aditz kategoria duten hiztegiko hitzekin egin dugu. Hala, kategoria bakoitzari dagozkion definizio-eskemak zeintzuk diren definitu ditugu. Kategoria bakoitzak bere-bereak diren eskemak ditu. Hau egiteak, hiztegiko balizko erratak harrapatzeke balio du. Adjektibo baten definizio-eremuan

Beste arazo bat da bi erlature berezi elkarren segidan agertzea. Alegia, zatietako bakoitza segida baldin badugu, bakoitza erlatureak adieraziko digu sarrera hitzaren *hiperonimoa* izango dela bere aurretik duen izena; kasu honetan zati da izen hori, baina hau, era berean, *meronimia* markatzen duen erlaturea dugu. Oraingoz lehen erlaturea besterik ez dugu markatuko halakoak ez direlako asko hiztegian barrena, baina hauek detektatzea eta lantzea etorkizuneko lana izango da.

liburuki. A0 iz. Idazlan hedatuak banatzen diren **zati****etako bakoitza**, bereiz josten eta argitaratzen direnak.

Oso zaila da halakorik erregeletan zehaztea, eta, berez, agertuko zaigu liburukiren *hiperonimoa* zati dela.

Ondorengo kapitulura pasa baino lehen komenigarria iruditu zaigu lortutako emaitzak azpimarratzea. Sinonimio, genus eta erlature bereziei esker 48.156 erlazio markatu ditugu.

sin	19.442
genus	14.985
erlz	10.494
erlt	13.729

54. taula. Erlazio-kopurua.

Hiztegian dauden definizioak kontuan hartuta, hona erlazio bakoitzarekin markatzen ditugunak.

erlazioa	markadun definizioak	guztiarikoko ehunekoa
sin	19.442	42,48
genus	14.985	35,98
erlz	10.494	26,19
erlt	13.729	31,58

55. taula. Markadun definizio-kopurua.

Ondorio hauek dira, beraz, EHtik informazioa erauzteko garaian lortu ditugun lehenak. Ondorengo atalean, sarrera-hitz diren elementuei beste modu batean erreparatuko diegu. Hemendik aurrera aztertuko ditugun hitzak eratorriak izango dira.

Amaia, haurrari hau erakusten saiatuko gara, ezta?

Muxuk.

irri	A0	iz. Barrea , bereziki isekazkoa.
irribarre	A0	iz. Barre xumea, begitartearen keinu sotil batez adierazten dena.
barre	A0	iz. Poz bizia eta bat-batekoa adierazten duten aurpegiko giharren mugitze bereziak, gehienetan hasperen zaratatsuekin batera egiten direnak.

Irria, lanik eztena
ondorio ezin errana...
Oin aberasten ukailea
biluzi gabe emailea...
Zoin lastergi den iragana,
betikotz daukalarik emana
Pobreenari aberatsenak
ezin kenduzko ondasuna

Akituaren eztigarri
etsituaren altxagarri
doluminetan goxagarri
nahigabeetan sendagarri
Irria ez da erosten
ez eta ere ebasten
Emitzean balioa hartzen
ez baitu diruak kundatzen

Irria, ezperantza xori
gure anaiantzat mezulari.
Gau egun beltzetan kantari
goroko bizi berriari,
Ezpain bihotzetan irria
kanta mutil libertatia!
zabal munduari egia!
nausituko da justizia

I-rrri-aaaa, i-rrri-aaaa, i-rrri-aaa, ii-rrri-aaaa.

(P. Ospital, P. Carrere)

V. Eratorpena

V.1. Sarrera

Aurreko atalean ikusi berri dugu hiztegi elebakar batean agertzen den erlazio nagusia: sarrera-hitza eta definizio hitz baten artekoa, alegia. IV.7.3. atalean ikusi dugunez, ordea, adjektiboen artean beste kategorietan dugun informazioa baino gutxiago dugu. Hau normala da genus informaziorik ez dagoelako adjektiboen artean, eta informazio gehiago erlatoeren bitartez lortzen delako. Baina, era berean, konturatu gara erlatoe berezi jakin batzuk, askotan eratorpen-atzizki jakin batekin daudela lotuta, eta eratorpenak euskarazko hitzen sorkuntzan indar handia duenez, aztertzea erabaki dugu.

Atal honetan bilatuko dugun erlazioa ez da sarrera-hitza eta definizioan agertzen den hitz baten artekoa, baizik eta sarrera hitzak berak bere barruan duen beste hitz batekin duen erlazioa. EHN agertzen diren sarrera-hitza eratorriak aztertu nahi ditugu.

Eratorpenaren barruan, hitz eratorriaren eta oinarriaren artean zenbait erlazio lexiko-semantiko berezi sortzen dira. Horiek eratorriaren esanahiaren arabera izaten dira, eta, eratorpen-atzizki bera duten hitzek esanahi diferenteak izan ditzaketenez, oinarriarekiko erlazioak ere desberdinak dira. Beraz, atzizki batekin sortutako eratorriek adiera bat baino gehiago baldin badute, eratorriaren eta oinarriaren arteko erlazioak ere bat baino gehiago izango dira, eta beraz, erlazioa ambigua izango da. Gure helburua erlazio hori zein den jakitea da, eta, horretarako,

hiztegian agertzen den definizioa baliatu dugu. Azken ideia hau azpimarratu nahi dugu tesi-lan honen aportazio nagusietako bat delako. Anbiguotasun semantikoa automatikoki ebaztea oso zaila da, eta guk, kasu honetan, definizioei esker ebatzi ahalko dugu. Eratorriaren esanahiaren arabera, definizioetan halako eskema jakinak azalduko direla pentsatu dugu, eta eskema horiek detektatu ondoren, erregelak egin eta sarrera-hitza den eratorriaren eta bere barruan duen oinarriaren artean zein erlazio dagoen zehaztu dugu.

Aurreprozesuaren kapituluan azaldu dugu hitz eratorriak zeintzuk diren nola jakiten dugun, eta landu beharreko atzizki lexikalak zeintzuk diren nola erabaki dugun.

Landu beharreko atzizki lexikalak zeintzuk diren erabaki ondoren, horiei buruz literaturan zer esan izan den aztertuko dugu (Azkarate, 1990b; Azkue, 1923-1925; Lafitte, 1944; Villasante, 1974; Euskaltzaindia, 1985; 1987a; 1987b; 1991; 1992; 1993). Atzizki bakoitzarentzat hainbat interpretazio agertu zaizkigu, eta hori izango da atal honetan azaltzen saiatuko garena. Atzizki bakoitza analizatu dugu, eta, nahiz eta irakurgarritasuna erraztearren hona informazio osoa ez dugun ekarri, B. eranskinean zehaztu dugu atzizki bakoitzari dagokion informazio osoa. B. eranskinean atzizki bakoitzarekin sortutako hitzak zenbat diren paratu dugu, hauen oinarria zein kategoriatakoa den, atzizki bakoitzak izan ditzakeen interpretazioak, oinarrien ezaugarriak zeintzuk diren, eta abar.

Eratortariaren adiera zein den definizio-eskemak adieraziko digu. Horretarako, definizio-eremuan deskribapen bat izatea garrantzitsua da, baina, zenbaitetan —dagoeneko IV. kapituluan ikusi dugun bezala— hitz bat definitzeko garaian sinonimo bat erabiltzen da. Halakoetan, sinonimo horren definizioa topatzen saiatu gara, bertatik nolabaiteko informazioa lortu ahal dugun ikusteko, eta, lortuz gero, hasierako sarrera-hitziari informazioa pasatzeko. Halako metodoari *sinonimiaren hedadura* deitu diogu, eta V.9.3. atalean zehatzago azaldu dugu.

Beraz, eratorri horien adiera zein den zehaztu nahi dugu; eta, hitzen adierei buruz ari garenez, pentsatu dugu eratorriaren eta oinarriaren arteko lotura

*EuroWordNet*⁶⁴ txerta daitekeela, eta hala *EuskalWordNet* aberastu. Lotzeko garaian jakin beharko dugu *EuroWordNet*eko zein *synseti* lotu beharko ditugun oinarria eta eratorria, eta horretarako, hiztegi elebidunak, eta lortutako itzulpenek *EuroWordNet*en bertan duten adiera-zenbakia erabiliko dugu. Pentsatuko dugu euskarazko hitzen eta beste hizkuntzetako hitzen adieren artean kontzeptu bera baldin badago, kontzeptu hori dela eratorpen-prozesu horretan lortzen dena, eta hala lortuko dugu oinarriak adierazten duen kontzeptua eta eratorriak adierazten duen kontzeptua lotzea, eta lotura hori, beste hizkuntzetarako ere baliagarria izatea.

Adierak desanbiguatzeko erabiltzen den metodo automatikoa da tesi-lan honen beste aportazioetako bat.

Hemendik aurrera ikusiko dugun kapituluaren hezurdura honokoa da: V.2. atalean morfologiaren oinarrizko zenbait ezaugarri emango ditugu; V.3. atalean eratorpen-atzizkiei buruz eta hauek sailkatzeko moduei buruz hitz egingo dugu; V.4. atalean eratorpenak hiztegian duen isla aztertuko dugu; V.5. atalean hiztegian agertzen diren eratorpen-atzizkiak eta hauek dituzten adierak aztertuko ditugu; V.6. atalean definizioak erabiliz adiera-desanbiguazioa nola egiten dugun erakutsiko dugu; V.7. atalean emaitza orokorrak paratuko ditugu; V.8. atalean lortutako emaitzak *EuroWordNet*en txertatzeko egin behar den lana analizatuko dugu; V.9. atalean etorkizuneko lanak zehaztuko ditugu; eta V.10. atalean ondorioen eta etorkizuneko lanen bilduma jarri dugu.

V.2. Morfologia

Eratorpena aztertzea erabaki genuen berez erlazio lexiko-semantiko bereziak daudelako oinarriaren eta hitz eratorriaren artean. Morfologian bi zati bereizten dira: hitz-eraketa eta flexioa. Bi zati hauen artean badago elkargune bat, dudazko kasuak, alegia.

Hitz-eraketaren barruan eratorpena eta hitz-elkarketa bereizten dira. Eratorpenean, parte bat beregaina da eta bestea ez; hitz-elkarketan, berriz, bi elementu beregain lotzen dira.

⁶⁴ *EuroWordNet*, aurrekarietan aipatu dugun bezala, hainbat hizkuntza batzen dituen ezagutza-base lexikala da. Sinonimo multzo bat kontzeptu baten isla da, eta sinonimo multzo horiei *synset* deitu ohi zaie.

Eratorpenaren eta flexioaren artean bada nahasketarik, baina eratorpena maila lexikoan gertatzen da, eta flexioa, aldiz, maila sintaktikoan eta perpaus mailan agertzen da.

Eratorpena	
maila lexikoa	zuritasun
	zuriketa

56. taula. Eratorpena.

Flexioa	
maila sintaktikoa	bertsoa
perpaus maila	datorrela

57. taula. Flexioa.

Hitz eratorria perpaus batean erabili ahal izateko, flexiora jo beharko dugu eta dagokion forma eman perpaus zuzena lortzeko. Beraz, morfologiaren barruan bada hierarkia bat:

hitz-eraketa → flexioa

Lehenik eratorpena aplikatuko da, eta ondoren flexioa.

V.2.1. Eratorpenaren eta flexioaren arteko ezberdintasunak

Eratorpenean atzizkiak askotan aldatuko du oinarriaren kategoria:

zuri (adj) → zuritasun (ize)

Kategoriaz gain, tasun semantikoak ere alda daitezke:

bertso (ize) → bertsolari (ize)

- bizidun + bizidun

+ konkretu + gizaki

Flexioak, aldiz, ez du kategoriarik aldatzen.

Eratorpen-atzizkiek esanahia aldatzen dute, eta semantikoki badu eraginik; badute eduki semantikoa. Horrezaz gain, klase irekia osatzen dute: atzizki berriak sor daitezke edo batzuk baztertu; adibidez, *-gailu* atzizkia ez zen mendebaldean erabiltzen, ekialdeko atzizkia baitzen.⁶⁵ Baina orain —eta euskara batuan garaile atera da— asko erabiltzen da mendebaldean.⁶⁶

⁶⁵ *Orotariko Euskal Hiztegia* (OEH) (Mitxelena eta Sarasola, 1987-) jartzen duenez, atzizki hau baztaneran, lapurteran, baxe-nafarreran, zubereran eta erronkarieran erabiltzen zen.

⁶⁶ Baina gaur egun beste adiera batekin erabiltzen da.

Gauza bertsua gertatzen da *-go* atzizkiarekin; gainera, atzizkia ekialdekoa da izatez, baina orain mendebaldean erabiltzen hasi da barra-barra.⁶⁷ Mendebaldean *-tza* atzizkia erabiltzen zen.

Flexio-elementuek, aldiz, klase itxia osatzen dute, eta klase hauen ezaugarria da elementu berriak ezin sortu ahal izatea.

Emankortasun maila dela eta, flexio-elementuak erabat erregularrak dira, emankortasun osoa dute. Eratorpen-elementuak, aldiz, ez dira erabat erregularrak. Eratorpen-atzizki bakoitzak oinarri berezia behar du. Batzuek besteek baino lege zurrunagoak dituzte, baina erregelak beti egongo dira.

Izen sintagma → izen multzoa + artikulua	Izena → aditz-oina + keta	Izena → izena + tza
mahaia	azterketa	jendetza
jendea	eginketa	aterkitza
erregela bat ematen denean, horrek ez du salbuespenik eta, daudenean, oso zehatzak dira	ez dute emankortasun osorik	

58. taula. Flexio- eta eratorpen-elementuen erregularutasun maila.

Aurreko taulan agertzen den *eginketa* ezingo da sortu; eta *aterkitzak* ere —lehenengo kolpean behintzat— aski arraroa dirudi.

Flexio-elementuak behin bakarrik aplika daitezke, ez dira errekursiboak —genitiboak izan ezik—; eratorpena, aldiz, errekursiboa da —oro har hitz-elkarketa errekursiboa da—.

Perpaua osatzeko, flexioa behar-beharrezkoa da, eta, eratorpena, aldiz, ez. Izan ere, dagoeneko aipatu dugun moduan, eratorpena maila lexikoan topatuko dugu, eta flexioa maila sintaktikoan eta perpau mailan.

Esanahi-kontuak aipatzen hasita, bai eratorpen-atzizkiek eta bai flexio-atzizkiek interpretazio bat baino gehiago izan ditzakete. Atal honetan aztertuko ditugu eratorpen-atzizkiek izan ditzaketen interpretazioak, eta VI. kapituluaren —definizio osoa aztertzen dugunean— flexio-atzizkien interpretazioak aztertuko ditugu.

⁶⁷ Erabilera ezezagun hau dela-eta, atzizki hau gaizki erabiltzen da mendebaldean.

V.3. Atzizki lexikalak

Elementu bat atzizki lexikala den edo ez erabakitzeke garaian kontuan izan diren kriterioak honakoak dira: maiztasuna, duten autonomia, esanahi-aldaketa, eta kategoria-aldaketa (Pérez Gaztelu *et al.*, 2004).

Hitz batek gero eta autonomia gutxiago baldin badu, eta maiztasun nahiko handiz beste hitz batzuekin batera —elkarrekin— agertzen bada, atzizki lexikalaren itxura duela pentsatuko dugu.

Horiez gain, kategoria-aldaketak ere garrantzia du. Oinarria kategoria batekoa izan, eta, delako atzizki horrekin juntatuz gero, kategoria aldatzen bada, atzizki lexikalaren zantzua izango du:

lodiera → dimentsioa adierazten duen izena

etorrera → ekintza adierazten duen izena

Guilbert-ek (1975) dioenez, atzizkia kategoria bat baino gehiagori eransten bazaio, honek esan nahi du atzizki hori oso indartsu eta aberatsa dela. Baina kategoria ezberdinetako oinarriak hartze honek ez du atzizki ezberdina denik adierazten. Hau da, oinarria kategoria batekoa izan edo bestekoa, interpretazio bera izango du, baina guk oinarriaren kategoriaren arabera bereizketa egin beharko dugu, eratorria eta oinarriaren arteko erlazioa ezberdinak direlako kasu batean eta bestean. Oinarriaren kategoriaren arabera, erlazio ezberdinak izango ditugu.

Demagun salerostegi eta behitegi hitz eratorriak ditugula. Bi kasuek *leku* bati egiten diote erreferentzia, baina eratorriaren eta oinarriaren arteko erlazioa ezberdina dute. Salerostegik salerosten den *leku* bati egingo dio erreferentzia, beraz erlazioa izena eta aditzaren artekoa da: *salerostegian salerosi egiten da*; eta behitegin, aldiz, oinarria behi izena da, eta erlazioa bi izenen artekoa da: *behitegian behiak gordetzen dira*.

Bauer-ek (1983) dioenez, berriz, atzizki bat baino gehiago dugun jakin ahal izateko esanahiari begiratu behar zaio. Atzizki lexikal bat zenbait oinarri lotu ondoren, hainbat esanahi baldin baditu, eratorpen-atzizkiaren interpretazio ezberdinak direla pentsatuko dugu.

eratorria	definizioa	interpretazioa
aitzurlaritza	aitzurlariaren lanbidea	lanbidea
sasitza	sasia dagoen lekua	lekua
eskaintza	eskaintzea	ekintza
mirabetza	mirabearen egoera	egoera
adiskidetza	adiskidearen nolakotasuna	nolakotasuna

59. taula. *-tzarekin* sortutako hitz eratorrien interpretazioak.

59. taulan, esanahiari begiratzen badiogu, badirudi bost multzo ditugula.

Hala ere, eta nahiz eta kriterio horiek izan, badira eratorpenaren eta konposizioaren arteko mugan diren elementuak. Guk, batik bat, eratorriaren eta oinarriaren arteko erlazioak lortu nahi ditugu, eta horretarako oso maiz agertzen diren konposizio-elementuek laguntzen bagaituzte, erabili egingo ditugu, hitz eratorria edo konposatua osatzen duten eztabaidan sakondu gabe. Muga-mugan dagoen atzizki horietako bat *-dun* da, baina *-dun* duten eratorrien eta hauen oinarrien arteko erlazioak erauzteko gai garenez, atzizki lexikal gisa landu dugu.

Behin elementu bat atzizki lexikala izan edo ez erabaki ondoren, arazoa izan ohi da atzizki horren lema zein den finkatzea.

-ari / *-lari* / *-tari* / *-kari*
-asun / *-tasun* / *-arzun* / *-tarzun*

Honen aurrean hainbat aukera ditugu: formarik zaharrena hartzea lematzat, edo gehien erabili den formari lehentasuna ematea, euskalki mailako kriterioa, alegia.

Lema zein den ezarri ondoren, aldaerak azaldu behar dira. Beste hizkuntzetan aldaerak azaltzeko baldintza fonetikoak aipatu izan dituzte, baina baita baldintza lexikoak ere, baldintza gramatikalak, eta abar.

Eratorpen-atzizkien kasuan baldintza fonetikoak hartu beharko ditugu kontuan. Testuinguru fonikoa begiratu beharko dugu, baina, hala ere, zaila izango zaigu asko zehaztea. Batez ere joerak finkatu beharko ditugu. Guk hemen azalduko dugun lanean ez ditugu aldaketa fonetikoak aztertuko,⁶⁸ eta eratorpen-atzizki batek hiztegian izan ditzakeen aldaera guztiak bilatzen saiatuko gara.

Ataltxo honekin amaitzeko, atzizkiak sailkatzeko erabili izan diren moduak ikusiko ditugu: Villasante (1974), Azkue (1923-1925), Lafitte (1944), eta Ithurri (1979), hain zuzen ere.

⁶⁸ Atzizki lexikalak aurreko hitzarekin lotzerakoan jasaten dituzten aldaketa fonetiko-fonologikoak aztertu ditu Izagirre-k (2004).

V.3.1. Alfabeto-hurrenkera (Villasante, 1974)

Alfabeto-hurrenkera proposatzen du Villasantek irizpide gisa. Hurrenkera honek baditu alde onak eta txarrak: onak, atzizkiak aurkitzea oso erraza dela; txarrak, zenbait kasutan atzizkia bera finkatzeko arazoak izango ditugula:

-(t)asun, -(t)ar, -(t)egi, -(k)izun, -(k)oi

Honelako kasuetan nora jo beharko dugu?

Gainera, balio bertsuko atzizkiak erabat sakabanaturik agertuko lirateke: *-le*, *-tzaile*. Askoz egokiagoa litzateke biak elkarren ondoan agertzea, beraien artean dauden berdintasunez nahiz desberdintasunez errazago jabetzeko.

V.3.2. Hitz berrien kategoria eta atzizkien esanahia (Azkue 1923-1925; Lafitte, 1944)

Beste batzuek bi irizpide hauen arabera sailkatu izan dituzte atzizkiak:

- Hitz berrien kategoria: izen-sortzaileak, adberbio-sortzaileak, aditz-sortzaileak, adjektibo-sortzaileak.
- Atzizkiaren esanahia. Izen-sortzaileen artean: *izen abstraktuen* sortzaileak, *agente* adierakoak, *leku* adierakoak, *denbora* adierakoak, eta abar; adjektibo-sortzaileen artean, berriz, *edukitzea* adierazten dutenak, *posizioa* adierazten dutenak, eta abar.

Era honetan adiera bertsua duten atzizkiak, edo antzekoak direnak, elkarrekin agertuko dira. Alfabetoaren arabeko sailkapenak ematen ez dituen aukerak ematen ditu sailkapen modu honek.

Hala ere, honek ere badu arazo bat: esanahia beti labainkorra izaten da. Adibidez, Azkuek *izen abstraktutzat* hartzen dituenak, Lafittek banatu egiten ditu: *egoera* adierazten dutenak, *ondorioa* adierazten dutenak. Beste nahasketarik ere sortzen dira. Azkuek *-dun*, *jabegoa* adierazteko atzizkia dela dio; *-ti*, *-tsu*, ugaritasuna adierazteko; *-kor* eta *-bera*, *pasionales* omen dira. Lafittek, berriz, *-tsu*, *-dun* ugaritasuna adierazteko atzizkiak direla dio.

V.3.3. Oinarriaren kategoria (Ithurri, 1979)

Beste batzuetan lehen kriterioa hitz berrien kategoria da, eta bigarrena oinarriaren kategoria; izen-sortzaileen artean, oinarrian izena dutenak, adjektiboa dutenak, adberbioa dutenak, eta abar; adjektibo-sortzaileen artean, oinarrian izena dutenak, adjektiboa dutenak, eta abar.

Baina Ithurri-ren kasuan, lehen kriterioa oinarriaren kategoria da: izenari lotzen zaizkion atzizkiak, adjektiboari lotzen zaizkion atzizkiak, eta abar. Oinarria zein den, halako taldean sailkatuko da.

V.3.4. Lanen ondorioak

Aukerak ikusi ondoren, aipatu behar da gaur egungo egoera ikusiz, egokiagoa dela lehen kriterio bezala eratorriaren kategoria hartzea; bigarren kriterioa interpretazioa; eta hirugarrena oinarriaren kategoria. Hala ere, honek garrantzi gutxi izango du tesiaren idazketan, sortuko den datu-basean galdeketa egiten duenak erabaki dezakeelako zein irizpideren arabera sailkatu atzizkiak. Sortuko dugun ezagutza-basean eratorriak eta oinarriak lotuta izango ditugu, eta atzizki lexikalak ere agertuko dira.⁶⁹

Hemendik aurrera atzizki bakoitza ikusiko dugu eta hauek EHn duten isla.

V.4. Eratorpena *Euskal Hiztegia*n

EHn ez da markatzen sarrera-hitz bat eratorria ote den edo ez. Beraz, lehen esan dugunez, hiztegia agertzen diren sarreren artean eratorriak zeintzuk diren ezagutu beharko dugu lehenik. Horretarako, aipatu dugun Ixa Taldean garatutako *Morpheus* analizatzailer morfologikoa (Urkia, 1997) erabili dugu, baina dituenak baino atzizki lexikal gehiago txertatuz. Kontua da *Morpheusek* oraingoz oso atzizki lexikal gutxi ezagutzen dituela,⁷⁰ eta lan honetarako berez ezagutzen dituenak baino gehiago erabili beharko ditugu.

⁶⁹ Gure helburuetako bat definizioei esker eratorriaren adiera ezagutzea denez, atzizkien taula azaltzen dugunean, atzizki bakoitzaren aukera guztiak elkarrekin jarriko ditugu, duen anbiguotasun maila ikus dadin.

⁷⁰ *Morpheusek* EDBLn landuta dauden atzizki lexikalak ezagutzen ditu, eta, gaur egun 21 dira: *ño, txo, tzar, garren, pe, na, ka, xe, te, tze, arazi, tsu, gile, dun, ki, tasun, dura, pen, men, mendu, garri*.

Lehenik zehaztu behar izan dugu zein atzizki lexikal txertatu analizatzailan. Horretarako sarrera-hitz guztiak hartu ditugu, eta zeintzuk diren usuen agertzen diren hitz-amaierak ikusi. Horiekin zerrenda bat osatu dugu atzizki lexikal direnak erauziz. Zerrenda horretan hitz konposatuak sortzen dituzten elementuak ere agertu zaizkigu —*aldi*, *bide*, eta abar—; halakoak hiztegi-sarrera baten elementu gisa askotan agertzen dira, eta hitz konposatua eta oinarriaren arteko erlazioa bereizteko gai izango garelakoan, modu berean landu ditugu.

Beraz, lehenik analizatzaille morfologikoa aplikatu diegu sarrera-hitzei eta hortik balizko hitz eratorriak erauzi ditugu. Behin horiek izanda, hauen definizioak aztertu ditugu eta ikusi ea zeintzuk diren agertzen dituzten ezaugarri bertsuak, eta ea horien bitartez jakin ahal dugun eratorri horren esanahia zein den. Esan dugun bezala, atzizki lexikal batek esanahi bat baino gehiago izan ditzake. Adibidez, *-tza* atzizkiarekin osatutako hitza *lanbidea* izan ahalko da —*olerkaritza* modukoetan— edo *lekua* —*sasitza* bezalakoetan—. Beraz gure helburua izango da eratorri horren interpretazioa zein den jakitea, eta, hori jakin ondoren, zehaztu ahal izango dugu zein den eratorriaren eta oinarriaren arteko erlazioa. Kasu honetan, *olerkaritza* jakin ahal izango dugu eratorriak adierazten duen *lanbidea* olerkariarena dela, eta bestean, *sasitzak* adierazten duen *lekuan*, *sasi* asko izango dugula. Bi hauek oso definizio mota ezberdinak dituzte hiztegian, eta guk, definizio-patroiak zeintzuk diren detektatuz, bat edo beste den zehaztu ahalko dugu.

olerkaritza A0 iz. Olerkariaren **lanbidea**.

sasitza A0 iz. Sasidia, sasiz beteriko **lekua** edo **saila**.

Behin definizioak aztertu ondoren, eratorriaren esanahia zehazteko garaian konturatu gara oso garrantzitsua izan daitekeela aurretik *sinonimiarekin* egindako lana. Askotan agertzen da hitz eratorri bat sinonimo baten bitartez definitua. Beraz, deskripziorik ez dagoenez definizio-eremu horretan, ezin zehaztuko dugu zein den sarrera-hitzaren interpretazioa. Baina, sinonimoa detektatzen badugu, eta suposatuko dugunez bi sinonimo horien interpretazioa bera dela, sinonimo hori hiztegian bilatu ahalko dugu eta deskribapen bat baldin badu definizio-eremuan, orduan gai izango gara zein interpretazio duen jakiteko. Eta, ondorioz, lehenengo sarrera-hitzaren interpretazioa ere zehaztu ahal izango dugu. Halako adibidea V.6. atalean ikusiko dugu.

Eta sinonimoa hitz eratorria dela ezagutuz gero —sinonimo horiei analizatzaile morfologikoa aplikatuz— interpretazioa ezagutzeko beste modu bat dugu. Sinonimo eratorriak duen atzizkiaren arabera oinarriaren eta eratorriaren arteko erlazioa zehaztu ahalko dugu. Adibidez, *-tza* atzizkiak interpretazio bat baino gehiago ditu: *nolakotasuna*, *lanbidea*, *lekua*, *ekintza*. Eta *simaurtzaren* definizioan *sinonimo* bat dugu: *simaurtegi*. *-tegik* beti *lekua* adierazten duenez, jakin ahalko dugu *simaurtza* ere *lekua* izango dela, eta jakin ahalko dugu *leku* horretan *simaurrea* izango dela.

Adibidean ikusi dugunez, gerta daiteke sarrera-hitzak duen atzizki lexikala eta definizioan agertzen den atzizki lexikala anbiguoak izatea, baina biek dituzten interpretazioen arteko ebakidura bakarra izatea. Orduan pentsatu beharko dugu ebakidura den hori izango dela bi sinonimo horien interpretazioa. Hala ere, adiera erabakitzeke metodo hau ez da definizioa erabiltzea bezain zehatza, eta bazterreko kasuetan besterik ez dugu aplikatuko.

Sinonimoen definizioak erabiltze honi *sinonimiaren hedadura* deitu dugu. Oraingoz ez dugu aplikatu, baina V.6. ataleko adibidean ikusiko dugunez, informazioa asko heredatzen denez, etorkizunean aplikatu beharrekoa izango da.

V.5. Euskarazko eratorpen-atzizkiak EHn

Hiztegiarekin lan egiten dugunez, guri gehien interesatuko zaizkigun eratorpen-atzizkiak hiztegiko sarreren artean agerpen nabarmena dutenak dira. Horretarako, lehenik jakin behar dugu hiztegiko definizioen artean zeintzuk diren amaiera usuenak, eta horietatik erauzi zeintzuk diren atzizki eta zeintzuk ez diren deus. Izan ere, kontuan izan behar da zerrenda honetan ez direla atzizki lexikalak bakarrik izango. Batzuetan aldaerak ere sartu behar izan ditugu, analizatzaileak erregela morfo-fonologikoak ezin dituelako baliatu —une honetan aztertu gabe daudelako—. ⁷¹

⁷¹ Erregela hauek aplikatu nahi izanez gero, atzizki bakoitzak dituen erregela fonologiko guztiak zehaztu beharko lirateke, eta lan honen helburu zuzena hori ez denez, modu sinpleago batean egitea erabaki dugu, nahiz eta zenbaitetan aldaerak atzizki lexikal gisa tratatu behar izan ditugun. Hauek gain, zerrendan konposizioa egiten duten elementuak ere agertu dira. Guk, zerrenda eratzeke garaian interesgarri izan daitezkeen maiztasun handieneko amaierak jarri ditugu.

Zenbait letra-segida askotan agertzen dira sarrera-hitzen amaieran —*sun* (825 aldiz) eta *ria* (509 aldiz)—, baina horiek ez dira atzizki lexikalak. *ari* —448 aldiz— eta *tza* —443 aldiz— letra-segidek, aldiz, badute atzizki lexikalen itxura, eta horiek ontzat emango ditugu azterketa honetarako. Modu honetako azterketa eginez, hona analizatzailean integratu ditugun atzizkien zerrenda:

ada	alde	aldi	antza	antzia	arazi	ari	aro
arte	asun	be	bera	bide	buru	dade	dar
date	degi	di	dore	du	dun	dura	era
eria	erraz	eta	etsi	ezia	ezin	gabe	gai
gailu	gaitz	garri	gile	gin	gintza	go	goa
gorri	grafia	gune	ismo	ista	kada	kaitz	kara
kari	keria	keta	ki	kide	kin	kintza	kizun
ko	koa	koi	kor	kume	kunde	kuntza	lari
le	leku	liar	logia	logia	men	mendu	mentu
ondo	ontzi	orri	pe	pen	tade	tar	tara
tari	tasun	tate	te	tegi	ti	tiar	toki
tsu	tu	tun	tura	txe	txo	tza	tzaile
tzar	tze	tzia	ume	zahar	zain	zale	zia
zio							

60. taula. Analizatzaileari gehitu dizkiogun eratorpen-atzizkiak eta konposizio-elementuak.

Oraingoz atzizki bakoitzari ez diogu inongo murriztapenik jartzen, baina etorkizunean, atzizki bakoitzak oinarrian hartzen duen elementua zein kategoriatakoa den zehaztu ondoren, zehaztu beharko da zeintzuk diren onargarri eta zeintzuk ez (ikus V.9.2. atalean etorkizuneko lana).

Definizioak aztertzen hasi baino lehen jakin beharko dugu atzizki bakoitzaren azpian zein erlazio dugun. Egokiena litzateke atzizkiak multzokatzea, eta gure helburua eratorriaren eta oinarriaren arteko erlazioa erauztea denez, agian egokiena litzateke horren arabera multzokatzea. Kontua da, oraingoz ez dugula atzizki lexikalen halako errepresentaziorik ikusi, eta zenbaitetan oso zaila dela erlazioa etiketa bakar batekin zehaztea. Beste hizkuntzetan egin diren lanei erreparatuta, interpretazioen arabera sailkatzea erabaki dugu, eta parafraseatu eratorriaren eta oinarriaren artean zein erlazio dagoen.

Normalean interpretazioak eratorriaren eta oinarriaren artean zuzenean sortuko dira, eta esango dugu halakoetan bi elementuren arteko erlazioa dugula. Zenbaitetan adjektibo eta aditz baten arteko erlazioa izango da, berez izenen eta aditzen arteko erlazioa dena, baina hala utziko dugu —aditz baten *berezko objektua adjektibo* bat dela esatea, kasu—.

Baina bestetan, tartean *aditz* edo *ekintza* bat dagoela sumatuko dugu. Dagoeneko azaldu dugu *–gileren* kasua. Halakoetan eratorriaren eta oinarriaren artean *berezko objektu* erlazioa dagoela esaten badugu, tarteko aditz baten beharra sumatuko dugu. Aditz hori zein den definiziotik erauzi beharko dugu —ahal denetan, behintzat—, eta beraz, erlazioa hiru elementuren artekoa dela esango dugu.

Hona eratorri eta oinarrien arteko erlazioak bi multzotan: bi elementuren arteko erlazio direnak eta hiru elementuren arteko erlazioak.

V.5.1. Bi elementuren arteko erlazioak

Atal honetan atzizki bakoitzak izan ditzakeen oinarriekin dituen erlazioak azalduko ditugu. Hemen taula laburtua paratu dugu, eta B. eranskinean, atzizki bakoitzari buruzko informazioa dagoen lekuan taula osoa eman dugu.

Hona ondoko taulan zutabeka eman dugun informazioa: atzizki lexikala⁷², honen interpretazio parafraseatua —*interpretazioa*—,⁷³ oinarriaren kategoria —*oinarria*—, eratorriaren kategoria —*eratorria*—, adibideak —*adibidea*—, eta definizio-patroi nagusia —*definizio-patroia*—.

Adibidez, *–ari*, *–lari*, *–kari*, *–tari* atzizki lexikalek edo aldaerek *ADI(t)zen duen pertsona* interpretazioa duenean, oinarria aditza da —parafraseoan agertzen den *ADI* hori, hain zuzen ere—, eta eratorria izena edo adjektiboa da. Eratorri hauen definizioetan *–t(z)en duena* definizio-patroia agertu ohi da.

beilari A1 iz. Beilat**zen duena**.

marraskari H2A0 izond. Marraskat**zen**, hortzikat**zen duena**.

beilariren interpretazioa *beilatzen duen pertsona* da, eta marraskarirena *marraskatzen duen pertsona*. Nahiz eta definizioetan bestelako aditzak agertu, oinarrian dagoen aditza hartzen dugu interpretazioa emateko garaian.

Askotan, interpretazioa eta definizio-patroia bat datoz, baina ez beti. Definizio-patroietan, zenbaitetan, izartxoak paratu dugu (*) edozer ager daitekeela adierazteko.

⁷² Atzizkiak eta aldaerak bildu egin ditugu espazio-kontuak direla-eta. Eranskinean emandako tauletan, berriz, leku bera dute bai atzizkiek eta bai hauen aldaerek.

⁷³ Parafraseatzerakoan agertzen diren kategoriak oinarriarenak dira. *–ariren ADI(t)zen duen pertsona* baldin bada *beilariren* interpretazioa, *beilatzen duen pertsona* dela adierazi nahi dugu.

atzizkia	interpretazioa	oinarria	eratorria	adibidea	definizio-patroia
arazi	ADIt(z)era behartu	adi	adi	bataiarazi	-ra behartu
ari, lari, kari, tari	ADIt(z)en duen pertsona	adi	ize / adj	beilari, marraskari, marrumalari	-t(z)en duena
ari, kari	ADIt(z)eko tresna	adi	ize	oinazekari	-t(z)eko tresna
	IZEren laguntzaz sortzen den soinua	ize	ize / adj	ezpainkari	
	ADIZ sortzen den soinua	adi	ize / adj	igurzkari	
	IZE denborako maiztasuna duen aldizkaria	ize	ize / adj	hilabetekari	-en aldizkaria
bera	ADIt(z)eko joera duena	adi	adj	maitabera	-t(z)eko joera duena
	IZE joera duena	ize	adj	gosebera	-t(z)eko joera duena
di	IZE asko dagoen lekua	ize	ize	pinudi	-z beteriko lekua
	IZEren multzoa	ize	ize	ontzidi	-en multzoa
dun	IZE duena	ize	ize / adj	kaskodun, gorpuzdun	* d(it)uena
dura	ADIt(z)ea	adi	ize	ebakidura	-t(z)ea
	ADJ izatearen nolakotasuna	adj	ize	hezadura	-tasun
era	ADIt(z)ea	adi	ize	irteera	-t(z)ea
	ADIt(z)eko modua	adi	ize	egoskera	-t(z)eko era
	IZE / ADJ ezaugarria duen neurria	ize / adj	ize	lodiera	-en neurria
	IZEko hizkuntza; IZEren hizkuntza	ize	ize	bulgariera	-ko hizkuntza
eria	IZEren multzoa	ize	ize	ikasleria	-en multzoa
gailu	ADIt(z)eko tresna	adi	ize	destilagailu	-t(z)eko tresna
garri	ADI daitekeena	adi	adj	gomendagarri	* merezi duena; * daitekeena
garri	ADIt(z)en duena	adi	ize / adj	harrigarri, sendagarri	-t(z)en duena
go, goa	IZE / ADJ izatearen nolakotasuna	ize / adj	ize	erkidego, eskalego	-en nolakotasuna
	IZEren lanbidea	ize	ize	sorgingo	-en lanbidea
	IZEren multzoa	ize	ize	udaltzaingo	-en multzoa
kada	IZEk duen eduki-neurria	ize	ize	basokada	-en edukia
	IZEz emandako kolpea	ize	ize	adarkada	-z jotako kolpea
	ADIt(z)ea	adi	ize	ziztada	-t(z)ea
keria	IZE / ADJ izatearen nolakotasun gaitzesgarria	ize / adj	ize	gozokeria	-tasun gaitzesgarria
	IZE / ADJren ekintza	ize / adj	ize	maltzurkeria, pailazokeria	-i dagokion egitea
keta	ADIt(z)ea	adi	ize	marrazketa	-t(z)ea
	IZEren multzoa	ize	ize	gazteketa	-en multzoa
ki	ADIt(z)eko tresna	adi	ize	epaiki	-t(z)eko tresna
	IZEren haragia	ize	ize	abelgorriki	IZE okela
	IZEren zatia	ize	ize	liburuki	-etako bakoitza
	IZEzko gauzakia	ize	ize	buztinki	-zko gauzaki

61. taula. Bi elementuren arteko erlazioak dituzten atzizki lexikalen ezaugarriak.

atzizkia	interpretazioa	oinarria	eratorria	adibidea	definizio-patroia
kin	ADIt(z)en dena	adi	ize	eranskin	-t(z)en den zera
ko	IZEn dagoen pertsona	ize	ize	itaurreko	* dagoena
	IZEn ematen den kolpea	ize	ize	ipurdiko	* kolpea
	IZEn eramaten den jantzia	ize	ize	buruko	* jantzia
koi	ADIt(z)eko joera duena	adi	adj	eskurakoi	erraz *-t(z)en dena
	IZEkiko joera	ize	adj	dirukoi	-zalea
kor	ADIt(z)eko joera duena	adi	adj	erorkor	-t(z)eko joera duena; erraz *-t(z)en dena
kunde	ADIt(z)ea	adi	ize	hazkunde	-t(z)ea
kuntza	ADIt(z)ea	adi	ize	bahikuntza	-t(z)ea
	IZEren lanbidea	ize	ize	medikuntza	-en lanbidea
le, tzaile	ADIt(z)en duen pertsona	adi	ize / adj	goresle, eraisle, gazitzaile, erretzaile	-t(z)en duena
men, pen	ADIt(z)ea	adi	ize	ukamen, bedeinkapen	-t(z)ea
men	ADIt(z)eko ahalmena	adi	ize	asmamen	-t(z)eko ahalmena
tar / dar	IZEn sortutako pertsona	ize	ize / adj	baserritar, bizkaitar	-ko biztanlea, -n sortua
	IZEren jarraitzailea	ize	ize / adj	mahomatar, kalbindar	-en jarraitzailea
tasun	IZE / ADJ izatearen nolakotasuna	ize / adj	ize	haurtasun, hauskortasun	-en nolakotasuna, ADJ denaren nolakotasuna
	IZE / ADJ izatearen egoera	ize / adj	ize	erbestetasun, drogazaletasun	-en egoera, ADJ dagoenaren egoera
te, tze	ADIt(z)ea	adi	ize		-t(z)ea
te	IZE ezaugarria duen denbora	ize	ize	berote	* denbora; * aldia
	IZE multzoa	ize	ize	zortzikote	* taldea
tegi	ADIt(z)eko lekua	adi	ize	salerostegi	-t(z)eko tokia
	IZEren multzoa	ize	ize	arautegi	-en multzoa
ti	ADIt(z)eko joera duena	adi	adj	izuti	erraz *-t(z)en dena
	IZE joera duena	ize	adj	setati	-t(z)eko joera duena
ti (di)	IZE asko dagoen lekua	ize	ize	sahasti	* lekua
tsu	IZE duena	ize	adj	adarrondotsu	-z betea; IZE (asko) dagoena
tza	IZEren lanbidea	ize	ize	aizurlaritza	-en lanbidea
	IZE asko dagoen lekua	ize	ize	sasitza	* saila; * lekua
	ADIt(z)ea	adi	ize	eskaintza	-t(z)ea
	IZE / ADJ izatearen egoera	ize / adj	ize	mirabetza, haurduntza	-en egoera
	IZE / ADJ izatearen nolakotasuna	ize / adj	ize	arintza, adiskidetza	-en nolakotasuna
zio	ADIt(z)ea	adi	ize	benedizio	-t(z)ea

62. taula. Bi elementuren arteko erlazioak dituzten atzizki lexikalen ezaugarriak.

V.5.2. Hiru elementuren arteko erlazioak

Bestetan, erlazioan hiru elementuk hartzen dute parte. Nahiz eta guk esan *pilotari* eta *pilotaren* artean *berezko objektu* erlazioa dagoela, hor tartean definiziotik erauzi behar dugun aditza dago, kasu honetan *jokatu* aditza.

pilotari A0 iz. Pilotan *jokatzen* duen pertsona; bereziki, pilota jokoa lanbidetzat duena.

Izan ere, *berezko objektua* aditz eta izen baten arteko erlazioa da, eta ez bi izenen artekoa, baina hitz eratorrian bi izen ditugunez, erlazioa bi izenen artekoa dela diogu. Honi buruzko informazio zehatza B. eranskinean duzue, atzizki lexikal bakoitzari dagokion atalean.

Aditza ez da bera kasu guztietan. talogileren kasuan, adibidez, *egin* da talo eta talogileren artean dagoen aditza. Eta urregileren kasuan, landu da urre eta urregileren artean dagoena.

talogile A0 iz. Taloak *egiten* dituen pertsona.
urregile A0 iz. Urrea *lantzen* duen pertsona.

atzizkia	interpretazioa	oinarria	eratorria	adibidea	definizio-patroia
ari, lari, kari, tari	IZEa ADIt(z)en duen pertsona	ize	ize	pilotari, neurtizlari	IZE ADIt(z)en duena
gile, gin	IZEak ADIt(z)en dituen pertsona	ize	ize	talogile, upagin	IZE ADIt(z)en duena
gintza	IZEak def-ADIt(z)eko jarduna (lanbidea)	ize	ize	bidegintza	IZE egiteko lanbidea
keta	IZE def-ADIt(z)ea	ize	ize	egurketa	IZE *-t(z)ea
tegi	IZE def-ADIt(z)en den lekua	ize	ize	behitegi	IZE *(z)en den lekua

63. taula. Hiru elementuren arteko erlazioak dituzten atzizki lexikalen ezaugarriak.

Aurrekoen artean, *-gile*, *-gin* eta *-gintza* talde honetan bakarrik daude —hiru elementuren arteko erlazioetan, alegia—, eta *-ari*, *-lari*, *-kari*, *-tari*, *-keta* eta *-tegi* bi elementuren arteko erlazioetan ere agertzen dira.

V.6. Atzizkien interpretazioak nola zehaztu definizioak erabiliz

Hiztegi-sarrera eratorriak ezagutu ondoren, oinarriaren kategoria zein behar duen aztertu ondoren, eta eratorriaren balizko adierak zeintzuek behar duten aztertu

ondoren, eratorri horiek eta hauen definizioak aztertuko ditugu. Oro har, adiera bakoitzerako hiztegigileak definizio-eskema jakin batzuk erabiliko ditu; horiek erabilia zehaztu ahalko dugu eratorriaren interpretazioa zein den, eta, horren arabera, eratorriaren eta oinarriaren arteko erlazioa zein den.

Prozesu osoa adibide bat erabiliz azalduko dugu. Demagun *-tzadun* eratorriak aztertu nahi ditugula. 64. taulan atzizki honek izan ditzakeen oinarriak eta interpretazioak daude.⁷⁴

atzizkia	interpretazioa	oinarria	eratorria	adibidea	definizio-patroia
tza	IZEren lanbidea	ize	ize	aitzurlaritzza	-en lanbidea
	IZE asko dagoen lekua	ize	ize	sasitza	IZE saila
	ADIt(z)ea	adi	ize	eskaintza	-t(z)ea
	IZE / ADJ izatearen egoera	ize / adj	ize	mirabetza, haurduntza	-en egoera
	IZE / ADJ izatearen nolakotasuna	ize / adj	ize	arintza, adiskidetza	-en nolakotasuna

64. taula. *-tza* atzizkidun hitz eratorrien balizko adierak.

-tza atzizkiak oinarrian izenak, aditzak eta adjektiboak izan ditzake, eta, beraz, horren arabera ezingo dugu analisi morfologikoan dugun loturarik baztertu. Hona *sasitza* hitzarentzat analizatzaile morfologikoak ematen dituen aukerak:

sasitza	sasitu	KAT_ADI	tza
sasitza	sasi	KAT_ADJ	tza
sasitza	sasi	KAT_IZE	tza

Horren arabera, eta 64. taulan ditugun interpretazioen parafraseak kontuan hartuta, 65. taulan jarri ditugu *sasitza* hitzaren balizko interpretazio guztiak.

oinarria	interpretazioa
aditza	sasitzea
adjektiboa	sasi izatearen egoera edo nolakotasuna
izena	sasi asko dagoen lekua
	sasi izatearen egoera edo nolakotasuna
	sasiaren lanbidea

65. taula. *sasitza* hitzaren balizko interpretazio guztiak.

Beraz, puntu honetara arte interpretazio guztiak zilegi dira. Ondoren, zilegitasun hori egiaztatzeko, hiztegiko definizioa erabiliko dugu. Erregelak egin ditugu interpretazio bakoitzerako. Horretarako III.7. atalean aipatu dugun *Murritzapen Gramatika* erabili dugu, eta morfologikoki analizatutako definizioaren gainean aplikatu dugu definizio-patroien arabera atzizki bakoitzaren interpretazioak ezagutzeko garatutako gramatika.⁷⁵ Gainera, lehendik ditugun markak erabiltzen saiatuko gara; izan ere,

⁷⁴ Hau 62. taularen zatia da.

⁷⁵ Atzizki lexikal bakoitzaren interpretazioak kontuan hartuta, atzizki bakoitzari dagokion gramatika sortu dugu, eta, guztiak, I.7. eranskinean eman ditugu.

askotan gertatuko da sarrera-hitz eratorri bat *sinonimo* baten bitartez definitzea, eta horiek ezagutzea oso garrantzitsua da *sinonimiaren hedadura* aplikatu ahal izateko.

Adibidearekin jarraituta, definizioan bilatuko dugu ea leku,⁷⁶ egoera, nolakotasun edo lanbide⁷⁷ agertzen diren, eta horren arabera, interpretazioa zein den zehaztu.

sasitza A0 iz. Sasidia, sasiz beteriko **lekua**
edo **saila**.

Beraz, kasu honetan leku eta sail agertzen direnez, sasitza horren interpretazioa *lekua* dela esango dugu, eta oinarriarekin duen erlazioa *sasia egotea*.

-tza atzizkiarekin lortu ditugun emaitzak begiratzuz, analizatzaile morfologikoak 327 —326 *izenen* artean eta 1 *adjektiboen* artean— bikote lortu ditu. Sarrera-hitz baten barruan, analizatzaile morfologikoak bikote bat baino gehiago dagoela erabaki dezake:

eratorria	oinarria	oinarriaren kategoria	atzizkia
lumatza	lumatu	ADI	tza
lumatza	luma	IZE	tza

66. taula. lumatza sarrera-hitzaren analisi morfologikoa.

Bikoizketa hauek kontuan izanez, -tza duten sarrerak 286 dira —285 *izen* kategoriakoak eta 1 *adjektibo* kategoriakoa⁷⁸—. Gainera, sarrera-hitz forma batek adiera bat baino gehiago izan ditzake, baina analizatzaileak behin ezagutuko du.

errektoretza A0N1 iz. Errektorearen lan egoitza.

errektoretza A0 iz. Errektorearen lanbidea eta kargua.

Analizatzaileak lortutako bikote horietatik, 87k dute oinarrian aditza, 45ek adjektiboa eta 195ek izena —1 *adjektibo* kategoriako eratorrien artean—. Atzizki bakoitzari buruzko halako informazioa B. eranskinean jarri dugu.

⁷⁶ leku diogunean, halako hitz bat bilatuko dugu: leku, toki, sail, eremu, baso, oihan. Etorkizunean multzo hauek definitzeko garaian, *EuskalWordNeteko synsetak* erabili ahalko ditugu.

⁷⁷ lanbide diogunean, halako hitz bat bilatuko dugu: lanbide, kargu, aginte, titulu, ogibide, antze, bizibide, agintaritza, jardun, maila, lana, betekizun, egiteko.

⁷⁸ espioitza da adjektiboen artean dagoen adibidea, eta errorea da, definizioa izenarena baitu: espioi lana.

Eratorriak	izen		izen-adjektibo	adjektibo	
Oinarriak	ADI	87	ADI	ADI	
	ADJ	45	ADJ	ADJ	
	IZE	194	IZE	IZE	1
	Denera	326	Denera	Denera	1

67. taula. –*tzadun* eratorriak: eratorriaren eta oinarriaren kategoriaren arabera kopuruak.

286 sarrera-hitz horien definizioak aztertuta, 127 sarrera-hitz sailkatu ditugu interpretazioen arabera. 68. taulan dago multzokaketaren emaitza. honekin:

nolakotasuna	9
egoera	4
lekua	13
lanbidea	101
denera	127

68. taula. –*tzadun* eratorrien ezagututako interpretazioen kopurua.

286 horietatik, definizioan sinonimoa duten sarrera-hitzak 102 dira. Horietako 19k, sinonimoaz gain, deskribapena dute, eta 19 horietatik 6 kasutan lortzen dugu sarrera-hitzaren interpretazioa; ondorioz, 96 sarrera-hitzen interpretazioa sinonimoen interpretazioen bitartez lortzen saiatuko gara, lan hau automatikoki egitea merezi duen baieztatzeko.

Sinonimodun 96 definizio horiek aztertu ondoren, hona dugun informazioa. Eratorpenaren bitartez ezagutu ditzakegun sinonimoak 82 dira; denera 66 definiziotan daude. Eratorriak ez diren sinonimoak, berriz, 41 dira, denera 30 definiziotan.

Ezagutzen ditugun 82 sinonimoetatik, 64 sinonimorentzat lortzen dugu informazioa —*informazioa bai* errenka—, denera 52 definiziotan daude. Guztira, 59 sinonimorentzat ez dugu informaziorik lortzen —*informazioa ez* errenka—, denera 44 definizio.

	eratorria		ez-eratorria	
	definizioak	sinonimoak	definizioak	sinonimoak
informazioa bai	47	58	5	6
informazioa ez	19	24	25	35
denera	66	82	30	41

69. taula. –*tzadun* eratorrien *sinonimiaren hedadura*.

Informazioa lortzen ez dugun hauen artean, badira 3 definiziotan agertzen diren 4 sinonimo hiztegian sarrerarik ez dutenak: aringarri, suntsidura, elur-erauntsi eta aurkientza.

69. taulan informazioa duten 52 definizio horiek dituzten interpretazioak 70. taulan daude.

Interpretazioa	definizioak	sinonimoak
jarduera	26	32
lanbidea	13	16
lekua	6	8
nolakotasuna	5	5
nolakotasuna/egoera	1	2
egoera	1	1

70. taula. *-tzadun* sinonimoen interpretazioak.

Eta informazioa duen elementu bakoitzaren informazioa:

Jarduera: ez-eratorria (1), keta (4), aldi (1), di (1), tze (13), asun (1), te (4), gin (1), ki (1), tza (2), aldi (1), eta (1), go(1).

Lanbidea: go (2), tza (8), bide (3), ez-eratorria (3).

Lekua: ez-eratorria (1), tegi (2), tza (5).

Nolakotasuna: tasun (2), ez-eratorria (1), go (1), tza (1).

Nolakotasuna/egoera: tasun (2)

Egoera: zio (1).

Eta informaziorik lortu ez dugun sinonimoen informazioa hau da: hitz eratorri gisa ezagutu ditugunak 24 dira, eta ez eratorriak 35. Eratorrien artean hauek dira agertzen diren eratorpen-atzizkiak:

tasun (3), te (1), tza (4), degi (1), garri (1), dura (1), ko (2), di (1), tegi (1), pen (3), men (2), arte (1), tura (2), keria (1).

Beraz, egindako sinonimoen azterketaren ondoren, honoko emaitza dugu: 96 sinonimodun definizio horietatik, hauen interpretazioa aztertuta 52rentzat lortu dugu informazioa, hau da, %54,17rentzat. Emaitza honek berak adierazten du, oso emaitza interesgarriak lortuko ditugunez, etorkizuneko lana izango dela sinonimiaren hedadura automatikoki aplikatzea.⁷⁹

V.7. Emaitza orokorrak

Atal honek bi zati ditu: batetik, estaldura aipatuko dugu (V.7.1. atala), eta, bestetik, doitasuna (V.7.2. atala).

⁷⁹ Lan hau oraingoz ez da sistematikoki atzizki guztiekin egin.

V.7.1. Estaldura

71. taulan eta 72. taulan aztertutako atzizki bakoitzarekin ditugun estaldura-datuak daude. *Denera* zutabea *Morpheusek* ezagututako eratorri-kopurua dago; *zenbat erlazio* zutabea, erlazioa zenbatetan ezagutu dugun; eta *sinonimoak* zutabea, sarrera-hitzaren sinonimoak zenbatetan ezagutu ditugun ondoren *sinonimiaren hedadura* aplikatu ahal izateko. Azken bi zutabeetan, erlazioak ezagutzeko garaian lortu dugun estaldura —*erlazioen estaldura*—, eta sinonimiaren bitartez definitutako sarrera-hitzen estaldura —*sinonimiaren estaldura*—.

	Denera	Zenbat erlazio	Sinonimoak	Erlazioen estaldura	Sinonimiaren estaldura
tasun	912	701	172	%76,86	%18,86
garri	488	342	170	%70,08	%34,84
tzaile	467	349	106	%74,73	%30,37
tsu	433	241	187	%55,66	%43,19
keria	370	205	147	%55,41	%39,73
ko	287	29	133	%10,10	%46,34
tza	285	127	102	%44,56	%35,79
aldi	285	95	130	%33,33	%45,61
dun	278	156	92	%56,12	%33,09
ki	251	84	87	%33,47	%34,66
keta	228	11	175	%4,82	%76,75
tegi	218	50	85	%22,94	%38,99
bide	214	81	74	%37,85	%34,58
kor	171	96	70	%56,14	%72,92
lari	127	39	45	%30,71	%35,43
zio	166	0 ⁸⁰	114	%0,00	%68,67
pen	160	0	112	%0,00	%70,00
kari	143	64	62	%44,76	%43,36
gile	117	85	31	%72,65	%26,50
men	137	24	79	%17,52	%57,66
te	123	22	54	%17,89	%43,90
dura	132	16	83	%12,12	%62,88
le	93	45	38	%48,39	%40,86
di	105	32	46	%30,48	%43,81
ti	105	36	57	%34,29	%54,29
era	108	23	67	%21,30	%62,04
go	94	34	53	%36,17	%56,38
gintza	97	69	15	%71,13	%15,46
tar	92	8	42	%8,70	%45,65
kada	75	33	32	%44,00	%42,67

71. taula. Aztertutako atzizki lexikalak eta kopuruak.

⁸⁰ Erlaziorik topatzen ez dugun hauen definizioetan sinonimia askotan agertzen da, eta definizio-patroi garbirik ez dugu topatu.

	Denera	Zenbat erlazio	Sinonimoak	Erlazioen estaldura	Sinonimiaren estaldura
gailu	71	26	28	%36,62	%39,44
kuntza	63	1	48	%1,59	%76,19
kin	48	10	24	%20,83	%50,00
tze	48	0	30	%0,00	%62,50
kunde	47	0	32	%0,00	%68,09
bera	34	10	18	%29,41	%52,94
gin	37	12	16	%32,43	%43,24
tari	27	11	7	%40,74	%25,93
eria	41	10	12	%24,39	%29,27
koi	39	8	21	%20,51	%53,85
ari	27	11	12	%40,74	%44,44
dar	9	0	5	%0,00	%55,56
goa	12	3	7	%25,00	%58,33
denera	7.264	3.199	2.920	%44,04	%40,20

72. taula. Aztertutako atzizki lexikalak eta kopuruak.

71. taulan eta 72. taulan dauden atzizki lexikalak kontuan hartuta, denera, 7.264 sarrera eratorri ditugu. Horietatik 3.199tan —sarreraren %44,04— lortu dugu nolabaiteko erlazioa markatzea, eta 2.920tan —guztien %40,20— *sinonimia* erabili ahal dugu erlazioa markatzeko, *sinonimoa* ezagutzen dugulako.

definizioak	7.264
erlazioa markatuta	3.199
sinonimoa markatuta	2.920

73. taula. Eratorri-kopurua: erlazio-kopurua, sinonimo-kopurua.

Aditzetan *-arazi* atzizkia landu dugu, eta hauek dira zenbakiak:

	Denera	Zenbat erlazio	Sinonimoak	Erlazioen estaldura	Sinonimiaren estaldura
arazi	309	242	67	%78,32	%21,68

74. taula. *-arazi* aditz-arazlearekin lortutako emaitzak.

Honek izango duen erlazioa beti *arazlea* da.

Aditzak sortzen dituen beste atzizkia, eta hiztegian agerpen handia duena *-tu / -du* da, eta ondoren honi buruz hitz egingo dugu.

V.7.1.1. *-tu / -du* atzizkia

-tu/-du atzizki ez-lexikalari buruzko azterketa txiki bat egin dugu, eta hona emaitzak.

–tu/-du atzizki ez-lexikaldun aditzak asko eta asko —ia denak— dira euskaraz. Eratorri hauen eta haien oinarrien artean dagoen erlazioa desanbiguatzea lortuz gero informazio handia lortuko genuke, baina berehala konturatu gara ez dela ahuntzaren gauerdiko eztula.

Gràcia *et al.*-ek (2000) –tu/-du atzizkiaren hainbat mota bereizten dituzte, eta saiatuko gara aztertzen ea bereizketa horiek ba ote duten eraginik aditzen definizioetan. Beraz, –tu/-du bakoitzari buruzko ezaugarriak banan-banan ematen saiatuko gara.

Oinarrian izena (*tu1*,⁸¹ *tu2*, *tu3*, *tu4*, *tu5*, *tu6*, *tu7*), adjektiboa (*tu8*), adberbioa (*tu9*, *tu10*) eta postposizio-sintagma (*tu11*, *tu12*) egon daiteke. Eta eratorriaren kategoria aditza da. Interpretazioei dagokienez, berriz, bat baino gehiago izango dugu:

- *tu1*, *tu2*, *tu3*, *tu8*, *tu9*: *egoera- edo nolakotasun-aldaketa*: izoztu, gizondu, lasaitu, mailakatu.
- *tu4*: *gabezia (gabetasun)*: larrutu, lumatu.
- *tu7*: *donazioa*: saritu, zigortu.
- *tu6*, *tu11*, *tu12*: *leku-aldaketa*: baztertu, saihestu, lurreratu.
- *tu10*: *errepikatzea, aktibitatea (instrumentuarekin)*: mailukatu, mokokatu.
- *tu5*: *lokazioa*: hostatu, loratu.

Kontuak honela, gure asmoa zenbait adibideren definizioak ikusi ondoren talde bateko edo besteko ziren erregelak sortzea izan da, eta batean edo bestean sailkatzea.

Berehala konturatu gara zeregin oso-oso zaila dela: *egoera- edo nolakotasun-aldaketa* multzokoak nahiko erraz harrapa ditzakegu, baina beste multzoetan irristatzeko arrisku handia dago. Erregela zenbait egin ondoren, lagin bat atera dugu jakiteko zenbat aditz etiketatzen ditugun metodo hau jarraituta. Hala, 68 adibideko laginetik, zuzen-zuzenean 11 markatu ditugu. Hauetatik 8k *egoera- edo nolakotasun-aldaketa* adierazten dute, 1ek *donazioa*, eta beste 1ek *leku-aldaketa*. Azkenik esan behar da 11 horietatik 1 dela errorea. Beraz, badirudi doitasun ona dugula (%90,91), baina estaldura benetan txikia denez (%14,71), lan hau alde batera uztea erabaki

⁸¹ –tu/du sailkatzeko erabilitako taldeen izenak dira.

dugu, nahiz eta zenbait multzo ateratzea lortu dugun. Kontuan izan behar da, hala ere, *-tu/-dudun* aditzak izugarritzko pila direla hiztegia —3.625 *-tudunak* eta 475 *-dudunak*—, eta, nahiz eta ehuneko txiki bat markatu, asko izan daitezkeela.

denera	markatuta	ondo	gaizki	estaldura	doitasuna
68	11	10	1	%14,71	%90,91

75. taula. *-tu/-duren* datu estatistikoak: estaldura, doitasuna.

Beraz, nahiz eta zenbait talde biltzea lortu, atzizki ez-lexikal hau oraingoz ez lantzea erabaki dugu, lortzen dugun estaldura oso txikia delako. Gainera taldeen arteko mugak ez dira batere garbiak.

Agian aurrera egiteko moduan *sinonimiaren hedadura* aztertu beharko litzateke. Ziur asko, *sinonimiaren hedadura* kontuan izanez gero, aditz gehiago sailkatzea lortuko genuke, baina lan hau oraingoz gerorako utzi dugu.

V.7.2. Doitasuna

Doitasuna neurtzeko hiru atzizki erabili ditugu. Lagin bat hartu dugu eta bi hizkuntzalariri eman diegu, eratorri eta oinarriaren arteko erlazioak marka ditzaten. Ondoren konparatu ditugu bi hizkuntzalariek euren artean duten adostasun-maila, hizkuntzalariek eta metodo automatikoak duten adostasun-maila, eta hiruren artean lortzen den adostasun-maila. Hona emaitzak:

	-garri	-keria	-tsu ⁸²
bi hizkuntzalari	%90	%82,8	%83
A hizkuntzalaria eta metodoa	%100	%96,2	%82
B hizkuntzalaria eta metodoa	%100	%84,6	%80
bi hizkuntzalariak eta metodoa	%100	%80,8	%72

76. taula. Eratorpenarekin lortutako emaitzen doitasuna.

Datu hau neurtzerakoan, kontuan izan ditugu bai metodoak eta bai hizkuntzalariek markatzen dituzten kasuak. Markatu gabeak kontuan hartzen ez ditugunez,⁸³ *-garriren* kasuan bi hizkuntzalarien arteko adostasuna hizkuntzalariek eta metodoak dutena baino txikiagoa da. Hizkuntzalariak ados ez dauden kasuren bat metodoak ez du ezagutu, eta guk doitasuna neurtzeko garaian metodoak markatzen ez duena ez

⁸² *-tsu* atzizkian saiatu gara *ezaugarriak* eta *egoerak* bereizten, eta nahiz eta sumatu nolabaiteko diferentzia badela, pentsatu dugu oso zaila dela bereizketa horrekin aurrera jarraitzea eta albo batera utzi dugu. Honokoak bereizketa baztertu aurreko datuak dira.

⁸³ Doitasuna neurtzeko markatutakoen artean ea zenbat dauden ondo begiratzen denez, ez markatuak ez dira kontuan hartzen.

dugu kontuan hartu. Kontuan hartzen ez ditugun horiek estaldura-arazoak dira, eta ez doitasun-arazoak. Beraz, lortutako doitasuna oso ona da.

Neurketa hauek egiteko laginak 30 eta 40 definizioen artekoak dira. Hor eskuz markatu dugu zer topatu nahi dugun, eta gero estaldura-datuak atera, markatu nahi dugun hori zenbatetan markatzen dugun jakiteko. Hona lagin horietan lortutako estaldura (Agirre eta Lersundi, 2001).

Atzizkiak	lagina	benetan dauden erlazioak	metodoak markatuta	Estaldura
-garri	30	36	27	%75,0
-keria	30	39	34	%87,2
-tsu	40	37	31	%83,8

77. taula. Aztertutako laginaren estaldura-kopurua.

V.8. Erlazioak EuroWordNeten txertatzea

Txosten honen hasiera-hasieratik aipatu dugunez, lan honen guztiaren helburua *Euskal Hiztegiaren Ezagutza-Basea* (EHEB) osatzea da. Bestalde, *EuskalWordNet*⁸⁴ osatzen ari garenez, egoki litzateke bi hauen arteko lotura gauzatu ahal izatea.

Baina guk hiztegitik erauzitako erlazioak *EuskalWordNeten* txertatu ahal izateko, bi baldintza behar dira:

- Guk dugun erlazioa ezagutza-base berezi horrek duen eskemaren arabera errepresentagarria izatea (V.8.1. atala).
- Erlazioa kontzeptuen artekoa izatea, eta ez hitzen artekoa. Kontzeptu horiek hitzen adierak izango dituzte oinarri (V.8.2. atala).

Bi hauek nola aplikatu daitezkeen aztertu dugu, eta hona azalpenak.

V.8.1. Lortutako erlazioak WordNet-en txertatzea

Dagoeneko esan dugun bezala, izendatzen ditugun erlazioak ez dira beti sinpleak. Hona *aztergarriaren* kasua.

⁸⁴ Aurrekarietan esan dugun bezala, *EuskalWordNet* hau *EuroWordNet* ezagutza-basea oinarri erabiliz osatzen ari gara, eta azken hau *ingeleza*, *espainiera*, *italiera*, *nederlandera*, *alemana*, *frantsesa*, *txekiera*, *estoniera* eta *katalana* hizkuntzak barneratzen dituen ezagutza-basea da. Beraz, *euskara* zerrenda horretan izanez gero, eta bertan loturak kontzeptu mailakoak direnez, hizkuntzartekotasunari begira urrats handia litzateke.

aztergarri. A0 izond. Aztertua izatea merezi duena.

Guk esango dugu azterturen *objektua* dela aztergarri. Baina hemen zehazten ari garen erlazioa aditz eta adjektibo baten artekoa da. *objektu* erlazioa, oro har, aditz eta izen baten artean gertatu ohi da. Beraz, benetako erlazioa litzateke aditzaren artean eta adjektibo horrek modifikatzen duen izenaren artean. Adibidez, *lan aztergarri* sintagma baldin badugu, azterturen *objektua* lana da. Oraingoz, erlazio konplexu hau ezingo da *EuroWordNet*en errepresentatu, bertan dauden erlazioak bi elementuren artekoak direlako.

Erlazioak markatzeko garaian beste arazoa izango da aditza errepresentazioan ez agertzea, baina erlazioan garrantzia izatea. Hau gertatzen zaigu *–gile* atzizkiarekin sortutako eratorriekin. Adibidez, *pilotagile* baldin badugu, erlazioa *pilotarekin* izango du eta esango dugu eratorriak adierazten duen pertsonaren *berezko objektua* *pilota* dela. Hala ere, *objektu* erlazioa aditzen eta izenen artekoa izan ohi da, eta erraza izan daiteke erauztea hor azpian dagoen aditza *egin* dela. Baina ez da beti horrela, urregilea ez baita *urrea* egiten duena, baizik eta *urrea* lantzen duena. Eta hau nahikoa erraza da hiztegieta definizioetatik ezagutzea, definizioan aditz hori esplizitu agertzen delako, nahiz eta gero —berriro ere erlazioan bikotea izan beharrean hirukotea dugunez— errepresentazioan arazoa izan.

Orain ezinezkoa bada ere, lanean dihardugu gerora erlazio konplexu hauek errepresenta daitezen. Kontzeptuen arteko loturei buruz ari garenez —ondoko atalean azalduko dugun bezala—, euskaraz lortutako erlazioak baliagarri izango dira aurretik aipatu ditugun hizkuntza guztietarako.

V.8.2. Adieren desanbiguazioa

Esan berri dugunez, kontzeptuen arteko erlazioak markatu ahal izateko, lehenik jakin behar dugu zein den sarrera-hitzarekin erlazionatuta dagoen hitz horrek errepresentatzen duen kontzeptua. Izan ere, hitz baten barruan adiera edo kontzeptu bat baino gehiago egon daitezke. Adibide gisa *amugile* hitza hartuz gero, oinarrian dugun *amu* horrek bi adiera ditu *EuskalWordNet*en:

- 02510398n zenbakidun kontzeptua, *amu* fisikoa, arrainak eta halakoak harrapatzeko erabiltzen dena.

- 03882076n zenbakidun kontzeptua, *amu* figuratua, norbait engainatzeko erabiltzen den gertaerari aplikatzen zaiona.

Guk jakin beharko dugu *amugileren* oinarrian bietako zein dagoen —edo biak ote dauden—, kontzeptuen arteko lotura gauzatu nahi izanez gero, behintzat; kasu honetan, lehenengo adiera da hor oinarrian dagoena, eta ez bigarrena.

Lehen esan badugu ere, hemen azalduko dugu guk markatutako erlazioak kontzeptu mailakoak badira, beste hizkuntzei ere baliagarri izango zaizkiela esaten dugunean zer adierazi nahi dugun. Demagun *garagardogile* hitza dugula, eta zehazten dugula *garagardogileren berezko objektua* dela *garagardo* —eta, definizioan egin *aditza* agertzen denez, *objektu* hori egitea—. Behin hau dugula, espainierako kontzeptuekin lotzen badugu, lotuta izango ditugu *cervecero* eta *cerveza* eta haien arteko erlazioa euskaraz zehaztutako bera izango da. Inork pentsatu ahalko du espainieraz ere erlazio hori argia dela *-ro* atzizkiaren azterketa eginez gero erauzi ahalko litzatekeelako, eta hala da. Baina ingelesez kontzeptu horiek adierazteko *beer* eta *brewer* ditugu. Eta hor, agerian behintzat, ez dago inongo lotura morfologikorik, baina beste hizkuntzetako lotura kontzeptuen artekoa izatea lortzen bada, bi horiek ere beste hizkuntzetan zehaztutako erlazio bera izango dute.

Kontua da kontzeptuen arteko loturak gauzatzeko adieren arteko desanbiguazioa egin behar dela, eta orain arte errore-tasa handia duen lanbidea izan da. Euskararen kasuan hau handiagoa izan daiteke *EuskalWordNet* sortzen ari den ezagutza-basea delako, eta zenbait kontzeptu oraindik txukundu gabe daudelako.⁸⁵

Guk adiera-desanbiguazioa lortu ahal izateko, kontzeptuen arteko *ebakidura*⁸⁶ erabili dugu. Euskaraz dugun hitz eratorria eta oinarria beste hizkuntzetako beraien itzulpenekin konparatu ditugu, eta haien arteko kontzeptu-ebakidura erauzten saiatu gara.

Demagun, *zilar* eta *zilar* hitzak ditugula:

⁸⁵ Hemen emango ditugun lotura-datuak 2001ekoak dira. Gaur egun *EuskalWordNet* landuagoa dagoenez, emaitza hobeak izango genituzkeela uste dugu.

⁸⁶ Sarreran esan dugun bezala, *ebakidura* hitzarekin, bi hitzek dituzten balizko kontzeptuen artean kontzeptu bera zein diren adierazi nahi dugu. Hitzak hizkuntza berekoak, nahiz hizkuntza ezberdinetakoak izan daitezke.

	eratorria	kontzeptuak ⁸⁷	oinarria	kontzeptuak
euskara	zilargile	06345632n	zilar	08206456n (eskudiru) 08819522n (substantzia)
espainiera	platero	06345632n	plata	02423695n (silverware) 08206456n (eskudiru) 08819522n (substantzia)
ingelese	silversmith	06345632n	silver	08206456n (eskudiru) 08819522n (substantzia) 03466093n (zilar koloreko) 02424132n (zilarrezko mahai tresneria)

78. taula. Hizkuntzen arteko loturak.

zilargilek, platerok eta silversmithek kontzeptu bakarrari egiten diotenez erreferentzia ez dago arazorik. Oinarriari erreparatzen badiogu, berriz, euskaraz zilar hitzak bi adiera ditu; espainierako plata hitzak, aldiz, hiru adiera ditu; eta ingeleseko silver hitzak lau adiera. Beraz, kontzeptuen arteko ebakidura erabiliz gero, ondorioztatuko dugu platero eta silversmith horiek oinarrian dituzten plata eta silver eskudiruaren edo substantziaren erreferentzia direla.

Metodologia honekin lortutako emaitzak ebaluatzeko, 100 *–giledun* eratorri hartu ditugu, eta laginean dauden oinarri-eratorri bikoteak aztertu ditugu:

Loturarik ez	%53,0
Eradorria	%22,0
Oinarria	%9,0
Biak	%16,0

79. taula. Loturen estaldura
100 adibideekin.

Erabilitako algoritmoak eratorriak %22an lotu ditu, oinarriak %9an eta biak %16an. Loturarik gabe gelditu dira kasuen %53a.

Doitasunari dagokionez, eratorrien kasuan, %97 zuzenak dira. Oinarriari dagokionez, %80 da erabat zuzena, eta %15ean kontzeptu zuzena lotu da, nahiz eta beste kontzepturen bat ere lotzen den. Honen adibide da *antzerkigile* hitza. Oinarrian dugun *antzerki* horrek bi kontzepturen ebakidura du, bata, *liburua* —*antzezlan*— eta bestea *antzekizuna*. Eta bigarren hau ez da izango antzerkigilek oinarrian duen kontzeptua. Beraz, esango dugu ona ezagutzen duela —*liburua*—, eta ona ez den beste bat ere bai —*antzekizuna*—.

⁸⁷ Kontzeptuak *EuroWordNet*en berez zenbakiak dira, eta kontzeptu horiek duten zenbakia paratu dugu.

	Eratorria (%)		Oinarria (%)	
	Denera	Adiera bakarra	Denera	Adiera bakarra
Zuzena	%97,3	%52,6	%80,7	%7,7
Batzuen artean zuzena			%15,4	
Ez-zuzena	%2,7		%3,8	

80. taula. Adieren desanbiguazioaren doitasuna.

Kontuan izan behar da eratorrien kasuan adiera bakarrekoak asko direla —%52,6—, eta horrek adiera-desanbiguazioan asko laguntzen du. Oinarrien kasuan, berriz, adiera bakarrekoen zenbakia izugarri jaisten da —%7,7— baina, hala ere, desanbiguazio-doitasuna ez da hainbeste jaisten.

Behin *EuroWordNet*ekin lotzeko ahalegina azaldu ondoren, etorkizuneko lanak aipatuko ditugu.

V.9. Etorkizuneko lana

Atal honetan etorkizunean garatzeko geratu diren hainbat lan azaldu ditugu: atzizki lexikal berriak lantzea, oinarriaren kategoria kontuan hartzea, sinonimiaren hedadura, eratorpenarekin egindako lana eta *EuskalWordNet*en arteko lotura masiboa egitea.

V.9.1. Atzizki lexikal gehiago lantzea

Oro har hiztegian agerpen handirik ez duten eratorpen-atzizkiak ez ditugu landu. Hala ere, zerrendan konposizio-elementuak ere agertzen dira, eta hauek ere aztertu beharko ditugu.

81. taulan dagoeneko analizatzaile morfologikoan oraingoz aztertu ez ditugun, baina dagoeneko analizatzaile morfologikoan integratu ditugun atzizki lexikalak —edo konposizio elementuak— paratu ditugu. Analizatzaileak ezagutzen dituen eratorri eta konposatuen kategorien arabera sailkatu ditugu.

Adibidez, analizatzaileak ezagutu dituen *-gabedun* hitzak 151 dira: 13 izen kategoriakoak, 127 adjektibo kategoriakoak, eta 11 bi kategoriatakoak.

	izen	ize/adj	adj	denera
gabe	13	11	127	151
txo	122	0	7	129
gai	94	3	12	109
zale	19	19	59	97
pe	91	1	0	92
mendu	85	0	0	85
gaitz	9	1	70	80
kizun	66	0	6	72
buru	65	0	2	67
ezin	8	0	57	65
zain	64	0	1	65
toki	63	0	1	64
gune	60	0	0	60
kide	44	2	9	55
alde	54	0	0	54
leku	51	0	0	51
ondo	45	0	1	46
ismo	39	0	1	40
ista	13	16	11	40
tura	36	0	0	36
kume	35	0	0	35
tzar	31	2	1	34
tun	17	8	6	31
arte	29	0	0	29
degi	29	0	0	29
eta	26	0	0	26
gorri	24	0	2	26
kaitz	10	0	15	25
be	18	1	5	24
logia	23	0	0	23

	izen	ize/adj	adj	denera
tiar	8	3	12	23
asun	22	0	0	22
ada	21	0	0	21
ada	21	0	0	21
ontzi	18	0	3	21
zia	21	0	0	21
tara	20	0	0	20
dore	18	0	1	19
kara	5	2	12	19
mentu	16	0	0	16
txe	13	0	1	14
erraz	1	0	11	12
tate	12	0	0	12
antza	11	0	0	11
grafia	10	0	0	10
kintza	8	0	0	8
tzia	7	0	0	7
aro	4	0	2	6
dade	5	0	1	6
liar	5	0	0	5
liar	5	0	0	5
zahar	3	1	0	4
ezia	3	0	0	3
orri	2	0	0	2
antzia	1	0	0	1
date	1	0	0	1
koa	0	1	0	1
tade	1	0	0	1
ume	1	0	0	1

81. taula. Landu gabe utzi ditugun eratorpen-atzizkiak eta konposizio-elementuak.

V.9.2. Oinarriaren kategoria kontuan hartzea

Oraingoz atzizki bakoitzari ez diogu inongo murriztapenik ezartzen, baina ondorengo pausoren batean, atzizki bakoitzak oinarrian hartzen duen elementua zein kategoriatakoa den zehaztu ondoren, zehaztu beharko da zeintzuk diren onargarri eta zeintzuk ez. Hartara, analizatzaileak emandako zenbait bikote trakets saihestuko ditugu.

-tegi adibide gisa hartuz gero, honek oinarrian izena izango duela esango dugu, eta halakoak 151 ezagutu ditu analizatzaileak. Oinarrian aditza ezagutu du beste 108 kasutan eta adjektiboa 29 kasutan. Kontua da, analizatzailea saiatzen dela ahal duen guztia ezagutzen, eta inurritegi moduko adibide batean, oinarrian inurritu aditza eta inurri izena dagoela esango du.

inurritegi	inurritu	KAT_ADI	tegi
inurritegi	inurri	KAT_IZE	tegi

Eta adjektiboekin kontu bera gertatzen da. Badira oinarri batzuk izen eta adjektibo kategoriatakoak izan daitezkeenak, eta horiek biak emango ditu analizatzaileak.

ongarritegi	ongarri	KAT_ADJ	tegi
ongarritegi	ongarri	KAT_IZE	tegi

Hala ere, guztiak ez dira modu honetakoak, eta badira oinarrian izenik ez dutenak. Halako kasua da *itxitegi*. Honek oinarrian ez du izenik, baina analizatzaileak oinarrian aditza duelakoan detektatzen du. Definizioan agertzen denez, hitz honek *toki itxia* adieraziko du.

itxitegi	itxi	KAT_ADI	tegi
----------	------	---------	------

Beraz, analizatzaileak ez du beti oinarrian izen bat bilatuko, eta atzizki bakoitzarentzat oinarriak nolakoak diren aztertu ondoren, egoki izan daiteke atzizki horrek onartzen ez dituen oinarriak baztertzea —eta agian *itxitegi* moduko kasu bereziak, alegia, dagokion oinarria ez dutenak zerrendatzea—.

V.9.3. Sinonimiaren hedadura

Dagoeneko aipatu dugun bezala, *sinonimiaren hedadura* erabiltzeak hitzen arteko erlazio gehiago ezagutzea dakar. Gainera, interpretazioak *ekintzak* —edo hauen *ondorioak*— edo *gertaerak* direnean, normalean sinonimoen bitartez definitzen dira, eta sinonimo horiek edo aditz-izenak edo *ekintza* adieraziko duten hitz eratorriak izan ohi dira.

jaiotza	Al	iz.	Jaiotzea , mundura etortzea .
---------	----	-----	---

Kasu honetan, ez *sinonimoa* bakarrik, baizik eta *hiperonimoa* ere aditz-izen bat da. Halakoetan, aurretik erauzitako informazioa oso garrantzitsua da; *sinonimo* hori eta *hiperonimo* hori aurretik markatuta ditugu, eta halako definizioei orain marka berririk jarriko ez diegunez, horiek zerrendatzea garrantzitsua izan daiteke, eta berriro ere analizatzaile morfologikoa aplikatu horien interpretazioak zehaztu ahal izateko. Demagun *baliokidetza* hitza dugula. Analizatzaileak hiru kategoriatako oinarriak izan ditzakeela zehaztuko du:

baliokidetza	baliokidetu	KAT_ADI	tza
baliokidetza	baliokide	KAT_ADJ	tza
baliokidetza	baliokide	KAT_IZE	tza

Beraz, V.6. ataleko adibidean bezala, *-tzak* dituen 5 interpretazioak dira posible oraingoz. Definizioa jo, eta hau topatu dugu:

baliokidetza A0 iz. Baliokidetasuna.

Kasu honetan, definizio-eremuan ez dugu deskribapen bat, baizik eta *sinonimo* bat. Lehendik egina dugun lanari esker, badakigu baliokidetasun hori baliokidetzaren sinonimoa dela, eta beraz, pentsatu behar dugu interpretazio bera izango dutela. Orain interpretazioak ezagutzeko egindako erregelek markatu ez dituzten definizio-eremuetatik dauden sinonimoak ateratzen baditugu, eta horiei analizatzaile morfologikoa berriro aplikatu, jakin ahalko dugu sinonimo horiek eratorriak diren edo ez. Eratorriak direnetan, atzizki horren interpretazioak zeintzuk diren jakin ahalko dugu, eta konparatu ahalko ditugu aurretik ditugun interpretazioak eta orain sortu zaizkigun berriak. Kasu honetan *-tasundun* hitz eratorria dugu, eta *-tasunen* interpretazioak hauek dira:

atzizkia	interpretazioa	oinarria	eratorria	adibidea	definizio-patroia
tasun	ADJ izatearen egoera	adj	ize	drogazaletasun	-en egoera, adj. dagoenaren egoera
tasun	ADJ izatearen nolakotasuna	adj	ize	hauskortasun	adj. denaren nolakotasuna

82. taula. *-tasun* atzizki lexikalaren interpretazioak.

Beraz, bi horietako bat izango da baliokidetza hitzaren interpretazioa: *egoera* edo *nolakotasuna*. Hala, balizko bost interpretazioetatik hauen ebakidura erabilita, zenbakia bira murriztu zaigu: *egoera* eta *nolakotasuna*. Baina, honen ondoren, *baliokidetasun* horren definizioa aztertu dezakegu:

baliokidetasun A0 iz. Baliokide**aren nolakotasuna**.

Eta definizioan nolakotasun forma agertzen denez, jakingo dugu baliokidetasuna izango dela baliokide izatearen nolakotasuna; eta hedaduraz, baliokidetza izango dela baliokide izatearen nolakotasuna.

V.9.4. Erlazioak EuskalWordNet-era lotzea

B. eranskinean agertu ditugun erlazioak oraingoz ez daude *EuskalWordNet*era lotuta, eta gainera, dagoeneko ikusi dugunez, hiru elementuren arteko erlazioak lotzea —urregile-urre-landu— ez da erraza izango oraingoz *EuskalWordNet*ek ez duelako aukera hori.

Lan hau *Meaning* —IST-2001-34460— proiektu europarraren ataza da, eta euskarazko hitzen artean lortutako erlazio lexiko-semanticokoak beste hizkuntzetarako baliagarri direnez, *EuskalWordNetera* lotzea garrantzitsua da. Hau etorkizunerako lan gisa utzi dugu.

V.10. Ondorioak eta etorkizuneko lana

Kapitulu honek hitz eratorri eta haien oinarrien artean dauden erlazio lexiko-semanticokoak erauzteko metodoa azaltzen du.

Lehenengo aportazioa izan da analizatzaile morfologikoari berez dituen baino atzizki lexikal gehiago txertatzea. Honek ahalbidetu du ondorengo lanetarako hitz eratorri gehiago ezagutzea. Tresna hau ezagutzarako oso garrantzitsua da, nahiz eta, sorkuntzan erabili nahiko bagenu, gainsorkuntza handia sortuko lukeen. Atzizki lexikalak sorkuntza egiteko erabili nahiko bagenu, oinarriaren ezaugarri zehatzagoak beharko genituzke, hala nola, *biziduntasuna*, *konkretutasuna*, *aditz-mota*, eta abar.

Hitz eratorri bat ezagutu ondoren, eratorri horren adierak zeintzuk diren zehaztu ditugu. Horretarako atzizki horri buruz egin izan diren lanak erabili ditugu, eta, hala, bere adierak zehaztu (ikus B. eranskina). Adieraren arabera, eratorriaren eta oinarriaren arteko erlazioa ezberdina da. Guk horien arteko lotura zein den zehaztu dugu. Baina lotura hori, ezin dugu zuzenean lortu atzizki jakin batekin sortutako eratorri batek interpretazio bat baino gehiago baldin badu. Interpretazio hori zehazteko, hiztegian duen definizio-patroia aztertu dugu, eta horren arabera erabaki dugu interpretazio bat edo beste duen. Dagoeneko aipatuta dugun bezala, desanbiguazioa LNPre arlo osoa da, eta gaur egun ataza aski zaila. Gure kasuan, desanbiguazioa definizioei esker egiteko gai gara.

V.7.1. atalean aipatu dugun bezala, aztertu ditugun atzizkiak dituzten 7.573 sarrera ezagutu ditugu. Eratorria eta erlazioaren arteko erlazioa 3.441 kasutan markatu dugu —%45,44—, eta sarreraren sinonimoa 2.987 kasutan —%39,44— ezagutu dugu.

Eratorriaren eta oinarriaren artean dagoen erlazioa desanbiguatzeke garaian lortu dugun doitasuna neurtzeko, anbiguoak diren hiru atzizki lexikalez osatutako

laginak eskuz etiketatu ditugu, eta etiketa horiek makinak emandakoekin konparatu. Hiru atzizkietatik bitan, bi hizkuntzalariren arteko adostasuna makinaren eta hizkuntzalari baten artekoa baino baxuagoa izan da, eta bestean, altuagoa. Honek adierazten du makinak markatutakoak pertsonak markatzen duenaren adinako doitasuna duela; atzizkien arabera, %100etik %80 ingurura. Kontuan izan behar da makinak ez dituela denak markatzen; dudazko kasuetan ez du adierarik ematen, eta agian dudazko horiek dira hizkuntzalarien artean ezadostasunak sortzen dituztenak. Hizkuntzalariek kasu ia guztiak etiketatu dituzte, eta, beraz, estaldura ia osoa dute; makinak lortzen duen batez besteko estaldura, aldiz, %82 da.

Aurretik esan dugun bezala, konposizioa osatzen duten hitzak ere aztertzen saiatu gara, hitz konposatua eta oinarriaren arteko erlazioa lortuko dugulakoan. Mugan dauden zenbait atzizkirekin oso emaitza onak lortu ditugu —*-dun* eta *-bera*, kasu—, baina beste zenbaitzurekin —*-aldi* eta *-bide*— ez dugu lortu eratorria eta oinarriaren arteko erlazioa zehaztea. Eratorriaren esanahia zein den desanbiguatzen dugu, baina ezin dugu oinarriarekiko erlazioa zehaztu.

Lan hau egin ondoren, saiatu gara oinarriaren eta eratorriaren adierak desanbiguatzen, eta hala, lortutako informazioa *EuskalWordNeten* txerta dezakegu. *EuskalWordNeten* kontzeptuen arteko loturak ditugu, *synseten* —sinonimo multzoen— arabera sailkatuta. Behin hau eginez gero, kontzeptuen arteko erlazio lexiko-semantikoak aberasten ditugu, ez bakarrik euskararentzat, baizik eta *EuroWordNeten* dauden hizkuntza guztientzat. Euskarazko bi kontzepturen arteko lotura lortzen badugu, eta kontzeptu horiek ingelesezko eta espainierazko hitzekin lotuta baldin badaude, beste hizkuntza horietako hitz horien artean erlazio bera dagoela pentsatuko dugu.

Honek dakartzan hobekuntzak ikusita, baieztatu dezakegu lan hau beste hizkuntzekin egingo balitz, eratorpenarekin lotuta dauden hainbat atazek estaldura handiagoa lortuko luketela.

Eratorrien eta oinarrien adiera desanbiguatzeko, hiztegiko eratorriak eta hauen oinarriak *EuskalWordNeten* ote dauden begiratu dugu. Hauen estaldura ez da oso altua: eratorrien kasuan %22, oinarria %9 eta biak %16. Kontuan izan behar dugu, lotura hau egiteko garaian *EuskalWordNet* sortzen ari ginela. Loturak egin ondoren, kontzeptuen ebakidura erabiliz gauzatutako desanbiguazioaren doitasunari

dagokionez, berriz, oso datu onak ditugu. Eratorrien kasuan %97 eta oinarrien kasuan %80 da erabat zuzena, eta %15ean kontzeptu zuzena lotuta dago, baina baita zuzena ez den beste bat ere. Eratorrietan emaitza hobeak lortzen dira adiera bakarreko eratorriak gehiago direlako: %52,6; oinarrien kasuan, berriz, adiera bakarrekoak %7,7 dira.

Etorkizuneko lanen artean, egindako ikerketari esker badakigu sinonimiaren hedadura aplikatzeak erauzitako erlazioen gorakada ekarriko duela. Ikusi dugunez, —*tza* atzizkidun hitzen kasuan (V.6 atala), sinonimoen bidez definitutako 96 definiziotatik, 52 kasutan lortzen dugu sinonimo horien informazioa hasierako sarrera-hitzari hedatzea. Beraz, sinonimiaren hedadura aplikatu gabe baino %54,17 interpretazio gehiago lortu ditugu.

Etorkizuneko beste lana erlazio konplexuen deskribapena fintzea da, eta tarteko lana egiten duen elementua definiziotik erauzi ahal dugunean errekuaratzea. Zenbaitetan erlazioak ez dira bi elementuren artekoak, eta definizioetan agertzen den aditzak indar handia du erlazioan bertan. Eratorria eta oinarria izenak direnean, eta erlazioa *berezko objektua* dela esaten dugunean, tartean aditza dagoela suposatu behar dugu, eta aditz hori zein den definiziotik erauzi beharko dugu —ahal den kasuetan, behintzat—. Erlazio hauek oraingoz ezin dira *EuskalWordNet*en txertatu, baina *Meaning* proiektuari esker lan hau gauzatu nahi da.

Aurrera begira ikasketa-teknikak erabili ahalko ditugu atzizkien interpretazioak desanbiguatu ahal izateko. Hiztegitik eta atzizkiekin ikasitakoa corpusetara hedatuko dugu, baliagarri izan baitaiteke hiztegian ez dauden hitz eratorriak ezagutu eta desanbiguatu ahal izateko.

Ondorengo kapituluan sarrera-hitza alde batera utzi, eta *differentiaren* azterketari ekingo diogu. Definizioan bertan dauden hitzen arteko erlazio lexiko-semanticokoak zeintzuk diren zehazten saiatuko gara.

Esker mila, irakasle, ikaskide eta lankideok

herri	A1	iz.	Lagunartean eta lurralde jakin batean bizi diren, eta elkarrekin hainbat ohitura eta erakunde dituzten gizakien multzoa.
hizkuntza	A1	iz.	Giza talde batek elkarrekin komunikatzeko duen hitzezko adierazpidea.

Iragan egun batez, Ostatu batean,
bi lagun ari ziren ez ta bai betean;
biak ziren euskaldun zintzoak ustean,
halere ezin adi elgarren artean;
entzuten egona naiz umore tristean.

Ez dut osorik hartu baten parabola,
erderaz mintzo baitzen, berak jakin nola:
gure herri maiteaz zuela axola,
hau, bere gain beharra, bertzen men dagola,
“Gora Euskal-Herria”, frantsesez ziola.

Anai-arrebak entzun ene ahotsa:
izaite bat ez daike hezur utsez osa;
herria da gorputza, hizkuntza bihotza;
bertzetik bereixtean bitarik bakotxa,
izaite horrendako segurra hilotza.

Batzu herriaz orroit, euskaraz ahanzi
bertzek euskara maite, herria gaitzetsi;
hikuntza ta herria berex ez doatzi,
berek nahi daukute konpreniarazi
bata bertzea gabe ez daizkela bizi.

(Xalbador)

VI. Patroi lexikalak eta postposizioak

VI.1. Sarrera

Tesi-lan honen ildoari jarraiki, sarrera-hitza eta definizioko hitz baten arteko erlazioa erauzi dugu; baita sarrera-hitza eratorria denetan eratorriaren eta oinarriaren arteko erlazioa ere. Kapitulu honetan, definizioan agertzen diren hitzen arteko erlazioa(k) markatuko d(it)ugu; zehazki, dagoeneko aipatu dugun *differentia* aztertuko dugu.

ildo A1 iz. Goldeaz lurra irauliz egiten den
 irekidura luzea.

ildoren definizio horretatik orain arte genusa ezagutu dugu: irekidura. Eta orain *differentia* aztertuko dugu: luzea, eta goldeaz lurra irauliz egiten dena. Erlazioak definizio barruko hitzen artekoak dira: irekidura-luze, irekidura-egin, egin-irauli, irauli-lurra, irauli-golde.

Erlazio horiek erauzteko, bi modu erabiliko ditugu: bata definizio barruko *patroi lexikalen* bitartez, eta bestea *definizio osoa aztertuz*. Bi hurbilpen hauek ez dira inondik inora maila berekoak, baina tesi-lan honetan daramagun eskemaren arabera, eta definizio barruko hitzen arteko erlazioak zehaztuko dituztenez, kapitulu berean emango ditugu.

Definizioan agertzen diren patroi lexikalak elementu zurrun gisa hartuko ditugu. Patroi horiek *token* segida jakinak izango dira (ikus III.8. atala), eta beti ordena berean agertuko dira. Halakoak aztertu dituzte bai Vanderwende-k (1995), eta bai Harabagiu-k (1997). Hala ere, azken honek garrantzi gehixeago eman die

preposizioei. Honi jarraituta, patroi lexikalak aztertu ondoren, definizioan ditugun postposizioei erreparatuko diegu *differentiaren* ezaugarriak lortzeko asmoz.

Postposizio baten bitartez definizioko zein hitz lotzen ditugun jakiteko, definizioaren azaleko sintaxi osoa ezagutu behar dugu. Sintaxi hau dagoeneko ezagututa izanda, loturak sintagmen artekoak izango lirateke, eta erlazio nagusiak sintagmen buruen artekoak, hain zuzen ere. Beraz, azaleko sintaxiak izugarri lagunduko luke, eta are gehiago buruen arteko dependentziak zeintzuk diren markatuta izango bagenitu. Adibidez, ildoren adibidean: *egin-irauli*, *irauli-golde*, *irauli-lur*.

Baina, helburua hori bada ere, oraingoz Ixa Taldean garapen prozesuan dauden dependentziak (Aduriz *et al.*, 2004) ez ditugu aplikatu, baina kapituluan ikusiko dugun teknikari esker, ondoz ondoko hitzak kontuan hartuz, postposizioa daraman hitzari interpretazioaren bat esleitzeko gai gara.

Beraz, eta sintagmen arteko erlazioa zein den jakin nahi dugunez, sintagmak lotzen dituzten elementuen balizko interpretazioei erreparatzea erabaki dugu. Horiek izango ditugu kasu-atzizki gramatikalak eta postposizioak. Kasu-atzizki gramatikalak aditz laguntzailean isla dutenak dira. Gurean *absolutiboa*, *ergatiboa* eta *datiboa* dira. Postposizioei dagokienez, bi motakoak ditugu: postposizio-atzizkiak —gramatika klasikoetan deklinabide-atzizki deitu izan direnak—, adibidez *-z* instrumentala; eta postposizio konposatuak, adibidez, *-i buruz*.⁸⁸

Lan honen helburua ez da elementu bat noiz den postposizio eta noiz ez ebaztea, baizik eta, beste batzuek egindako lanak erabiliz, postposizio-zerrenda bat lortu eta horiei balizko interpretazio batzuk ematea.

Era berean, ez da gure helburua zehaztea eta teorizatzea ea ematen dugun interpretazioen zerrenda *rol tematikoak* diren edo *interpretazio semantikoak* diren. Oinarri hartuko ditugun lanetan, alde batetik *interpretazio semantikoak* ditugu (Aldezabal, 2004), eta, bestetik, *egitura lexiko-kontzeptualak*, hauek dituzten elementu multzo guztiekin, hala nola, *primitiboak*, *eremuak*, *motak* eta *rolak* (Dorr,

⁸⁸ Postposizio-atzizkietan *genitiboak* ditugu, euskararen kasuan *edutezkoa* eta *lekuzkoa*. Hauek sortzen dituzten erlazioak, normalean, sintagmaren barruko erlazioak dira, eta gure helburua sintagmen arteko erlazioak zedarritzea denez, hauek ez ditugu kontuan izango.

1993). Gure helburua horiek guztiek barne hartzen dituzten *interpretazio multzo* bat sortzea da, hizkuntzen arteko lotura gauzatu ahal izateko.

Guk lana hiztegiko definizioetan dauden hitzen arteko erlazioak ateratzeko erabiliko dugu, baina behin postposizio horien interpretazioa lortuz gero, bada aukerarik lortutako jakintza testu errealetan aplikatzeko (Agirre *et al.*, 2004). Honexegatik, hain zuzen ere, ez gara hiztegian agertzen diren postposizioen interpretazioak ematera mugatuko, baizik eta ahal den postposizio kopuru handiena ezagutzen saiatuko gara. Postposizioen adibideak emateko garaian, ez gara EHn agertzen direnei mugatu, eta corpusetan topatutakoak ere testura ekarri ditugu.

Beste hizkuntzetako preposizioen interpretazioak izanez gero, hizkuntzartekotasuna garatu ahal izango dugu. Euskaraz postposizioak ditugun leku berean, espainieraz eta ingelesez preposizioak ditugu —eta zenbait argumenturen kasuetan, preposiziorik-eza, hala nola aditzen subjektu eta objektuetan—. Azterketa honetan ez diogu erreparatuko elementu baten argumentu-tasunari edo modifikatzaile-tasunari, eta interpretazioak agertzen den postposizioaren edo preposizioaren arabera esleituko dira.

Gure helburuetako bat preposizio edo postposizioentzat emandako interpretazioak bateragarriak izatea da, eta, hala, hiztegien definizioak erabiliz, preposizio edo postposizio horien anbiguotasunari aurre egiteko teknika azalduko dugu. EHko sarrera-hitza beste hizkuntzetara itzuliko dugu hiztegi elebidunak erabiliz, eta ondoren, lortutako itzulpen horiek ea ingeleseko eta espainierako hiztegietan sarrera-hitz gisa dauden bilatuko dugu. Behin horiek topatuta, hizkuntza ezberdinetako definizioek hitz bertsuak dituzten konparatuko dugu, eta hala testu paralelo lerrokatuak sortuko ditugu. Honela, kontzeptu berari buruzko definizioak direnez, eta definizioko hitzak bertsuak direnez, agertzen diren preposizio eta postposizioen interpretazioak ere bertsuak direla pentsatuko dugu. Honekin, hizkuntzartekotasuna erabiliz, postposizioen interpretazioak desanbiguatzeke gai izango gara.

Euskarazko postposizio-atzizki eta kasu gramatikalentzat, Aldezabal-ek (2004) bere lanean ehun aditzen argumentu-egitura aztertzen duenean, hainbat interpretazio semantiko aipatzen ditu. Hori izango da gure oinarrietako bat. Eta beste hizkuntzetako —espainiera eta ingelesa— preposizioen deskribapenak kontuan

hartuta, guk aukeratu duguna Dorr-ena (1993) da, ingeleseko eta espainierako preposizio multzo handi baten deskribapena eginda duelako.

Beraz, guk sortuko dugun interpretazio-zerrendak bien —Aldezabal eta Dorr— arteko lotura izan beharko du, hartan guztien informazioa erabili ahal izango dugulako, eta euskara, ingelesa eta espainiera lotzea lortuko dugulako.

Interpretazio bat zenbat postposizioek edo preposizioek duten zehazterakoan, sortzen diren ebakidurak oso handiak dira. Adibide gisa, esaten dugunean *inesiboak leku* interpretazioa duela, eta ingelesez, *in the middle of*ek *leku* interpretazioa duela, bi elementuak duten ebakiduraren arabera lotzen ditugunean, *inesiboa* eta *in the middle of* lotuko ditugu. Hau hala agertzeak, nahikoa arraroa dirudi, baina lehenengo hurbilpena denez, sinpletasunaren alde egin dugu. Hala ere, postposizio askok hiztegietan dituzten itzulpen-preposizio bere-bereak zeintzuk diren emango dugu F., G. eta H. eranskinetan. Adibidez, *in the middle of*en kasuan, bere-bere gisa agertzen dira: *-0⁸⁹ erdian* eta *-en erdian*.

Baina, gure lehen helburua ez da itzulpen automatikoa. Guk testu paraleloekin lan egiten dugula esan dezakegu; ditugun euskarazko definizioen paraleloak lortu behar ditugu bai ingelesez eta bai espainieraz. Beraz, guk ez dugu sorkuntza erabiltzen, baizik eta, dauden postposizioen edo preposizioen interpretazioen ebakidurak erabiliko ditugu agertzen direnen interpretazioak desanbiguatzeko.

Eta aurreko bera hiztegietan egin beharrean corpus paraleloetan egiten badugu, interesgarria izan daiteke postposizio eta preposizio guztien interpretazio-ebakidurak izatea. Hona testu paralelo honekin nola lan egingo genukeen.

Plazan elkartu ginen.

We met in the middle of the square.

Halako itzulpen libreak corpusetan sarri agertzea ez da hain arraroa izango, eta *plaza* / *square* horren interpretazioa zein den jakin ahal izateko, *inesiboaren* (-n) eta *in the middle of*en interpretazioak zeintzuk diren jakin beharko dugu, eta haien arteko ebakidura zein den erauzi. Kasu honetan, biek izango dute *leku* interpretazioa postposizio-preposizio guztien interpretazioak lotuta ditugulako; baina, postposizioen lotura zuzenak besterik ez bagenitu izango, ezin izango genituzke *inesiboa* eta *in the*

⁸⁹ Aurreko elementuak postposiziorik ez duela hartzen adierazten du.

middle of lotu, *inesiboaren* itzulpen zuzen-zuzena ez delako inoiz *in the middle of* izango. Eta halakoak corpus errealean gertatu, gertatzen dira.

Hauex da, bada, kapitulu honen hezurdura: VI.2. atala EHko patroi lexikalei eskaini diegu; VI.3. atalean lehenik erabili ditugun iturriak zeintzuk diren aipatu dugu; VI.4. atalean postposizio bakoitzari interpretazioa emateko metodologia zein izan den azaldu dugu, eta adibide batekin garatu VI.4.1. atalean; VI.5.ean postposizio-atzizki eta kasu gramatikal bakoitzaren interpretazioak emango ditugu. Hauek banan-banan eman ditugu anbiguotasun handia dutelako. Baina, postposizio konposatuen kasuan (VI.6. atala), berriz, anbiguotasuna asko gutxitzen denez, lehenengo anbiguen zerrenda eman dugu, eta, ondoren, interpretazio bakoitzaren azpian hari dagozkion postposizioen zerrenda. VI.7. atalean postposizio-atzizki instrumentalarekin egindako esperimentua erakutsi dugu. Bertan postposizio-atzizki honen interpretazioak aztertu ditugu. Ingeleseko eta espainierako hiztegietatik pareak atera ditugu eta ikusi dugu nola funtzionatzen duen gure metodoak eta nolako emaitzak lortzen ditugun, beste hizkuntzak erabiliz, instrumentalaren interpretazioa erabakitzeke garaian. Eta VI.8. atalean ondorioak aipatu ditugu.

VI.2. Patroi lexikalak EHn

Vanderwende-ren (1995) lanari jarraiki, EHn agertzen diren patroi lexikalen bilaketari ekin diogu. Patroi hauek oso zehatzak izan ohi dira —*tzeko erabiltzen den*—, eta hiztegian halako gutxi agertzen dira. Atal honetan horien zerrenda eman dugu eta C. eranskinean bakoitzari buruzko ezaugarrien berri eman dugu.

Patroi hauek erauzteko garaian, III.8. atalean aipatutako *token*-segidak izan ditugu kontuan. Han usuen agertutako segidak aztertu ditugu, eta, dagoeneko erlatokeen atalean (IV.4. atala) ikusi dugun bezala, batzuk erlatoretzat hartu ditugu, baina beste batzuk patroitza hartuko ditugu. Erlatore eta patroiaren arteko diferentzia oso txikia da. Erlatokeen kasuan ikusi dugunez, definizioaren gune sintaktikoaren posizioan ez da esanahi garrantzitsua duen hitzik agertzen —askotan elipsia agertzen da, bestetan adiera oso ahuleko hitzak—, eta erlatore honek erlazioan jartzen ditu sarrera-hitza eta definizioan agertzen den beste hitzen bat. Patroiaren kasuan, aldiz, genusak azaldu ditugunean ikusitako definizio motak agertzen dira. Genusaren

informazioa izateaz gain, definizioan agertzen den segida oso jakin bati esker, genus hori eta definizioko beste hitzen bat jartzen da harremanetan. Hona adibide bat:

galbahe A0 iz. Aleak, eta hondarra, ikatza eta abar
bahetzeko erabiltzen den **bahe** xehea.

Aurreko adibidean, lehendik egindako azterketari esker, badakigu galbaheren genera *bahe* dela, eta -t(z)eko erabiltzen den patroiarri esker *bahe* horren *berezko ekintza-helburua* bahetu dela jakingo dugu. Kasu honetan *patroi lexikal* bat dugula esango dugu.⁹⁰

Beraz, dagoeneko ikusi dugunez, erlatoareak eta patroii lexikalak *token*-segiden azterketari esker lortuko ditugu, eta diferentzia nagusia izango da erlatoaren kasuan lortzen dugun erlazioa sarrera-hitzaren eta definizioko hitz baten artekoa dela, eta patroien bitartez, berriz, definizioan dauden bi hitzen arteko erlazioa ezagutuko dugula.

Patroia beti *jakina* da, eta beti ditu elementu lexikal berak eta orden berean, gainera. Hiztegia 718 patroii markatu ditugu.

patroia	interpretazioa	zenbat aldiz hiztegia
-z mintzatuz ⁹¹	erabileraren eremu semantikoa	264
-t(z)eko erabiltzen den	berezko ekintza helburua	140
egiten duen	berezko objektua	161
ematen duen	berezko objektua	25
-en familiako	meronimia	75
-i dagokion	berezko egile ⁹²	53

83. taula. Patroiak eta erlazioak.

Berezi moduan hartuko ditugu *ez*, *bereziki* eta *batez ere* partikulak. Hauek kasu guztietan markatu ditugu, bestelako markarik har ez dezaten, arazo handiak ekartzen dituztelako.

Emaitzei dagokienez, IV. kapitulan emaitzak ateratzeko (ikus IV.7. atala) erabilitako lagin bera aztertu dugu. Izenen laginean 2 patroii daude, eta 2ak ondo topatu ditugu: *-en familiako* eta *-ak egiten dituen*. Adjektiboen laginean 6 harrapatu ditugu eta beste 2k egin digute ihes —*-z mintzatuz*—. Aditzen laginean 1 dugu —*-z*

⁹⁰ Erlatoarekin sortutako definizioa izango bagenu, bahetzeko erabiltzen dena —elipsidun forma— izango genuke.

⁹¹ Honek ez ditu erlazioan jartzen definizioan dauden bi hitz, baizik eta sarrera-hitzaren erabileraren eremu semantikoa zein den zehazten du. Ez da beste patroiiak bezalako, baina informazio interesgarria ematen du eta multzo honetan sartu dugu.

⁹² Hala ere, patroii hau dudazkoa izan daiteke, beti ez baitu erlazio bera adierazten.

mintzatuz— eta ez dugu harrapatzen. Izen eta adjektiboen laginean 6 harrapatzen ditugu —*-z mintzatuz*—. Azpisarreraren laginean ez da patroirik agertzen.

	definizioak denera	patroiak	ondo	gaizki	harrapatu gabe
izenak	108	2	2	0	0
adjektiboak	108	6	6	0	2
aditzak	100	1	0	0	1
izen / adjektibo	124	6	6	0	0
azpisarrerak	92	0	0	0	0

84. taula. Patroiak laginetan.

Beraz, badirudi *differentia* ezagutzeko garaian patroiek ez digutela gehiegi laguntzen oso gutxi topatu ditugulako, eta ondoren postposizioekin egindako lana azalduko dugu.

VI.3. Postposizioen interpretazioak zehazteko oinarriak

Postposizioek definizioan dauden hitzen arteko loturak gauzatzen dituzte, eta horien interpretazioak ezagutuz gero, lotzen dituzten hitzen arteko erlazioa zein den ezagutu ahalko dugu. Espainiera eta ingelesa bezalako hizkuntzetan, postposizioek betetzen duten leku berean preposizioak agertzen dira, eta, gure helburua euskarazko, espainierazko eta ingelesezko postposizio eta preposizioei interpretazio-zerrenda bera esleitzea da.

Esan dugun bezala, postposizioen interpretazioak zehazteko, bi oinarri nagusi erabili ditugu. Alde batetik zenbait euskarazko postposizio-atzizkirentzat eta kasu gramatikalentzat hainbat interpretazio semantiko ematen dituen Aldezabalen (2004) lana, eta ingelesezko eta espainierazko preposizioen egitura lexiko-kontzeptualak sortu dituzten *Marylandeko Unibertsitateko CLIP taldearen* lana (Dorr, 1993; Dorr eta Habash, 2002).

Aldezabal-ek (2004) 100 euskal aditzen azterketa sakona egin du. Bertan zehazten du aditz multzo jakin batzuen argumentu-egitura eta baita zenbait elementuren interpretazio semantikoa ere.⁹³ Guk bere lanean agertzen diren postposizio-atzizkien agerpenak hartu ditugu, eta berak aipatutako interpretazio semantikoekin lotu. Hala, postposizio-atzizki bakoitzarentzat interpretazio-zerrenda

⁹³ Aldezabalen datu-basea <http://ixa2.si.ehu.es/nuke/index.php> (2004-11-30ean atzitua) web-helbidean ikus daiteke.

jakina lortu dugu. Hau da gure lanaren iturri nagusia. Gure helburu nagusia euskara da, eta garrantzitsua iruditu zaigu iturri gisa hizkuntza bera oinarri duen lana hartzea.

Adibidez, *laztandu* aditzarentzat aipatzen ditu kasu azpikategorizatu gisa *ergatiboa* eta *absolutiboa*; kasu onartugabe gisa: *-ela konpletiboa* eta *adlatiboa*; eta onartzen dituen kasuen artean: *inesiboa*, *instrumentala*, *soziatiboa*, *ablatiboa* eta *datiboa*. Eta agertzen den kasu bakoitzarentzat interpretazio semantiko hauek ematen ditu:

ERG (ergatiboa) entitatea beti gizakia da, eta ekintza burutzen du. ABS (absolutibo) entitatea, izaki biziduna da askotan, edo bestela izaki bizidun baten gorputz atala; (...)

INE (inesibo) zilegi dela esan dugu, baina ez dugu adibiderik aurkitu. Hala ere, ekintza non edo nola gertatzen den adieraziko luke.

SOZek (soziatibo) eta INSek (instrumental) modua adierazteko gaitasuna dute, nahiz eta adibide bakarra aurkitu dugun; zehazki SOZ duena, eta horretan baliabidea adierazten du (...) SOZek konpainia ezin dezakeela adierazi ematen du

ABLEk (ablatibo), adibiderik agertu ez bazaigu ere, ERG (ergatibo) entitatearen kokalekuari erreferentzia egin diezaioke (*aulkitik laztandu*), edo ABS (absolutibo) entitateari dagokion atalen bati (*gainetik...*). Nolanahi ere ez da batere arrunta.

Azkenik, DATi (datibo) dagokionez, esan behar da hau azaltzen denean, ABS (absolutibo) entitatearekiko jabetzazko harreman baten berri ematen duela beti. Jabetza hau berariazkoa da ABS entitatea gorputz-atala denean. Gorputz-atala ez denean, ABS entitatea DAT entitatearen jabea da besterik gabe, hau da, jabetza besterengarria da. Beraz, horrelakoetan, ekintza zuzentzen zaion entitatea baino gehiago, gure ustez jabetzazko harreman hori da azpimarratu nahi dena. (...) (Aldezabal, 2004: C eranskina: 30. or.)

Aditz honen balio sintaktiko/semantikoak aipatzen dituenean, kasu azpikategorizatuei *esperimentatzaile* eta *gaia* balio semantikoak esleitzen dizkie, jardueratzat jotzen duelako aditz honek adierazten duen predikatu mota.

laztandu-DU-0: *esperimentatzailea* [+giz]_ERG; *gaia*_ABS
(Aldezabal, 2004: C eranskina: 30. or.)

Beraz, *laztandurentzat* emandako informaziotik erauzi ahal izango ditugu zenbait kasurentzat hainbat balizko interpretazio, hala nola: *ergatiboa*, *esperimentatzailea*; *absolutiboa*, *gaia*; *inesiboa*, *lekua* eta *modua*; *soziatiboa*, *modua*, *baliabidea* eta *konpainia*; *instrumentala*, *modua* eta *baliabidea*; *ablatiboari* eta *datiboari* buruz ezin

izango dugu interpretaziorik atera aipatzen dituenak perpausean dauden beste izen-sintagma batzuekiko erlazioak direlako eta ez ekintzarekiko erlazioak.

Aldezabalen lanean agertzen diren aditz guztien informazioa modu honetan aztertu ondoren guri baliagarri zaigun postposizio-atzizki eta interpretazioen arteko lotura dugu.

Beste hizkuntzetako preposizioen interpretazioak aztertu dituzten lanei dagokienez, ingelesezko eta espainierazko preposizioen nahiko deskribapen zehatza egin dute Marylandeko Unibertsitateko *CLIP* laboratorian. Guk, esan bezala, Dorr eta Habashen (Dorr, 1993; Dorr eta Habash, 2002) lanak hartu ditugu iturri gisa.

Hauek ingelesezko eta espainierazko preposizioen interpretazioak emateko egitura lexiko-kontzeptualak —*Lexical Conceptual Structure*, LCS— erabiltzen dituzte. Deskribapen mota hau Jackendoff-engandik jaso da (Jackendoff, 1983; 1990) eta aditzen jokaera sintaktiko-semantikoa deskribatzeko erabili izan da. Dorr-ek bere tesi-lanean (1993) ingelesezko aditzen egitura lexiko-kontzeptualak (ELK) eman zituen eta itzulpen automatikoari begira oso egitura erabilgarriak direla defendatu zuen. Lan-ildo honi jarraiki, preposizioen egitura lexiko-kontzeptualak sortu zituzten.

Egitura lexiko-kontzeptual hauek *mota*, *eremu*, *primitibo* eta *rolen* arteko lotura dira. Hona *with* preposizioak izan ditzakeen errepresentazioetako bat:

```
:DEF_WORD "with"
:COMMENT "with: 1. (a) He stabbed the burglar with a knife"
:LANGUAGE English
:LCS (WITH Instr (nil 27) (* Thing 20)))
```

Adibide honetan *with* preposizioaren egitura bat dugu, eta egitura horren adibide bat azaltzen da *comment* lerroan. Ondoren zein hizkuntzatako preposizioa den azaltzen da, eta azkenik egitura lexiko-kontzeptuala bera —LCS—: *primitibo* gisa *with* dugu eta *eremu* *instrumental* batean gertatzen da *ekintza* (27). *Ekintza* horrek rol tematiko *instrumentala* (20) du, eta *thing* motako elementua izan behar du. Hizkuntzen arteko lotura gauzatu ahal izateko, deskribapen horretan rol tematikoei erreparatzea pentsatu dugu, Aldezabalen *interpretazio semantikoek* ere hauen antza dutelakoan.

Mota honetakoa dugu erabiliko ditugun bi iturriek emango diguten informazioa, eta guk, bi horiek lotuko dituen interpretazio-zerrenda lortu nahi dugu, hurrengo pausoetan guztien lana erabilgarri izan dakigun.

Baina erabiliko dugun metodoa azaltzen hasi aurretik, badira lan hauen izaera dela-eta, kontuan hartu beharreko hainbat datu. Aldezabalen lanaren helburua aditzen argumentu-egiturak deskribatzea da, nahiz eta, zenbaitetan, aditzen modifikatzaileak ere aipatzen dituen. Beraz, kontuan izan behar dugu bere interpretazio-zerrendan zenbait modifikatzailearen interpretazioak kanpoan gera daitezkeela. Eta bestetik, Aldezabalek 9 kasu —postposizio— aztertu ditu. Gera daitezkeen balizko zulo hauek tapatzeko Dorr-en deskribapena begiratu dezakegu.

Behin postposizio-atzizki bat aukeratzen dugunean, kontuan izan behar dugu Aldezabalen lanean ez dugula zertan postposizio-atzizki horren interpretazio guztiak izan. Bere lanean 100 aditz horiekin atera zaizkion argumentu-egituretan eta modifikatzaile diren adibide nagusiak aztertu ditu, eta ez balizko egoera guztiak. Interpretazio hauek bilatzeko garaian euskarazko gramatikak eta hiztegiak baliatuko ditugu.

Gainera, kontuan izan behar dugu gerta daitekeela —eta zenbaitetan hala da— Aldezabalen lanean zenbait euskarazko postposizio-atzizki ez agertzea, 9 direlako aztertu dituen kasuak. Guk gramatikak baliatuko ditugu postposizio-atzizki guztiak agertzen diren edo ez zehazteko, eta postposizio konposatuekin —*i buruz* modukoak— ere lan egingo dugu.

Espainierazko eta ingelesezko preposizioei dagokienez, haien deskribapenak aztertu ditugu, eta bakoitzari dagozkion rolak bildu. Era berean, Habash-engandik (2002) jaso dugu ingelesezko preposizioen —gehi preposizio ezaren, \emptyset preposizioa esango dioguna— balizko rol tematikoen zerrenda.

Behin bi zerrendak izanda, idealena litzateke bi horiek zuzen-zuzenean lotu ahal izatea, alegia Dorr-en lanean ditugun *rol tematikoak* eta Aldezabalen lanean ditugun *interpretazio semantikoak* zuzenean lotzea, baina berehala jabetu gara hori erabat ezinezkoa zaigula. Bi zerrendak konparatzerakoan, zenbaitetan, Dorren interpretazioak Aldezabalenak baino zehatzagoak dira —hala gertatzen da *perc* (*perceived item*) kasuan—, baina, oro har, Aldezabalen interpretazioak Dorrenak

baino zehatzagoak dira.⁹⁴ Adibidez, Aldezabalen lanean topatuko ditugu *gai ukitua* eta *gai sortua*; *gai* mota ezberdinak, alegia.

Jabetza erlazioa ere berezia dela esan behar dugu. Badirudi erlazio hau bi izenen artekoa dela, eta ez izenaren eta aditzaren artekoa. Hau guztia argitzeko, adibide bat azalduko dut: *gizonak kamioia lastoz bete zuen*. Ekintza burutu ondoren *kamioia* —*edukitzailea*— izango da *lastoaren* —*edukiaren*— *jabe*, eta ezaugarri berria izango du *kamioi* horrek; baina aditzarekiko *lastoa gaia* izango da, eta *kamioia helburuko leku kokapena*. Dorren deskribapenean ere, biak aipatzen dira, eta *kamioia gaia* izango da —erlazioa eremu *identifikazional*ean gertatzen bada—, eta *helburuko leku kokapena* —erlazioa *leku* eremuan gertatzen bada—. Halako aditzek *alternantzia lokatiboa* dutela aipatzen dute, bi egiturak, alegia. Beraz, *jabetza* erlazioa albo batera utziko dugu lan honetan.

Aldezabalek *interpretazio semantiko* deitzen duenaren eta Dorrek *rol tematiko* deitzen duenaren artean bada beste diferentziarik: Aldezabalentzat *kausa* eta *bidea* *interpretazio semantikoak* dira, eta Dorren errepresentazioan halakoak *primitibo* eta *mota* gisa agertzen dira, eta *ez rol tematiko* gisa. Beste kontu berezia da *modu* interpretazioa, zeren nahiz eta Dorren rol tematikoen zerrendan agertzen den, gero ez zaio inongo preposiziori lotzen. Gainera, hauek euskaraz, askotan hitz beregainak izatera iritsi direnak ditugu, eta kontu izan beharko dugu.⁹⁵

Hauek guztiak kontuan izanez gero, zail izango da bataren eta bestearen lanak lotzea adibideei erreparatu gabe. Gainera, kontuan izan beharko dugu hizkuntza ezberdinentzat egindako deskribapenak direla, eta honek adibideak aztertzea eskatuko du. Beraz, bi lanak lotuko dituen tarteko interpretazio-zerrenda ematen saiatuko gara, bien hurbilpenak erabili ahal izateko. Hona erabiliko dugun metodoa.

VI.4. Interpretazioen zerrenda zehazteko metodoa

Atal honetan dagoeneko aipatu ditugun lanak lotuko dituen metodoa azalduko dugu. Esan dugunez, Aldezabalek eta Dorrek hurbilpen ezberdinak erabiltzen dituzte

⁹⁴ Hemen Dorren interpretazioak aipatzen ditugunean, bere egitura lexiko-kontzeptualaren rol tematikoari egiten diogu erreferentzia. Lehen aipatu dugun bezala, lotzen saiatzen garen elementua interpretazio horietako rola da.

⁹⁵ *buruz ikasi zuen* adibidea baldin badugu, *buruz* horrek izena gehi instrumental itxura duela pentsa dezakegu, baina, agian, dagoeneko adberbio gisa lexikalizatutako elementua izan daiteke (Pérez Gaztelu *et al.*, 2004).

postposizioen eta preposizioen interpretazioak emateko, eta, nahiz eta bata bestearengandik urrun daudela dirudien, bateragarriak direla ikusiko dugu.

Erabaki beharreko lehenengo kontua izango da zein den postposizio eta preposizio guztientzat erabiliko dugun interpretazio-zerrenda. Hau lortzeko, dagoeneko esan dugunez, bi iturri nagusi ditugu, baina baita hiztegiak eta gramatikak ere.

- Aldezabalek zenbait postposizio-atzizkirentzat proposatzen dituen interpretazio semantikoak (Aldezabal, 2004).
- Dorrek ingeles eta espainierazko preposizioentzat proposatutako egitura lexiko-kontzeptualak (Dorr, 1993).
- Hiztegi elebidunak: (Morris, 1998) eta (Elhuyar, 1996); gramatikak (Euskaltzaindia, 1985; 1987a; 1987b; 1991; 1992; 1993).

Aurreko atalean esan dugun bezala, gure hasierako helburua bi iturrien interpretazioak lotzea zen, baina, dagoeneko azaldu dugunez, ezinezkoa da zuzenez zuzenean lotzea eta, adibideei erreparatuta lotu behar izan ditugu.

Beraz, interpretazioen zerrenda zehazteko eta taula eleaniztuna eraikitzeko, euskararekin hasiko gara eta beste hizkuntzetara egingo dugu jauzi, eskuz etiketatutako adibideak beste hizkuntzetara itzuliz hiztegi elebidunen laguntzarekin. Lehenik Aldezabalen zenbait interpretazio —zehatzenak zirenak, alegia— biltzea erabaki dugu. Honen ondoren, euskarazko adibideetako postposizioak etiketatu ditugu, eta etiketa hori bera jarri zaio itzulita dagoen beste hizkuntzako adibideko preposizioari. Azken pauso gisa, etiketatutako adibideak Dorrengandik jasotako interpretazioekin konparatzen dira.

Hona pausoz pauso postposizio bakoitzarekin erabilitako metodoa:

- Postposizioa aukeratu.
- Postposizio honi dagozkion adibideak hiztegi elebidunetatik atera. Hala, postposizio honen itzulpen onak lortuko ditugu beste hizkuntzetan eta testuinguru bera dutenak, gainera.
- Postposizio honen balizko interpretazioak topatu Aldezabalen lanean.
- Postposizioaren interpretazioak aztertu, eta oso zehatzak direla iruditzen zaigunean biltzen saiatu.

- Markatu adibideak interpretazioekin eta zuloak ote dauden begiratu:
 - Aztertu ea baden adibideen artean Aldezabalen lanean agertzen ez den interpretazioen bat.
 - Adibideen artean gerta daiteke postposizio jakin baten interpretazio guztiak ez azaltzea, beraz hauen balizko itzulpenak ere lortu beharko ditugu.
- Behin puntu honetara iritsita, dagoeneko badugu euskarazko postposizio horri dagozkion interpretazioak eta interpretazio bakoitzarentzat adibideak euskaraz, espainieraz eta ingelesez
- Adibideetan agertzen diren ingelesezko eta espainierazko preposizioek orain arte erabilitako interpretazioa badute dagoeneko, eta hau Dorren interpretazioen zerrendatik lortutakoarekin konparatuko dugu
- Puntu honetan, halako laukoteak izango ditugu: *postposizioa – Aldezabalen interpretazioa – Dorren interpretazioa – preposizioa*. Eta laukote hauek aztertu ondoren, hauek ondorioztatuko ditugu:
 - Bi deskribapenen arteko adostasun-maila. Gainera, hemendik lortuko dugu ingelesaren edo espainieraren zuloak topatzea.
 - Lortuko dugun etiketa Aldezabalen eta Dorren artekoa da, eta biekin izango du lotura. Halako hirukotea izango dugu: *postposizioa – interpretazio bateratua – preposizioa*.
 - Aldezabalen eta Dorren arteko lotura interpretazio bateratuaren bitartez.

Metodoa postposizio-atzizki guztiei aplikatu ondoren, interpretazio-zerrenda bateratua lortu dugu, eta baita Aldezabalen eta Dorren interpretazioen arteko lotura. Ondoren metodoa erakutsiko dugu, baina kasu zehatz batekin. Hona, postposizio-atzizki instrumentala hartuz gero, nolako emaitza lortu dugun.

VI.4.1. Metodoa *instrumentalari* aplikatuta

Metodoa adibide baten bitartez argiago geratuko delakoan, postposizio-atzizki *instrumentala* hartu eta honekin nola jokatu dugun pausoz pauso ikusiko dugu.

Aldezabalen lana hartu eta bertan atzizki honek dituen balizko interpretazioak ikusiko ditugu. Lan honetan 9 interpretazio semantiko hartzen ditu postposizio-

atzizki honek: *modua*, *lekua*, *gai ukitua*, *zertaz jardun*, *egoera*, *baliabidea*, *kausa*, *gaia*, *denbora*.

Guk denak aztertu ondoren, *gaia*, *baliabidea*, *kausa*, eta *denbora* hautatu ditugu. Dagoeneko esan dugunez, *modua* erlazio berezia da, baina beste hizkuntzetan ere preposiziodun sintagmak agertzen direnez, eta gure ondorengo lanetarako hauek ezagutzea interesgarria izan daitekeenez zail izango delako preposizio-sintagma hauek denak lexikalizatu gisa agertzea hiztegieta, *modua* erlazioa ere txertatu egin dugu aurrekoen artean. *egoerak* oso arraroa dirudi, eta aditzen argumentuei oso lotuta dago —*aldarte txarrez egin*—; *lekuak* gehiago dirudi *gaia* izan daitekeen aldaturen argumentuari lotuta —*etxez aldatu da*—; eta *gai ukitua* eta *zertaz jardun* ere *gai* gisa tratatuko ditugu.

Beraz, bost balizko interpretazio ditugu postposizio-atzizki instrumentalarentzat, Aldezabalen lanetik abiatuta. Ondoren, hiztegi elebidunetatik hartutako adibideak —61 adibide— aztertu ditugu, eta ikusi ea postposizio-atzizki honek ba ote duen orain arte ikusi ez dugun interpretaziorik. Kasu honetan bada interpretazio bat instrumentalarentzat Aldezabalek aipatzen ez duena bere tesi-lanean. Gure iritziz, *itsasoz joan zen Donostiara* adibidean, instrumentala daraman elementuak *bidea* adieraziko du. Beraz, denera sei interpretazio izango ditugu atzizki honentzat.

Interpretazioak	Euskarazko adibidea	Ingeleseko itzulpena
kausa	beldurrez isildu ziren	they shut up out of fear
baliabidea	hirira autobusez joan zen	she went to the city by bus
bidea	lehorrez joan zen	she went by land
gaia	zutaz asko daki	he knows a lot about you
denbora	hiru urtez egon ziren han	they were there for three years
modua	eskuz idatzi zuen	she write it by hand

85. taula. Euskarazko *instrumental* postposizio-atzizkiaren interpretazio-zerrenda.

Behin adibide guztiak etiketatu ondoren halako hirukoteak ditugu: *interpretazioa* – *postposizio-atzizkia* – *preposizioa*.

Interpretazioak	Ingeleseko preposizioak	Espainierazko preposizioak
baliabidea	Ø, by , for, in , on , with	Ø, a, con , en, por
bidea	Ø, along, by , by way of, on , round, through	a través de, por , por delante de, por encima de, sobre
denbora	Ø, at , by , during , for, in , on	Ø, a, de , durante , en
gaia	Ø, about , at, for, of, on , in , with	Ø, a, acerca de , de , en, sobre , con
kausa	Ø, because of , by , due to, for, from , in, of, on account of, out of , on behalf of	Ø, a causa de , con, de , por , a
modua	Ø, at, by , in , on , with	a, de , en

86. taula. Postposizio-atzizki instrumentala eta erdarazko preposizioen arteko lotura interpretazio jakin bati dagokionez.⁹⁶

Hirukote hauekin datu-basea osatu ondoren, ingelesezko eta espainierazko preposizioen interpretazio hauek Dorren lanean topatu ditugunekin konparatu ditugu. Konparaketa honekin, Dorren eta gure interpretazioen lotura lortu dugu, eta, hala, laukoteak lortzen ditugu. Horrezaz gain, eskuz lortutakoa eta Dorren zerrenda konparatu ditugu eta emaitzak 87. taulan paratu ditugu: zenbat laukote sortu zaizkigun —*laukoteak*—, ingelesezko preposizioentzat eta espainierazko preposizioentzat eskuz lortu ditugun interpretazioak zenbatetan datozen bat egitura lexiko-kontzeptualetako rol tematikoekin —*lotura ona*—, gure interpretazioa egitura lexiko-kontzeptualetako primitibo bat zenbatetan den —*primitibo-interpretazio arazoa*—, ingelesezko edo espainierazko preposizio batentzat eskuz lortu dugun interpretazioen bat zenbatetan ez den egitura lexiko-kontzeptualetan agertzen —*azaltzen ez diren interpretazioak*—, eta eskuzko azterketan atera zaigun preposizioen bat zenbatetan ez duen egitura lexiko-kontzeptualik —*azaltzen ez diren preposizioak*—.

	Ingelesa	%	Espainiera	%
laukoteak: postposizioa – interpretazioa - preposizioa - LCS	45		28	
Lotura ona	18	40,00	8	28,57
primitibo-interpretazio arazoa	21	46,67	8	28,57
azaltzen ez diren interpretazioak	4	8,89	5	17,86
azaltzen ez diren preposizioak	2	4,44	7	25,00

87. taula. Instrumentalaren interpretazioak eta LCSen arteko lotura.

Loturak egin ondoren, 45 laukote ditugu ingelesezko preposizioak eta euskarazko instrumentala kontuan hartuz gero, eta 28 laukote espainierazko preposizioak kontuan hartuz gero. Hauetatik 18tan esango dugu lotura ona dela ingelesaren

⁹⁶ Letra lodiz dauden preposizioak dira adibideen azterketatik zuzenean lortu ditugunak —25 ingelesezko eta 19 espainierazko—; gainontzekoak, beste postposizio-atzizkiak aztertzerakoan interpretazio berak dituztenen ebakiduretatik sortu dira.

kasuan, eta 8 espainieraren kasuan. Guk loturari egoki deritzogu gure interpretazioa Dorren errepresentazioan rol tematikoa denean (ikus 88. taula).

Gure interpretazioak	Dorr-en LCSetako rolak
gaia	perc
gaia	th
gaia	info
instrumentua	ins
denbora	time

88. taula. Dorren rol tematikoekin egin dugun lotura.

Beste arazoa da, aipatu dugun bezala, Aldezabalek interpretazio semantiko deitzen dituen zenbait Dorren deskribapenean primitibo gisa agertzea. Ingelesaren kasuan 21 aldiz gertatzen da kontu hau eta espainieraren kasuan 8 aldiz. Postposizio-atzizki honetan bi interpretazio dira arazo hau sortzen dutenak, *kausa* —8 ingelesez eta 3 espainieraz— eta *bidea* —7 aldiz ingelesez eta 2 espainieraz—. (Ikus 89. taula).

Gure interpretazioak	Dorr-en interpretazioak
kausa	cause
bidea	path (TO, TOWARD, VIA)

89. taula. Dorren primitiboekin lotura.

Gainera, *kausa* primitiboa dela-eta esan behar da aditzen egitura lexiko-kontzeptualetan bai agertzen dela, baina preposizioen egituretan ez. Gure *kausa* etiketa Dorren errepresentazioarekin lotu nahi izanez gero, haren primitibo eta rolak biltzen dituen etiketa bat beharko dugu.

Moduari dagokionez, Dorrek errepresentazioa azaltzen duenean, interpretazio hau rol tematiko gisa azaltzen du baina rol tematiko hau ez da inoiz agertzen haien errepresentazioan; egitura lexiko-kontzeptualetako *typen* artean ere badago *manner* irakurketa, eta hau bai agertzen da egitura lexiko-kontzeptualetako errepresentazioan. Honen arrazoia bilatzen hasita, pentsa dezakegu *modu* hau ez dela, oro har, argumentu baten interpretazioa, eta haien lanaren muina aditzen argumentuak ezagutzean datza; edo beste irakurketa izan daiteke halako preposizio-sintagmak dagoeneko lexikalizatuta dauden sintagmak direla, eta hiztegian agertzen diren elementuak direnez, ez dutela errepresentaziorik behar. Guk *modu* hau galdutako interpretazio gisa kontatu dugu, eta 6 aldiz agertzen da ingelesezko preposizioekin, eta 3 aldiz espainierazko preposizioekin.

denbora interpretazioarekin ere zerbait bertsua gertatzen dela esan daiteke. Hau Dorren errepresentazioaren zulorik handienetakoa dela pentsa dezakegu. Kasu honetan galdutako interpretazioen artean 3 dira *denbora* — \emptyset , *by*, *in*— adierazten dutenak, eta bestea *instrumentala* da —*in*—. Espainieraren kasuan 4 dira *denbora* interpretaziokoak —*a*, *de*, *desde*, *en*— eta *instrumentala* kasu batean —*a*—.

Ingelesaren edo espainieraren atalean aztertu gabe dauden preposizioak kontuan izanez gero, ingelesaren kasuan aztertu gabe daude *on account of* eta *on behalf of*; eta espainieraren kasuan, \emptyset ,⁹⁷ *a causa de*, *acerca de*, eta *durante*.

Aurreko lotura guztiak lortu ditugu datu-basean hiztegi elebidunetatik lortutako pare guztiak jarri ondoren. Honen ondoren, Dorrek eta Habashek duten preposizioen zerrenda hartu dugu eta gure datu-basea osatu. Haien rol tematikoak gure interpretazioekin modu honetan lotu ditugu:

rol-tematikoa	interpretazioa
agent	kausa
experiencer	esperimentatzailea
information	gaia
theme	gaia
source	abiapuntuko kokapena
goal	helburuko kokapena
perceived	gaia
predicate	ezaugarria
locational	kokapena
time	denbora
benefactive	helburuko kokapena
instrument	baliabidea

90. taula. Dorren rolen eta gure interpretazioen arteko lotura.

Aurrekoak lotzen ditugu bai argumentu direnetan, eta bai modifikatzaile direnetan. Ondorengo lau rol tematikoek, aldiz, ez dute gure interpretazioekin lotura garbirik, eta, oraingoz, halakoak ez lotzea erabaki dugu; beraz, beheko interpretaziorik duten preposizioak ez ditugu gurera pasatuko.

possessional
proposition
purpose
particle

91. taula. LCSetatik esportatuko⁹⁸ ez ditugun interpretazioak.

⁹⁷ Dagoeneko esan dugun bezala, \emptyset honek *preposizio eza* adierazten du, hau da, beste hizkuntzetan agertzen diren subjektu eta objektuak. Izan ere, hauek preposiziorik gabe agertzen dira.

⁹⁸ *Esportatzea* haien egitura lexiko-kontzeptualetatik gure datu-basera informazioa automatikoki pasatzea da.

Rol horiek dituzten preposizioak taulan txertatu ondoren, instrumentalaren ebakidura asko handitzen da. Hona ebakidura-aula nola geratuko litzatekeen:

Instrumentalaren interpretazioak	Ingeles-preposizioak	Espainiera-preposizioak
Baliabidea	Ø, as, by, for, from, in, of, on, out of, with, without	Ø, a, con, de, en, por
Bidea	Ø, along, by, by way of, on, round, through	a través de, por
Denbora	Ø, about, after, ahead of, around, as, as of, at, back to, before, behind, between, beyond, by, close to, during, following, for, from, in, in relation to, near, on, per, previous to, prior to, pursuant to, related to, relative to, round, since, through, throughout, till, to, until, with respect to, within	Ø, a, de, desde, durante, en
Gaia	Ø, about, after, against, around, at, before, from, in, into, of, on, over, that, through, to, with	Ø, a, acerca de, ante, con, contra, de, en, por, que, sobre
Kausa	Ø, at, because of, by, due to, for, from, in, on account of, on behalf of, out of	Ø, a causa de, con, de, desde, por
Modua	Ø, at, by, in, on, with	a, de, en

92. taula. LCSetako preposizioak gura pasa ondoren instrumental eta preposizioen arteko ebakidura.

Lehenengo ebakidura-aulari erreparatuz gero, ingelesez, 45 hirukote —*instrumentala-interpretazioa-preposizioa*— genituen, eta orain 90 ditugu. Espainieraz, berriz, hasieran 28 hirukote genituen, eta orain 34. Beraz, nabaria da Marylandeko unibertsitateko sailkapenean ingelesaren preposizioak nahikoa landuta dituztela, eta espainierarenak, berriz, ez dituztela hain landuta.

Honen ondoren, euskarazko postposizio konposatuak landu ditugu, eta hor ere beste hizkuntzetako hainbat pare sortu zaizkigu. 93. taulan aurreko ez gain sortu zaizkigun ebakidura berriak eman ditugu.

Instrumentalaren interpretazioak	Ingelesezko preposizioak	Espainierazko preposizioak
Baliabidea	by means of, through, via	a cambio de, a cuenta de, a través de, de la mano de, en compensación de, por obra de, por medio de, sin
Bidea	above, across, around, from behind, from this side of, over, past, throughout, up	a lo largo de, al lado de, al otro lado de, arriba, detrás de, enfrente de, hacia arriba, junto a, por al lado de, por debajo de, por delante de, por dentro de, por detrás de, por el centro de, por el otro lado de, por encima de, por este lado de, sobre, tras de
Denbora	across, inside, next, over, past, towards, up	a continuación, a lo largo de, a través de, al cabo de, anterior a, antes de, cerca de, dentro de, después de, detrás de, en adelante, en el plazo de, en la temporada de, en tiempo de, hacia, hasta, inmediatamente después de, mientras, por, sobre, tras
Gaia	concerning, regarding, with respect to	con respecto a
Kausa	thanks to	a petición de, gracias a, por motivo de
Modua		

93. taula. Postposizio konposatuak landu ondorengo ebakidura berriak.

Ingelesaren kasuan, 23 hirukote berri ditugu. Espero bezala, ez dira asko berriak, nahiz eta euskarazko postposizio konposatu asko landu, ingeleseko preposizioen deskribapena ona delako.

Espainierari dagokionez, berriz, 53 hirukote berri ditugu. Honek adierazten du Dorren deskribapena ez dela hain zehatza espainierako preposizioei dagokienez.

Kasu berriei erreparatuz, nabaria da kasu gehienak *bide* eta *denbora* interpretazioetan topatuko ditugula, izan ere, *bideri* dagokionez, halako rol tematikorik ez dagoenez Dorren errepresentazioan, ezin izan dugu interpretazio hori duen preposiziorik esportatu;⁹⁹ eta *denbora* interpretazioari dagokionez, halakorik gutxitan agertzen da, errepresentazioan aditzen argumentuak dituztelako helburu, eta ez aditz baten inguruan agertzen den edozein elementu.

kausa interpretazioak ez du aldaketa nabarmenik jaso, postposizio konposatuek gutxitan dutelako irakurketa hau. *moduari* dagokionez, berriz, oraingoz *modua* interpretaziodun postposizio konposaturik ez dugu topatu.

Lan hau postposizio guztiekin egin ondoren, hona gure deskribapenean izango dugun interpretazio-zerrenda:

abiapuntuko kokapena
baliabidea
bidea
denbora
esperimentatzailea
ezaugarria
gaia
helburuko kokapena
kausa
kokapena
konpainia
modua
neurria
xedea

94. taula. Interpretazio-zerrenda.

Hemendik aurrera ikusiko dugu interpretazio bakoitza zein postposizio —-atzizkik nahiz konposatuk— hartzen duen.

⁹⁹ Gogoan izan *bide* irakurketa primitibo gisa errepresentatzen dela egitura lexiko-kontzeptualean.

VI.5. Postposizio-atzizkiak eta interpretazioak

Atal honetan postposizio-atzizki bakoitzak —baita kasu gramatikalek ere— izan ditzaketen interpretazioen zerrenda laburra jarriko dugu. D. eranskinean postposizio bakoitzari buruzko informazio osoagoa paratu dugu: esleitu dizkiogun interpretazioak eta adibideak, eta alde batera utzitako zenbait interpretazioen azalpenak.

95. taulan xxx agertzeak zutabe horretan dagoen postposizioak errenka horretan dagoen interpretazioa baduela adierazten du.

	-tik	-ra	-rantz	-raino	-entzat	-n	-z	-gatik	-kin	abs	dat	erg
abiapuntuko-kokapena	xxx							xxx		xxx	xxx	xxx
baliabidea	xxx					xxx	xxx		xxx			xxx
bidea	xxx						xxx					
denbora	xxx	xxx				xxx	xxx			xxx		
esperimentatzailea										xxx	xxx	xxx
ezaugarria										xxx		
gaia						xxx	xxx		xxx	xxx	xxx	xxx
helburuko-kokapena		xxx	xxx	xxx		xxx			xxx	xxx	xxx	xxx
kausa	xxx						xxx	xxx	xxx		xxx	xxx
kokapena						xxx						
konpainia									xxx			
modua	xxx	xxx				xxx	xxx		xxx			
neurria										xxx		
xedea					xxx							

95. taula. Postposizio-atzizkiak eta interpretazioak.

VI.6. Postposizio konposatuak

Orain arte ikusi ditugun postposizioak elementu bakarra dutenak dira —atzizkiak—, eta, hemendik aurrera elementu bat baino gehiagoz osatuta daudenak ikusiko ditugu. Talde honetakoak ditugu aurretik deus hartzen ez duten postposizioak, aurretik postposizio-atzizki bat hartzen dutenak, eta aurretik eta ondoren hartzen dutenak postposizio-atzizkiren bat:

etxe gabe geratu zen
 etxetik at zegoen zakurra
 etxetik hurbilean hiruzpalau zuhaitz zeuden

Lehenengo adibidean postposizio konposatua *-Ø gabe* dela esango dugu, bigarreanean, *-tik at*, eta hirugarrenean, *-tik hurbil-ean*.

Postposizio konposatu hauek atzizki direnak baino askoz gehiago dira, eta anbiguotasun txikiagokoak. Lehenik anbiguoak diren postposizio konposatuen zerrenda emango dugu —hau da, interpretazio bat baino gehiago izan ditzaketenak—, eta, ondoren, interpretazioen araberako multzoak emango ditugu.

Denera 172 postposizio konposatu zerrendatu ditugu; hauetatik, 31k dute bi interpretazioko anbiguotasuna, eta 4k hiru interpretazioko anbiguotasuna.

VI.6.1. Anbiguoak

Anbiguotasuna, *lekua* —eta *leku* adieraren ingurukoak— eta *denbora* interpretazioen artean maiz agertzen da. Anbiguotasun hau duten postposizio konposatuen artean, leku-denbora adierazten duten postposizio-atzizkiak hartzen dituzten elementu asko ditugu, eta, horiek maiz dute anbiguotasun hau:

Interpretazioak	Postposizio konposatuak
denbora / kokapena	-0 / -en aitzinEAN ¹⁰⁰ , -0 / -en aurreAN, -0 / -en ondoAN, -0 / -en osteAN
denbora / helburuko kokapena	-tik lekora
denbora / bidea	-0 / -n / -tik zehar0
denbora / bidea / kokapena	-0 / -en aurreTIK, -0 / -en gibelETIK, -0 / -en ondoTIK, -0 / -en osteTIK

96. taula. *denbora* / *lekua* anbiguotasuna postposizio konposatuetan.

Beste multzo bat osatzen dute *leku* mota ezberdinen arteko anbiguotasunek. Hemen nagusi izango da *abiapuntuko kokapena* eta *bidearen* artekoa; leku-denborazko atzizkiak hartzen dituzten elementuek *-tik* hartzen dutenean, ia beti bi interpretazio horiek izango dituzte.

Interpretazioak	Postposizio konposatuak
bidea / abiapuntuko kokapena	-0 / -en aitzinETIK, -0 / -en atzeTIK, -0 / -en azpiTIK, -0 / -en barneTIK, -0 / -en barrenETIK, -0 / -en barruTIK, -0 / -en erdiTIK, -0 / -en gainETIK, -0 / -en inguruTIK, -en / -z bestaldeTIK, -en / -z beste aldeTIK, -z honaindiTIK
bidea / helburuko kokapena	-0 / -en / -n / -tik barna
bidea / kokapena	-0 / -en aldamenETIK
kokapena / helburuko kokapena	-en / -i kontra, -en aurka
helburuko kokapena / gaia	-i buruz

97. taula. *leku* moten arteko anbiguotasuna postposizio konposatuetan.

Azkeneko bi multzo anbiguoak *ezaugarria* interpretazioaren ingurukoak dira.

¹⁰⁰ -0 / -en aitzinEAN jartzen dugunean, adierazi nahi dugu aurreko hitza markarik gabea dela (0), edo genitiboa duena (-en), postposizio-hitza aitzin dela, eta honek inesiboa hartzen duela (-ean). Halako postposizioak bat bezala kontatu ditugu.

Interpretazioak	Postposizio konposatuak
baliabidea / ezaugarria	-0 / -en orde, -0 / -en truke, -0 / -ik barik, -0 / -ik gabe, -en aldeRA, -en lekuAN, -en trukeAN
kausa / ezaugarria	-0 / -en eskeZ

98. taula. ezaugarriaren inguruko anbiguotasuna postposizio konposatuetan.

VI.6.2. Ez-anbiguoak

Azterketa egin ondoren, hona ez-anbiguo diren postposizioak interpretazioen arabera multzokatuta. Hemen taula jarriko dugu, eta E. eranskinean interpretazio bakoitzaren barruan dauden postposizio konposatuak eta adibideak paratu ditugu. H. eranskinean postposizio bakoitzaren interpretazioa eta espainierazko eta ingelesezko preposizioekin dituzten lotura zuzenak jarri ditugu.

Interpretazioak	Postposizio konposatuak
abiapuntuko kokapena	-0 alboTIK, -0 / -en aldeTIK, -0 / -en arteTIK, -0 / -en baitaTIK, -z haraindik, -z haraindiTIK, -0 kanpoTIK, -en menETIK, -en mendeTIK, -en menpeTIK, -ko parteTIK, -0 / -en peTIK
baliabidea (bitartekoa)	-0 / -en bideZ, -en bitarteZ, -en eskuTIK, -en lekuAN, -en medioZ, -en trukeAN
bidea	-0 / -en alboTIK, -n / -tik barrena, -n / -z gaindi, -0 / -n goiti, -0 / -n / -tik / -z gora, -tik / -n landa, -0 / -en pareTIK
denbora	-0 / -en aitzin, -ak aldeAN, -ak aldeRA, -0 / -ak arte, -0 arteAN, -0 arteraino, -0 barru, -0 / -ra bitarteAN, -ak bitarte, -ak bitarteAN, -tik goiti, -0 / -ak inguru, -0 / -ak inguruAN, -0 / -ak irian, -0 iria, -0 / -en ondoren, -0 parteAN
ezaugarria	-en alde, -0 / -en antzERA, -0 antzEAN, -0 / -en antzo, -en arabera, -0 arau, -en begira, -i begira, -ra begira, -en beha, -i beha, -ra beha, -0 / -en beharREAN, -0 bezala, -0 / -en bila, -en bizkar, -en bizkarRETIK, -en eraN, -en eraRA, -ko eraRA, -en ereduRA, -en ereduZ, -0 / -en eske, -0 / -en eskeZ, -en esku, -en fabore, -en faboreTAN, -0 gabeZ, -0 / -rik gabarik, -z gaindi, -0 / -en gisa, -0 / -en gisaN, -0 / -en gisaRA, -0 keta, -0 legez, -en lekuAN, -en mende, -en menpe, -en moduan, -en modura, -en ordean, -0 / -en ordeZ, -en pare, -0 / -en trukeAN, -0 / -en xerka, -0 / -en zain
gaia	-en gain, -en gainEAN
helburuko kokapena	-en / -i aitzin, -0 / -en aitzinERA, -0 / -en alboRA, -0 / -en aldamenEAN, -0 / -en aldamenERA, -0 / -en aldeRA, -0 / -en arteRA, -0 / -ra arte, -0 / -en atzeRA, -0 / -en aurreRA, -0 / -en azpiAN, -0 / -en azpiRA, -0 / -en baitaRA, -0 / -en barneRA, -0 / -en barrenERA, -0 / -en barruRA, -0 / -n / -tik beheiti, -0 / -n / -tik / -z behera, -en / -z bestaldeRA, -en / -z beste aldeRA, -ra bitarte, -ra buruz, -0 / -en erdiRA, -0 / -en gainERA, -0 / -en gibelERA, -z haraindiRA, -z honaindiRA, -0 / -en inguruRA, -0 kanpoRA, -tik kanpoRA, -en menERA, -en mendeRA, -en menpeRA, -0 / -en ondoRA, -0 / -en osteRA, -0 / -en pareRA, -ko parteRA, -0 / -en peRA
kausa	-0 / -en ariaZ, -0 / -en eskeZ, -i esker, -i eskerRAK, -0 / -en kariAZ, -en kausaZ
kokapena	-0 / -en alboAN, -0 / -en aldeAN, -tik alde, -0 / -en arteAN, -tik at, -0 / -en atzeAN, -tik aurreRA, -0 / -en baitaN, -0 / -en barneAN, -0 / -en barrenEAN, -n barrena, -0 / -en barruAN, -en / -z bestadeAN, -z bestalde, -en / -z beste aldeAN, -en bitarteAN, -0 / -en erdiAN, -0 / -en gainEAN, -0 gaindi, -tik gertu, -0 / -en gibelEAN, -z haraindi, -z haraindiAN, -z honaindi, -z honaindiAN, -tik hur, -tik hurREAN, -i hurbil, -tik hurbil, -tik hurbileAN, -0 / -en inguruan, -0 inguru, -tik / -z kanpo, -0 kanpoAN, -tik kanpoAN, -en kontraN, -tik landa, -en menEAN, -en mendeAN, -en menpeAN, -0 / -en pareAN, -ko parteAN, -0 / -en peAN, -kin batERA

99. taula. Postposizio konposatuak eta interpretazioak

Zerrenda osoak ikusi ondoren, ikus dezagun egindako esperimientua.

VI.7. EHko definizioetan dauden postposizioen desanbiguazioa

Gure helburua EHko definizioetan azaltzen diren postposizioen interpretazioa ezagutzea da, eta horretarako, interpretazio hori desanbiguatu egin behar da. Desanbiguazio-lan hori egiteko erabili dugun metodoa azalduko dugu.

Postposizioen interpretazioa lortu ahal izateko, hizkuntza ezberdinetako hiztegi-tako definizioak parekatu ditugu, bertan agertzen diren preposizioak eta postposizioak lotuz, eta haien interpretazioa dagozkien interpretazioen ebakiduratik erauziz. Hona erabili dugun metodoa pausoz pauso:

- EHko definizioetan dauden postposizioak ezagutu.
- Sarrera-hitzen eta postposizioaren aurrean eta atzean dauden hitzen itzulpena egin.
- Egindako itzulpenak baliatuz, hizkuntza ezberdinetako hiztegi-tan pareko diren definizioak lotu, eta postposizioak eta preposizioak parekatu.
- Postposizio eta preposizioen interpretazioak erabiliz, desanbiguazioa egin.

VI.7.1. EHko definizioetan postposizioak bilatu

Esperimentu honetarako euskarazko postposizio-atzizki instrumentala aukeratu dugu, eta lehenik bilatu beharko dugu instrumentala EHko zein definizio-tan agertzen den. Horretarako dagoeneko egin dugun definizioen analisi morfologikoa erabili dugu. Hartara, instrumentala duten definizioak erauzi ditugu, eta bertatik informazio hau gorde dugu: sarrera-hitza, instrumentala duen lema, postposizio-atzizkia bera¹⁰¹ eta ondoan eskuinean duen lema.

Azpian dugu *ildo* sarrera hiztegian agertzen den bezala:

ildo. iz. A1 **Goldeaz** lurra **irauliz** egiten den irekidura
luzea.

Hona aurreko definizioaren bertsio lematizatua:

¹⁰¹ Esperimentu honetarako ez du zentzu handirik, badakigulako gure helburua instrumentala dela, baina postposizio guztiekin lan egiten hastean, beharrezkoa izango dugu eremu hau, eta honexegatik paratu dugu.

```
/<@@lema      ildo>/<ID>/  
/<@@Adiera_string  A1.>/<ID>/  
/<@@Kategoria  iz. >/<ID>/  
"<Goldeaz>"  
  "golde"  IZE ARR DEK INS NUMS MUGM  
"<lurra>"  
  "lur"    IZE ARR DEK ABS NUMS MUGM  
"<irauliz>"  
  "irauli" ADI SIN AMM PART DEK INS MG  
"<egiten>"  
  "egin"   ADI SIN AMM ADOIN ASP EZBU  
"<den>"  
  "izan"   ADL A1 NOR NR_HU ERL MEN ERLT  
"<irekidura>"  
  "irekidura" IZE ARR DEK ABS MG  
"<luzea>"  
  "luze"    ADJ IZO DEK ABS NUMS MUGM  
"<$.>"  
  PUNT_PUNT
```

Eta, hona aurreko definiziotik erauzitako bi hirukoteak:

```
golde#INS#lur  
irauli#INS#egin
```

Dagoeneko jabetuko zinetenez, orain arte esaten aritu gara kapitulu honetan aditz eta izen baten arteko erlazioei erreparatuko diegula, eta adibide honetan bi izen ditugu —golde eta lur—. Esperimentu hau egiteko ez dugu inongo analizatzaile sintaktikorik erabili, eta are gutxiago dependentziak markatuko dituen analizatzailerik. Ondo-ondoan dauden hitzak hartu ditugu lehen esperimentu honetan ea deskribatutako metodoak balio ote duen probatzeko, eta baliagarritasunaren arabera, beste jakintza motak txertatu, hala nola, gaur egun Ixa Taldean garatzen ari diren dependentziak sortzeko gramatika (Aduriz *et al.*, 2004).

Elkarren ondoan dauden hitzak hartuz, informazioa galduko dugu, beste hizkuntzetako definizioetan bi hitz horiek elkarren ondoan agertu behar dutelako, baina adibide honetan ikusiko dugunez, zenbaitetan postposizioaren interpretazioa desanbiguatzeko gai gara.

VI.7.2. Itzulpenak egin

Behin sarrera-hitza eta hirukotea dugula, hauen itzulpena lortu behar dugu hiztegi elebidunak erabiliz. Lehenik sarrera-hitza itzuliko dugu eta itzulpen guztien definizioak bilatuko ditugu hiztegi elebakarretan —ingelesez eta espainieraz—.

Itzulpenak egiteko, bi hiztegi elebidun erabili ditugu: ingelesera itzultzeko Morris (1998), eta espainierara itzultzeko Elhuyar (1996). Erabili ditugun hiztegi elebakarrak, berriz, sei dira. Ingelesekoak, *Cambridge*,¹⁰² *Heritage*¹⁰³ eta *Wordsmith*.¹⁰⁴ Eta espainierazkoak, *Colmex*,¹⁰⁵ *RAE*¹⁰⁶ eta *Vox*.¹⁰⁷

Bi hiztegi elebidunen bertsio elektronikoa gure zerbitzarietan dugu, eta, beste sei hiztegi elebakarrak saretik atzitu ditugu.

Adibideari helduz, hona *ildoren* itzulpenak:

ildo: furrow, groove, flute, fluting, path, way
ildo: surco, camino, dirección, sentido, derrotero, estría

Guk itzulpen horien guztien definizioak bilatu ditugu aipatutako sei hiztegietan, eta hona lortu ditugun batzuk:

furrow#A long, narrow, shallow trench made in the **ground** by a **plow**.
surco#Excavación alargada, angosta y poco profunda que se hace paralelamente en la **tierra** con el **arado** , para sembrarla después.
surco#Hendedura que se hace en la **tierra** con el **arado**.

VI.7.3. Definizioak lotu

Definizioak euskaraz, espainieraz eta ingelesez izan ondoren, paraleloak direnak ezagutu behar ditugu. Horretarako, hirukotearen itzulpenak lortu behar ditugu, eta aztertu ea beste hizkuntzetako definizio horietan agertzen diren eta ea elkarren ondoan ote dauden. Hala bada, ingelesez eta espainieraz ere hirukoteak izango ditugu: ingelesezko edo espainierazko lehenengo hitza —*ground*, *tierra*— da euskarazko hirukotean agertzen den azken hitzaren itzulpena —*lur*—; tartean preposizioa dugu, eta hori da euskarazko postposizio-atzizkiaren itzulpena; eta ingelesez eta espainieraz dugun azken hitza —*plow*, *arado*— da euskarazko hirukotean dagoen lehenengo hitzaren itzulpena —*golde*—.

Adibidez, aurreko definizioak aztertuta, euskarazko hirukoteetako batekin lotu ahal izango ditugu:

¹⁰² <http://dictionary.cambridge.org/> (2004-11-30ean atzitua).

¹⁰³ <http://www.bartleby.com/61/> (2004-11-30ean atzitua).

¹⁰⁴ <http://www.wordsmyth.net/> (2004-11-30ean atzitua).

¹⁰⁵ <http://mezcal.colmex.mx/Scripts/Dem/principal.htm> (2004-11-30ean atzitua).

¹⁰⁶ <http://www.rae.es/> (2004-11-30ean atzitua).

¹⁰⁷ <http://www.diccionarios.com/index.phtml?redirect=true> (2004-11-30ean atzitua).

golde#INS#lur
ground#by#plow
tierra#con#arado

*golde*ren itzulpen ditugu *plow* eta *arado*; eta *lurre*nak *ground* eta *tierra*.

Halako hirukoterik ez duten definizioak baztertu egingo ditugu. Adibide gisa aipatuko dugu euskarazko definizioan beste hirukote bat bagenuen, eta halako hitz-segidarik ez dugu beste hizkuntzetako definizioetan topatu:

irauli#INS#egin

Hirukote hau, beraz, albo batera geratu da.

Ingeleseztan eta espainieraz dauden definizioetatik hirukote hauek atera ahal izateko, definizio horiek lematizatu behar izan ditugu eta elementu bakoitzaren kategoria zein den jakin. Ingeleserako *TnT* kategoria-etiketatzailerak erabiltzen dugu (Brants, 2000) eta lematizaziorako, *WordNet* (Miller *et al.*, 1990). Espainierarako Atserias *et al.*-ek (1998) azaldutako tresna erabiltzen dugu.

Etorkizunean dependentziak ematen dituzten analizatzaileak erabiliko ditugu definizioen arteko lotura hauek lortzeko.

VI.7.4. Desanbiguazioa

Halako laukoteak izanik, suposatu behar dugu hitz horiek testuinguru berean gertatzen badira, haien arteko erlazioa berdina dela hizkuntza guztietan, eta erlazio hori markatuko duela postposizioak edo preposizioak.

ildo#golde#ins#lur
furrow#ground#by#plow
surco#tierra#con#arado

Gure lanaren arabera, postposizio bakoitzak eta preposizio bakoitzak interpretazio-zerrenda bat du, eta kasu honetan interpretazioa *instrumentalaren*, *by*ren eta *con*en ebakidura da. Berez, *by* preposizioaren eta *instrumentalaren* arteko ebakidura hiru interpretaziokoa da, eta *con* preposizioaren eta *instrumentalaren* arteko ebakidura lau interpretaziokoa.

Laukotea	Interpretazioak
furrow#ground#by a#plow#	modu / instrumentu / denbora-iraupen
surco#tierra#con el#arado#	modu / instrumentu / kausa / eduki

100. taula. *Instrumental*ak eta *by*k, eta *instrumental*ak eta *con*ek dituzten interpretazio-ebakidurak.

Kontuan izan behar dugu espainieraz eta ingelesez, preposizioaren eta bigarren hitzaren artean artikulak —*a*, *el*— onartu ditugula.

Interpretazio hauen ebakidura bikoa dugu: *instrumentu* eta *modu*.

golde#INS#lur#	instrumentu modu
----------------	------------------

101. taula. *Instrumentala by* eta *conen* interpretazio-ebakidura.

Dakigunez, *instrumentu* da egokia, baina oraindik balizko bi interpretazio baditugu ere, beste bost baztertzea lortu dugu.

Esperimentu hau egin genuenean, interpretazio-zerrenda ez zegoen erabat zehaztua, eta orduko interpretazioak eta oraingoak ez dira berdinak. Hona esperimentua egiteko instrumentalari esleitutako interpretazioak (Agirre *et al.*, 2002):

	Euskara	Ingelesa
gaia	Seguru nago horretaz Matematikaz asko daki	I'm sure of that He's an expert in maths
denbora-iraupena	Arratsalde lasai egon nahi dut Gauetz egin dut	I like to relax of an evening I did it by night
instrumentua	Autobusez etorri naiz Belarra segaz moztu Euskaraz hitz egin	I have come by bus To cut grass with a scythe To speak in Basque
modua	Animali baten hestea betez egindako haragia Ahots ozen batez	A meat preparation made by filling an animal intestine In a loud voice
kausa	Haren aitzakiez nekatuta nago Beldurrez zurbildu Kanpoan lan egitea baztertu zuenez, lan-aukera ederra galdu zuen	Sick of his excuses To turn white of fear In refusing to work abroad, she missed an excellent job opportunity
edukitzailea	Edalontzia ardoz beteta dago Txapelaz dagoen gizona Ilez estalia	The glass is full of wine The man with the beret on Cover in hair
materia	Armairua egurrez egina dago	The wardrobe is made of wood

102. taula. Egindako esperimentuan instrumentalari esleitutako interpretazioak eta ingelesarekiko lotura.

	Euskara	Espainiera
gaia	Mariaz aritu dira Honetaz ziur naiz	Han mencionado a Maria Estoy seguro de esto
denbora-iraupena	Gauetz egin dut	Lo he hecho de noche
instrumentua	Belarra segaz moztu Euskaraz hitz egin Hiria harresiz inguratu dute	Cortar la hierba con la guadaña Hablar en vasco Han cubierto la ciudad de murallas
modua	Oinez etorri zen Ahots ozen batez Bere familiaren laguntzaz erosi zuen	Vino a pie En voz alta Lo compró con la ayuda de su familia
kausa	Beldurrez zurbildu Maitasunez hil	Con el miedo me quedé pálido Morir de amor
edukitzailea	Edalontzia ardoz beteta dago Txapelaz dagoen gizona ikusi dut	El baso esta lleno de vino He visto a un hombre con boina
materia	Armairua egurrez egina dago	El armario está hecho de madera

103. taula. Egindako esperimentuan instrumentalari esleitutako interpretazioak eta espainierarekiko lotura.

Interpretazio-zerrenda definitiboan ez dira *edukitzailea* eta *materia* agertzen. Lehenengoa *gaitzat* hartu dugu, eta bigarrena *baliabide (instrumentu)* bezala. Eta *bide* interpretazioa gehitu diogu.

Esperimentua egin genuenean, adibideetan atera zitzaizkigun espainierazko eta ingelesezko preposizioei egin genien kasu, eta ez preposizio guztiei. 104. taulan daude xekin markatuta lortu genituen ebakidurak (Agirre *et al.*, 2002):

	Euskara	Ingelesa				Espainiera			
	-z (ins.)	of	by	with	in	de	con	a	en
gai	x	x			x	x		x	
denbora-iraupen	x	x	x			x			
instrumentu	x		x	x	x	x	x		x
modu	x		x		x	x	x	x	x
kausa	x	x		x	x	x	x		
edukitzaile	x	x		x	x	x	x		
materia	x	x				x			

104. taula. Instrumentalaren interpretazioak eta erdarazko baliokideak.

Lan egiteko modua ikusi ondoren, emaitzak aipatuko ditugu.

VI.7.5. Emaitzak

Lehen esperimentu honekin metodoaren zehaztasuna neurtu nahi dugunez, eta lan honekin segi edo ez erabaki, adibide garbiak besterik ez ditugu kontuan hartu, eta zaborra sortzen dutenak baztertu. Halakoak dira postposizio konposatuetan agertzen diren instrumentalak dituztenak —*en bidez*—,¹⁰⁸ lokailu baten zati direnak —*batez ere*—, atzizki konposatu baten parte direnak —*zko*—, eta lematizazioak sortutako akatsak —*gizonezko* = *gizon* + *zko*—.

EHko 4.004 definiziotan agertzen da instrumentala. Erdarazko hiztegietan hauen parekoak bilatu ondoren, 91 hirukoterentzat lortu dugu parekorik.¹⁰⁹ Ateratako hirukoteak eskuz etiketatu ditugu eta erdua izango den lagina sortu dugu, makinak lortzen duen emaitza eta eredu hau konparatu ahal izateko.

Beraz, aurreprozesu hau egin ondoren, desanbiguazio-kopuruak atera ditugu. 105. taulan hiztegi bakoitzarekin lortu ditugun lotura-kopurua eman dugu —*denera*—, eta horietatik zenbatetan ezagutzen dugun postposizioaren interpretazio

¹⁰⁸ Esperimentu hau egin genuenean, oraindik ez genituen postposizio konposatuak landuta.

¹⁰⁹ 4.004 hauetan instrumental garbiak ez direnak ere agertzen dira. Izan ere, lotura 125 kasutan lortu badugu ere, goian paratu dugunez, horietatik instrumental hutsa 91 kasutan agertu da. Hau da, instrumental gisa ezagutzen dituenetatik, %72,8 da ona.

zuzena —*zuzenak*—. Datu hauekin metodoaren doitasuna lortu dugu, eta hasieran 7 interpretaziotakoa zen anbiguotasuna zenbatera murriztu den —*anbiguotasuna*—.

Hiztegia	denera	zuzenak	doitasuna	anbiguitasuna
cambridge	16	15	%93,8	4,0
Am. heritage	34	32	%94,1	3,2
wordsmith	26	26	%100	3,7
Colmex	10	9	%90	2,6
vox_ya	7	7	%100	2,8
Rae	26	25	%96,2	2,8
denera	91	86	%94,5	3,1
usuenak diren kasuak	91	85	%93,4	3,0

105. taula. Hiztegi bakoitzaren emaitzak eta guztiak batera ateratakoak.

Azkenean 91 definizio ditugu aztertzeko, eta hauetatik 86 markatzen ditugu ondo; beraz, %94,5eko doitasuna dugu, eta anbiguotasuna 3,1. Hau da, instrumental bat agertzen den bakoitzean, 7 balizko interpretazio baldin baditugu, metodo hau aplikatu ondoren interpretazioen kopuruaren batez bestekoa 3,1 da.

Kontuan izatekoa da gehien agertzen diren hiru interpretazioekin etiketatuz gero, doitasuna %93,4koa litzatekeela, baina hau ez da batere erreala; eredu den lagina eta probetarako erabili den lagina bera da, eta honek prezisioetan alerazioak sortzen ditu. Hau guztia kontuan izanda, lortutako emaitzak oso onak dira.

Eskuz etiketatutako laginean, hona interpretazioen maiztasuna:

Interpretazioa	Maiztasuna
instrumentua	37
edukitzailea	35
instrumentua / modua	7
modua	6
gaia	5
kausa	1
materia	0
denbora-iraupen	0
Denera markatuta	91

106. taula. Eredu den laginean dauden instrumentalaren interpretazio kopuruak.

Hirugarren etiketa usuena bikoitza da; honek esan nahi du eskuz etiketatu duen hizkuntzalaria ez dela gai izan hor azpian zein interpretazio dagoen zehazteko, eta etiketa bikoitza erabili duela.

Emaitza on hauek ikusita erabaki dugu interpretazio-zerrenda zehaztea euskarazko postposizioentzat, eta beste hizkuntzetan egindako preposizioen interpretazioen zerrendak aztertzea interpretazioak bateragarriak izan daitezzen. Hala,

aurretik aipatutako lanak hartu ditugu oinarri (Aldezabal, 2004; Dorr, 1993), eta azaldutako prozesu osoa abian jarri. Prozesu honen emaitzak aurki izango ditugu, eta tesi-lan honen etorkizuneko lana gisa aipatu dira.

VI.8. Ondorioak eta etorkizuneko lana

Atal honetan patroï lexikalei eta postposizioei dagozkion ondorioak paratu ditugu.

Patroiei dagokienez, EHko definizioetan dauden patroï lexikalak ezagutzen saiatu gara. Horretarako III.8. atalean aipatutako *token*-segidak hartu ditugu oinarri. Patroï itxura duten *token*-segida usuenak zerrendatuta izan ondoren, inguruko hitzen artean erlaziorik ezartzen dutenak hartu ditugu patroï lexikaltzat. Halako asko ez dira agertzen EHn —718 markatu ditugu—, eta definizioan dauden erlazioak zehazteko hitzen inguruko sintaxiari erreparatu diogu.

Definizioen *differentiaren* azterketa semantikoaren lehenengo pauso gisa postposizioen interpretazioak desanbiguatzen saiatu gara. Postposizio horien interpretazioa desanbiguatu ahal izateko, beste hizkuntzetako hiztegi elebidunetako pareko definizioak erabili ditugu. Bi definizio pareko izanez gero, euskarazko postposizioek eta erdarako preposizioek interpretazio bera dutela pentsatu dugu, eta hala beraien interpretazioa bi hizkuntza ezberdinetako elementu horien ebakiduraren bat izango da.

Interpretazio horiek desanbiguatzeke, beraz, hizkuntzartekotasuna erabili dugu, eta horretarako ezinbesteko baldintza izan da hizkuntza guztietako elementuentzat interpretazio bateratuak izatea.

Interpretazio-zerrenda bateratua lortzea nahikoa lan konplikatua izan da, gure abiapuntuak aski ezberdinak izan direlako, baina azkenik tarteko interpretazio-zerrenda bat sortu ahal izan dugu, eta horretarako orain arte arlo honetan egin izan diren lanak baliatu ditugu.

Perpaus baten barruan dauden loturak aztertzen hasita, maiz jo izan da aditzaren azpikategoriazioa eta argumentu-egitura definitzera. Lotura-elementuak erabiltzearen kontu hau nahikoa berria da, eta emaitza baikorrak ematen dituela ikus ahal izan dugu dagoeneko egin den esperimentuari esker. Beste hizkuntzetan ere metodo hau baliatzen ari dira, Dorr ingeleserako, Saint Dizier frantseserako, eta

nahiz eta erabiltzen dituzten interpretazioen errepresentazioak gureak baino sakonagoak izan —egitura lexiko-kontzeptualak erabiltzen baitituzte—, gurea haietara iristeko pauso bat da, eta gerora haienekin bateragarriak izango direnak erabili nahi izan ditugu.

Interpretazioen desanbiguazioari dagokionez, kontuan izan behar dugu egindako esperimentuaren arabera, postposizio-atzizki baten interpretazio-kopurua erdira jaisten dela. Egindako esperimentuan, euskarazko postposizio-atzizkiak zazpi interpretazio ditu, eta beste hizkuntzetako preposizioekin parekatu ondoren, interpretazio-kopurua 3,1era jaitsi da. Era berean, lortu dugun doitasun maila %94,5koa izan da.

Atal honekin hasi ginenean, postposizio-atzizkiak ezagutzeko ez genuen inongo arazorik. Horretarako aurretik aipatu dugun *Morpheus* erabili dugu, eta postposizio-atzizki bakoitzaren informazio osoa EDBLn dugu. Arazoa sortzen zitzaigun postposizio konposatuak ezagutzeko garaian. Horiek ezagutzeko tresnarik ez genuen, eta *Murriztapen Gramatika* erabiliz, halakoak ezagutzeko gramatika garatu behar izan dugu. Gramatika berri honekin lortzen diren emaitzak honokoak dira: %90,09ko estaldura, eta %93,62ko doitasuna (Aduriz *et al*, argitaratzear).

Denera euskarazko 12 postposizio-atzizki eta 211 postposizio konposatu aztertu ditugu, eta beste hizkuntzetako datuei erreparatuz, ingeleseko 149 preposizio eta espainierako 132 ditugu gure interpretazio-zerrendaren arabera etiketatua. Postposizio eta preposizio horien zerrendak, interpretazioen araberrako adibideekin eta hizkuntzen arteko loturekin F., G. eta H. eranskinetan paratu ditugu.

Lan honen aportazio nagusietako bat euskarazko postposizioentzat, ingelesezko eta espainierazko preposizioentzat interpretazioen inbentario bateratua da. Inbentario hori izan ondoren ingelesa eta espainiera erabili ahal izan dugu euskarazko postposizioen interpretazioak desanbiguatu ahal izateko. Hiru hizkuntzetako inbentarioa ezberdina balitz, ezinezkoa litzateke.

Etorkizuneko lanak hainbat dira atal honetan. Hiztegiko definizioen analisi sintaktikoa egin ahal izateak lana erraztu egingo du, eta sortutako hitzen arteko erlazioak handitu egingo dira.

Hemen instrumentalarakin egindako azterketa jarri dugu. Emaizak oso onak izan direnez, metodo bera hiztegiko definizioetan agertzen diren postposizio guztiei aplikatu nahi diegu, eta horretarako zehaztu ditugu postposizio guztien interpretazioak.

V.8.2. atalean azaldu dugun metodo bera erabili dezakegu hemen agertzen zaizkigun hitzen adierak desanbiguatu ahal izateko: *EuskalWordNetera* lotu, alegia.

Eta, azken helburua, *ExtendedWordNet* (Harabagiu *et al.*, 1999) egiten duten moduan, definizio guztien interpretazio semantikoa lortzea izango da.

Eskerrik asko, familia. Eskerrik asko lagunak.

Zuek gabe ez zen txosten hau jaioko.

gu	A0	izord.	Pluraleko lehen pertsonaren izenordaina, hitz egiten ari dena barnean duen lagun bi edo gehiagoren multzoa adierazten duena.
bazter	A2	iz.	Aldea, alderdia, lurraldea, bereziki bakartua.

Maite ditut maite gure bazterrak lanbroak
izkutatzen dizkidanean
zer izkutatzen duen
ez didanean ikusten uzten
orduan hasten bainaiz
izkutukoa
nire baitan pizten diren bazter miresgarriak ikusten

(J. Artze)

ostiral	A0	iz.	Astearen seigarren eguna, ostegunaren eta larunbataren artekoa.
---------	----	-----	---

Azken urte honetan bi egunen arteko baina askoz gehiago izan zara niretzat.

VII. Ondorioak eta etorkizuneko lanak

Helburua hiztegi batetik abiatuz bertan dauden hitzen arteko erlazio lexiko-semanticoko guztiak erauztea izan da, ondoren *Euskal Hiztegiaren Ezagutza-Basea* (EHEB) sortuz *Euskararen Ezagutza-Basearen* (EEB) bidean lehen urrats gisa. Guk erauzten dugun informazio guztia izango da ezagutza-base horren elikagaia.

VII.1. Oinarriak eta prestakizunak

Hitzen arteko erlazio lexiko-semanticokoak ezagutzeko aztergai izan dugu Sarasolaren *Euskal Hiztegia* (EH) (Sarasola, 1996). Hiztegi hau, berez, paperezko hiztegia da, eta gure azterketarako makinak irakurtzeko moduko hiztegia behar dugunez, TEI ekimenak proposatutako gidalerroak erabiliz markatu dugu hiztegia.

Lanerako erabili ditugun eremuak ez dira sarrera-hitz baten barruan agertzen diren guztiak, baizik eta, lanerako interesgarriak direnak: *sarrera-hitza*, *kategoria*, *adiera-zenbakia* eta *definizioa*.

Izenak, *adjektiboak* eta *aditzak* aztertu ditugu. Kategoria bakoitza bere aldetik aztertu dugu definizio-patroi ezberdinak dituztelako. Baina multzoak ez dira hain zorrotzak eta bada *izen* eta *adjektibo* izan daitezkeen hitzak biltzen dituen multzo bat.

Azpisarrerak ere multzo berezitzat hartu ditugu. *Azpisarrerak* sarrera-hitz baten barruan daude, baina ohiko definizio-patroiak erabiltzen direnez, guk sarrera-hitz gisa tratatu ditugu. *Azpisarrera* kategoria gabe agertzen denean, sarrera-hitzaren kategoria esleitu diogu, eta kategoria hori ez denez beti zuzena, *azpisarrerak* multzo

gisa lantzea erabaki dugu, eta ez *azpisarrera* bakoitza dagokion kategoriaren multzoan.

Beraz, bost talde landu ditugu: *izenak*, *adjektiboak*, *aditzak*, *izenak eta adjektiboak*, eta *azpisarrerak*.

VII.2. Analizatzaile morfologikoa

Gure helburua hiztegiko hitzen erlazioak automatikoki erauztea denez, hiztegian dagoen informazio ezkutua agerian jarri behar izan dugu makinak lan egin dezan. Horretarako *Morpheus* analizatzaile morfologikoa erabili dugu. Definizio guztiak analizatu ondoren, definizioan ez ditugu hitz multzoak bakarrik izango, baizik eta *token* multzoak. *Token*-segida usuenak zeintzuk diren ikasi dugu. Segida horiek eman digute definizio-patroiei buruzko informazioa eta horien bitartez bilatu ditugu definizioetan dauden erlatore berezi eta patroiz lexikalak.

Definizioekin ez ezik, sarrera-hitzak eratorriak ote diren ezagutu ahal izateko ere analizatzaile morfologiko bera erabili dugu. Sarrera-hitza eratorria denean, definizioaren arabera zehaztu dugu oinarriarekiko duen erlazioa.

VII.3. Hitzen arteko erlazioak eta gramatika

Hitzen arteko erlazioak erauzteko *Murriztapen Gramatika* erabili dugu. Gramatika honek testuingurua hartzen du kontuan, eta, erauzi nahi ditugun erlazioetarako oso erabilgarria da. Hitz bat beste batekin erlazionatuta egongo baita inguruan dituen elementuen arabera. Lortutako erlazioak hiru multzotan —tesi-lanaren kapituluek hori erakusten dute— bildu ditugu: sarrera-hitza eta definizioko hitz baten artekoak, sarrera hitzaren barruko erlazioak, eta definizioko bi hitzen arteko erlazioak.

VII.3.1. Sarrera-hitza eta definizioko hitz baten arteko erlazioak

Lehenik, sinonimoak eta genusak erauzi ditugu. Horretarako, sinonimiaren bitartez eta genusaren bitartez definitzen diren sarrera-hitzen definizio-patroi usuenak zeintzuk diren aztertu dugu. Sinonimoak eta genusak azaleko erlazioak dira, baina horien azpian erlazio lexiko-semantiko bereziak zehaztu ditugu. Sinonimoen kasuan erlazioa *sinonimia* da, eta genusen kasuan *hiperonimia* / *hiponimia*.

Hiperonimia / hiponimia erlazioa hierarkikoa da —goitik beherakoa edo behetik gorakoa, alegia—, eta bi izenen edo bi aditzen artean topatuko dugu. Adjektiboen antolakuntza, berriz, ez da hierarkikoa, eta haien definizio-patroien artean ez dugu genusik topatuko. Adjektiboen definizioetan usuen erlature bereziak daude.

Erlature bereziak definizioaren gunean dauden elementuak dira, baina, esanahiz ahulak direnez, definizioko beste hitz bat —erlazonatutako hitza deitu duguna— jartzen dute sarrera-hitzarekin harremanetan. Beraz, definizioetan erlature berezia eta sarrera-hitzarekin erlazonatutako hitza markatu ditugu. Eta erlature berezi hauek sortzen dituzten sakoneko erlazioak, berriz, honakoak dira: *ia-sinonimia*, *antonimia*, *hiperonimia*, *ezaugarria*, *edukia*, *jasale*, *berezko ekintza*, *meronimia*, *berezko objektu*, *berezko egile*, *erlaturearen arabera*.

Zerrenda ikusi ondoren, nabaria da erlature berezien azpian hainbat erlazio lexiko-semantiko daudela eta definizio-patroien bitartez horiek desanbiguatzen ditugu, eta —ahal dela— bi hitzen artean erlazio bakarra eman.

Kategoria bakoitzerako berari dagozkion definizio-patroiak daudenez, kategoria bakoitzean erregela ezberdinak aplikatu ditugu. Izen eta adjektibo izan daitezkeen hitzak biltzen dituen multzoari bi kategoria horietako erregelak aplikatu dizkiegu, eta erregela horiek berek erabaki dute zein kategoriatakoa den lexikografoak erabili duen definizio-patroia. Hala ere, banaketa ez da erabat zorrotza, bi kategoriatakoak izan daitezkeen erregelak egon badaudelako.

Azpisarreraren kasuan, zein kategoriatakoak diren jakin ahal izateko hiru kategorietako erregelak aplikatu ditugu, eta erregelek erabaki dute azpisarrera horren kategoria zein den.

Etorkizuneko lana izango da hemen lortutako emaitza *EuskalWordNetera* lotzea.

VII.3.2. Sarrera-hitzaren barruko erlazioak

Ondoko azterketa sarrera-hitzari erreparatuz egin dugu. Sarrera-hitza eratorria denean, eratorri horren eta oinarriaren arteko erlazioa zein den zehaztu dugu.

Hiztegiko sarrera-hitz eratorriak zeintzuk diren ezagutzeko *Morpheus* analizatzaile morfologikoak berez dituen eratorpen-atzizkiak baino gehiagorekin lan egin behar izan dugu. Horretarako hiztegiko sarrera-hitzen amaiera usuenak aztertu

ditugu, eta erabaki ea eratorpen-atzizki ote diren edo ez. Azterketa horretan, eratorpen atzizkiez gain konposizioa osatzen duten hainbat elementu ere agertu zaizkigu, baina hiztegian agerpen-maila handia dutenez, ezagutzea merezi duela erabaki dugu.

Analizatzailea prestatu ondoren, atzizki horiei buruko informazioa bildu dugu. Atzizki zenbaitek adiera ezberdina duten eratorriak sortzen dituzte. Gure helburua, atzizki bakoitzak sortzen dituen adierak zehaztu ondoren, adiera batekoa edo bestekoa ote den definizioaren arabera zehaztea izan da. Adieren arabera, oinarriarekin erlazio bat edo beste duela esan dugu. Beraz, atzizki bakoitzaren adierei buruzko bilketa-lana egiteaz gain, adieren arabera, oinarriarekin duen erlazioa zein den zehaztu dugu.

Hiztegian dauden 7.264 sarrera eratorrientzat 3.348tan (%46,09) lortzen dugu eratorria eta oinarriaren arteko erlazioa markatzea, eta 2.920 kasutan (%40,20) sarrera-hitzaren sinonimoaren definizioa erabili ahalko dugu —sinonimiaren hedadura, alegia— erlazioa markatzeko.

Sinonimiaren hedadura aplikatzea etorkizuneko lana izango bada ere, lagin bat hartu eta aplikatuz nolako emaitzak lortzen ditugun aztertu dugu. Sinonimoen bitartez definitutako hitz eratorriek duten erlazioa %54,17a da. Kasu orokor honi aplikatuz gero, 2.920 horietatik laginean lortutako datuen arabera 1.581 markatuko ditugu.

Doitasunari dagokionez, atzizki lexikalaren arabera asko aldatzen da. Zenbait atzizkiekin %100eko doitasuna neurtu dugu, eta beste batzuekin %80a. Beraz, %80 eta %100 arteko doitasuna dugu.

Eratorpenaren azterketa hau egin ondoren, eratorriak eta oinarriak *EuroWordNet*ekin lotzeko ahalegina egin dugu. Horretarako *EuskalWordNet* erabili dugu. Guk erauzitako erlazioak kontzeptuen artekoak izatea lortuz gero, eta *EuroWordNet*en txertatuz gero, beste hizkuntzentzako baliagarri izatea lortuko dugu. Ahalegin honen aukera *EuskalWordNet*ekin lanean hasi eta berehala neurtu genuen, eta hona lotura egiteko garaian lortutako emaitzak: estaldurari dagokionez, eratorriak *EuskalWordNet*era lotzea %22an lortu dugu, oinarriak %9an eta biak —eratorria eta oinarria— %16an. Beraz, %53 geratu zaizkigu lotu gabe. Eta doitasunari dagokionez, kontzeptua desanbiguatzea %97an lortu dugu eratorriaren kasuan, eta

oinarriaren kasuan %80an erabat ondo, eta %15ean ondo dagoenaz gain, errorea den beste kontzeptu bat lotuz. Eratorrietan emaitza hobeak lortzen badira ere, kontuan izan behar dugu anbiguitasun maila askozaz handiagoa dela oinarrien kasuan: eratorrien artean adiera bakarrekoak %52,6 dira, eta oinarrien kasuan %7,7. Emaitza hauek oso doitasun ona adierazten dutenez, *EuskalWordNeta* garatuago dagoenean lotura berriro egingo dugu, kontzeptuen arteko erlazioak izan ditzagun.

VII.3.3. Definizio barruko hitzen arteko erlazioak

Sarrera-hitza albo batera utzita, definizio barruan dauden hitzen arteko erlazioak zehaztu ditugu. Horretarako bi modu erabili ditugu: patroi lexikalak eta postposizioen analisia.

Patroi lexikalak *token*-segiden azterketatik erauzi ditugu; EHN oso gutxi agertzen dira. Denera 6 patroi lexikal bilatu ditugu eta 718 aldiz markatu ditugu. Lortutako estaldura %93,34 da eta doitasuna %100.

Definizio barruko erlazioak aztertzeke beste modua postposizioak aztertzea izan da. Postposizio izenaren azpian postposizio-atzizkiak —kasu gramatikalak eta bestelako kasuak deitu izan direnak— eta postposizio konposatuak aztertu ditugu.

Postposizio-atzizkiei balizko hainbat interpretazio esleitu dizkiegu gramatiketan eta Aldezabalen (2004) lanean oinarrituta. Baina interpretazio horiek esleitzeaz gain, buruan izan dugu momentu oro, interpretazio horiek bateragarriak izan daitezela beste hizkuntzetako preposizioei emandako interpretazioekin. Lan hau aski konplexua izan da, preposizioei emandako interpretazio-zerrenda itxirik ez dagoelako. Beraz, euskarazko iturria hartuta, eta ingeleserako eta espainierarako dagoen preposizioen errepresentazio zabalena hartuta —Dorr-ena— tarteko bidea egin dugu horiek guztiak erabili ahal izateko. Euskarazko postposizioak eta espainierazko eta ingelesezko preposizioak elkartzeko gai izango den interpretazio-zerrenda erabili dugu.

Postposizio edo preposizio baten interpretazioa zein den zehazteko, hiztegiak erabili ditugu. Lehenengo hiztegi elebidunak erabili ditugu hitzak euskaratik espainierara eta ingelesera itzultzeko. Lortutako itzulpenak espainierazko eta ingelesezko hiztegi elebakarretan bilatu ditugu eta modu honetan testu paraleloak sortu. Metodoak emaitza zuzena ematen duenean, euskaraz eta espainieraz edo

ingelesez kontzeptu berari dagokion definizio paraleloa lortzen dugu. Eta, definizio paralelo horiekin postposizioen eta preposizioen interpretazioak desanbiguatzea lortu dugu. Testuinguru berean agertzen direnez eta kontzeptu berari egiten diotenez erreferentzia, postposizio edo preposizio horien interpretazioa hauen arteko ebakidura da. Ebakidura preposizio eta postposizio horiek amankomunean dituzten interpretazioak dira.

Denera euskarazko 12 postposizio-atzizki eta 211 postposizio konposatu aztertu ditugu. Eta beste hizkuntzei dagokienez, espainierazko 132 preposizio eta ingelesezko 149 preposizio etiketatu ditugu.

Metodoaren zuzentasuna neurtzeko euskarazko postposizio-atzizki *instrumentala* hartu dugu eta beste hizkuntzak erabiliz desanbiguazio-emaizak neurtu ditugu. Hasiera batean *instrumentalak* 7 interpretazio ditu, eta metodoa erabili ondoren, batez beste 3,1 interpretazio ditugu. Doitasuna %94,5 da. Lan hau ez dugu oraindik modu zabalean aplikatu.

VII.4. Emaizak orokorrak

Hiztegian dauden 35.549 definizioetatik %93,32k du erlazio-markaren bat. Erlazio guztiak eta lortutako emaitzak <http://ixa.si.ehu.es/Ixa/resources/EH-erlazioak> web-orrian daude.

Datuei dagokienez, hona laburpena. Ondoko taulan hiztegian dauden definizioak —*Definizioak*—, horitako zenbat markatu ditugun —*Markatuta*—, eta markatu gabe —*Markatu gabe*— zenbat dauden jarri dugu —*kopurua*—. Kopuru osoaz gain, guztiekiko ehunekoa —*%*— ere eman dugu.

	Definizioak		Markatuta		Markatu gabe	
	kopurua	%	kopurua	%	kopurua	%
Izenak	21.511	100	20.876	97,05	635	2,95
Aditzak	6.132	100	6.014	98,08	118	1,92
Adjektiboak	4.310	100	3.853	89,40	457	10,60
Izenak-adjektiboak	1.020	100	961	94,22	59	5,78
Azpisarrerak	2.576	100	1.469	57,03	1.107	42,97
Guztira	35.549	100	33.173	93,32	2.376	6,68

107. taula. Hiztegian dauden definizioak eta markatutakoen kopurua

Erlazioak markatzeko elementuak kontuan hartuta, hona elementu bakoitzarekin markatu dugun definizio-kopurua —*kop.*—, eta dauden definizio guztiekiko erlazioa —*%*—.

	Izenak		Aditzak		Adjektiboak		Izenak-adjektiboak		Azpisarrerak		Guztira	
	kop.	%	kop.	%	kop.	%	kop.	%	kop.	%	kop.	%
Sin	9.381	43,61	2.826	46,09	1.921	44,57	388	38,04	586	22,75	15.102	42,48
Genus	9.946	46,24	2.018	32,91			195	19,12	633	24,57	12.792	35,98
Erlz	4.283	19,91	1.873	30,54	2.135	49,54	592	58,04	426	16,54	9.309	26,19
Erlt	5.455	25,36	2.380	38,81	2.256	52,34	621	60,88	516	20,03	11.228	31,58
Patr	445	2,07	54	0,88	138	3,20	48	4,71	29	1,13	714	2,01
Erat	2.107	9,79	243	3,96	808	18,75	190	18,63			3.348	9,42
Guztira	20.876	97,05	6.014	98,08	3.853	89,40	961	94,22	1469	57,03	33.173	93,32

108. taula. Hiztegiko definizioak dituzten erlazioen arabera.

Hiztegian markatuta dauden erlazioak —*denera*— 52.783 dira; hau da, markaren bat duten definizioek batez besteko —*BB*— 1,59 erlazio dituzte.

	Izenak		Aditzak		Adjektiboak		Izenak-adjektiboak		Azpisarrerak		Guztira	
	denera	BB	denera	BB	denera	BB	denera	BB	denera	BB	denera	BB
Sin	11.926	1,27	3.828	1,34	2.553	1,33	442	1,14	673	1,15	19.422	1,29
Genus	11.650	1,17	2.428	1,20			212	1,09	695	1,10	14.985	1,17
Erlz	4.768	1,11	2.044	1,09	2.477	1,16	720	1,22	485	1,14	10.494	1,13
Erlt	6.768	1,24	2.729	1,15	2.878	1,28	765	1,23	589	1,14	13.729	1,22
Patr	446	1,00	54	1,00	139	1,01	50	1,04	29	1,00	718	1,01
Erat	2.574	1,22	243	1,00	898	1,11	214	1,13			3.929	1,17
Guztira	33.364	1,60	9.282	1,54	6.468	1,68	1.683	1,75	1.986	1,35	52.783	1,59

109. taula. Hiztegiko definizioetan dugun erlazio-kopurua.

Marka bakoitzaren batez bestekoa lortzeko, marka hori duten elementu-kopurua zati marka hori duten definizio-kopurua eginda lortzen da.

VII.5. Etorkizuneko lanak

Etorkizuneko lanei dagokienez, garrantzitsuena izango da azaleko sintaxia eta dependentziak erabiltzea. Bi hauek erabili ahal izateko hiztegiko definizioetan zenbait aurreprozesu egin beharko dira definizioak esaldi oso bihurtzeko, baina emaitzak hobetuko direla uste dugu. Dependentziak erabiliz, gainera, definizio barruko hitzen arteko erlazio-kopurua asko handituko da, ondoz ondoko hitzekin lan egiteari utziko diogulako.

Beste lan nagusietako bat sinonimiaren hedadura aplikatzea izango da. Lan honetan, nahiz eta aipatu, ez dugu sinonimiaren hedadura modu automatikoa aplikatu, baina eskuz egindako ikerketari esker, hedadura aplikatuz gero %54

interpretazio gehiago lortuko genituzkeela jakin dugu, eta horrek adierazten du etorkizunean aplikatu beharreko metodoa dela.

Sortzen ditugun erlazio konplexuen deskribapena fintzea eta tarteko lana egiten duten elementuak definiziotik erauzteak ere garrantzia izango du etorkizunean. Guk zehaztutako erlazioak beti bi hitzen artekoak direla esan badugu ere, zenbaitetan erlazioa konplexuagoa da. *garagardogilearen berezko objektua* garagardoa dela esaten dugunean, tartean aditz baten beharra sumatzen da, *berezko objektu* erlazioa izen eta aditz baten artekoa izan ohi delako. Nahiz eta lan honetan automatikoki ez erauzi aditz hori zein den, etorkizunean definiziotik erauzi ahal izango dugu tarteko aditz hori *egin* dela. Deskribapen konplexu hori *EuroWordNet*en txertatzea oso garrantzitsua izango da.

Eratorpenarekin egindako lana *EuroWordNet*ekin —modu zabalean— lotzen ari gara, baina postposizioekin egindako lana oraingoz ez dugu era masiboan aplikatu.

Hau izan da *Euskal Hiztegia*n dauden hitzen arteko erlazio lexiko-semanticoen ustiapena lortzeko egindako lana. Definizio-patroiak bilatu ditugu eta zerrendatu ditugu; eratorpen-atzizkiei buruzko informazioa bildu dugu eta eratorpena aztertu dugu; euskarazko postposizioen interpretazioak aztertu ditugu eta beste hizkuntzetako preposizioen parean jarri ditugu. Azterketa sakon honen ondoren, euskarazko hitzak biltzen dituen datu-basea izateaz gain, euskarazko kontzeptuak lotzen dituen ezagutza-basearen lehen oinarria sortu dugu. Dagoeneko kontzeptuak euskaraz izendatzeko erabiltzen diren hitzak ez dira isolaturik bizi, baizik eta hurbileko familiakideei loturik.

Bibliografia

- (ACL-SIGSEM, 2003) ACL-SIGSEM, 2003, *Workshop on The linguistic Dimensions of prepositions and their use in computational linguistics formalisms and applications*, IRIT, Toulouse, Frantzia. Argitaratzailea P. Saint-Dizier.
- (Aduriz *et al.*, 1998) Aduriz I., Aldezabal I., Ansa O., Artola X., Díaz de Ilarraza A., Insausti J., 1998, “EDBL: a Multi-Purposed Lexical Support for the Treatment of Basque”, *Proceedings of the First International Conference on Language Resources and Evaluation (LREC)*, Granada. II. liburukia, 821-826 or.
- (Aduriz *et al.*, 2004) Aduriz I., Aranzabe M., Arriola J., Díaz de Ilarraza A., Gojenola K., Oronoz M., Uria L., 2004, “ A Cascaded Syntactic Analyser for Basque”, *Computational Linguistics and Intelligent Text Processing*, Berlin. 124-135 or.
- (Aduriz *et al.*, forthcoming) Aduriz I., Aldezabal I., Alegria I., Aranzabe M., Arriola J.M., Atutxa A., Balza I., Ceberio C., Esparza I., Izagirre E., Lersundi M., Mendizabal K., Oronoz M., Pociello E., Quintian M., Uria L., Urizar R., argitaratzear, *Postposizio konposatuak ezagutzeko gramatika*.
- (Ageno *et al.*, 1992) Ageno A., Carcoze S., Castellón I., Martí M.A., Rigau G., Rodríguez H., Taulé M., Verdejo M.F., 1992, “Un entorno para la extracción de información semántica de diccionarios”, *Asociación española para la inteligencia artificial (AEPIA)*, Madrid.
- (Agirre *et al.*, 1998) Agirre E., Ansa O., Arregi X., Arriola J.M., Diaz de Ilarraza A., Lersundi M., Soroa A., Urizar R., 1998, “Extracción de relaciones semánticas mediante gramáticas de restricciones”, *Congreso SEPLN*, Alicante.
- (Agirre *et al.*, 2000a) Agirre E., Ansa O., Arregi X., Artola X., Díaz de Ilarraza A., Lersundi M., Martínez D., Sarasola K., Urizar R., 2000, “Extraction of semantic relations from a Basque monolingual dictionary using Constraint Grammar”, *EURALEX International Congress*, Stuttgart, Universität Stuttgart. 641-650 or.

- (Agirre *et al.*, 2000b) Agirre E., Ansa O., Aranzabe M., Arregi X., Arriola J.M., Artola X., Diaz de Ilarraza A., Lersundi M., Martinez D., Sarasola K., Soroa A., 2000, *Hiztegietan oinarritutako hitzen adiera-desanbiguazioa euskaraz*, Barne txostena: UPV/EHU/LSI/TR 10-2000.
- (Agirre *et al.*, 2001) Agirre E., Ansa O., Arregi X., Arriola J.M., Artola X., Diaz de Ilarraza A., Lersundi M., Martinez D., Sarasola K., Soroa A., Urizar R., 2001, *Euskal Hiztegiatik hitzen arteko erlazio semantikorantz: sinonimoen, genusen eta erlatore berezien bilaketa*, Barne txostena: UPV/EHU/LSI/TR 2-2001.
- (Agirre *et al.*, 2002) Agirre E., Lersundi M., Martinez D., 2002, “A Multilingual Approach to Disambiguate Prepositions and Case Suffixes”, *ACL Workshop: Word Sense Disambiguation: recent successes and future directions*, University of Pennsylvania, Philadelphia. 1-8 or.
- (Agirre *et al.*, 2003) Agirre E., Ansa O., Arregi X., Artola X., Díaz de Ilarraza A., Lersundi M., 2003, “A Conceptual Schema for a Basque Lexical-Semantic Framework”, *Conference on Computational Lexicography and Text Research (Complex)*, Budapest.
- (Agirre *et al.*, 2004) Agirre E., Atutxa A., Gojenola K., Sarasola K., 2004, “Exploring portability of syntactic information from English to Basque”, *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC)*, Lisboa, Portugal.
- (Agirre eta Lersundi, 2001) Agirre E., Lersundi M., 2001, “Extracción de relaciones léxico-semánticas a partir de palabras derivadas usando patrones de definición”, *Procesamiento del Lenguaje Natural (SEPLN)*, Jaen. 27. zenbakia, 165-172 or.
- (Agirre, 1999) Agirre, E., 1999, *Kontzeptuen arteko erlazio-izaeraren formalizazioa ontologiak erabiliaz: dentsitate kontzeptuala*, Informatika Fakultatea, UPV / EHU, Donostia. Doktorego-tesia.
- (Aldezabal *et al.*, 1999) Aldezabal I., Ansa O., Artola X., Ezeiza A., Gojenola K., Insausti J.M., Lersundi M., 1999, *Euskararen Datu-Base Lexikala (EDBL): eskema berriaren proposamena*, Barne txostena: UPV/EHU/LSI/TR9-99.
- (Aldezabal *et al.*, 2001) Aldezabal I., Ansa O., Arrieta B., Artola X., Ezeiza A., Hernández G., Lersundi M., 2001, “EDBL: a General Lexical Basis for the

- Automatic Processing of Basque”, *Proceedings of the IRCS Workshop on linguistic databases*, Philadelphia (USA).
- (Aldezabal *et al.*, 2002) Aldezabal I., Aranzabe M., Atutxa A., Lersundi M., 2002, *Levin-ek English verb classes and alternations (1993) liburuan proposatzen dituen ingeleserako alternantziak euskararekin parekatuz*, Barne txostena. UPV/EHU / LSI / TR 13-2002.
- (Aldezabal, 2004) Aldezabal, I., 2004, *Aditz-azpikategorizazioaren azterketa sintaxi partzialetik sintaxi osorako bidean. 100 aditzen azterketa, Levin-en (1993) lana oinarri hartuta eta metodo automatikoak baliatuz*, Euskal Filologia saila, UPV / EHU, Leioa.
- (Alegria *et al.*, 1996) Alegria, I., Artola, X., Sarasola, K., Urkia, M., 1996, “Automatic morphological analysis of Basque”, *Literary & Linguistic Computing*, Oxford, Oxford University Press. 11. liburukia, 4. zk., 193-203 or.
- (Alegria *et al.*, 2004) Alegria I., Ansa O., Artola X., Ezeiza N., Gojenola K., Urizar R. 2004, “Representation and Treatment of Multiword Expressions in Basque”, *ACL workshop on Multiword Expressions*, Barcelona.
- (Amsler eta White, 1979) Amsler, R.A., White J.S., 1979, *Development of a computational methodology for deriving natural language semantic structures via analysis of machine-readable dictionaries*, Technical Report MCS77-01315, Nacional Science Foundation, University of Texas, Austin.
- (Amsler, 1980) Amsler, R.A., 1980, *The structure of the Merriam-Webster pocket dictionary*, Computer Science, University of Texas, Austin. Doktorego-tesia.
- (Amsler, 1981) Amsler, R.A., 1981, “A taxonomy for English nouns and verbs”, *Proceedings of the 19th Annual Meeting of the Association for Computational linguistics (ACL)*, Stanford, California. 133-138. or.
- (Appelt *et al.*, 1993) Appelt D., Hobbs J., Bear J., Israel D., Kameyama M., Tyson M., 1993, “Fastus: a finite-state processor for information extraction from real-world text”, *Proceedings of international joint conference of Artificial Intelligence (IJCAI)*, Chambery, France. 1172-1178 or.

- (Arregi *et al.*, 2002) Arregi X., Arriola J., Artola X., Díaz de Ilarraza A., García E., Laskurain B., Sarasola K., Soroa A., Uria L., 2002, “Semiautomatic conversion of the Euskal Hiztegia Basque Dictionary to a queryable electronic form”, *T.A.L. journal “Electronic dictionaries: for humans, machines or both”*. 45-57 or.
- (Arriola *et al.*, 1996) Arriola, J.M., Artola, X., Soroa, A., 1996, “Hauta-Lanerako Euskal Hiztegiaren analisi erdiautomatikoa”, *Anuario del Seminario de Filología Vasca ‘Julio de Urquijo’*, XXX-2, 621-629 or.
- (Arriola *et al.*, 1997) Arriola, J.M., Artola, X., Gojenola, K., Soroa, A., 1997, “TEI: testu-kodeketarako gidalerroak”, *EKAIA*.
- (Arriola, 2000) Arriola J.M., 2000, Euskal Hiztegia-ren azterketa eta egituratzea ezagutza lexikalaren eskuratze automatikoari begira. Aditz-adibideen analisisa Murritzapen-gramatika baliatuz, azpikategorizazioaren bidean, EHU, tesi-lana.
- (Artiagoitia, 1995) Artiagoitia X., 1995, “-garri atzizkiaren izaera bikoitzaz. Zergatik den maitagarria bezain mingarria”, *Anuario del Seminario de Filología Vasca ‘Julio de Urquijo’*. XXIX-2, 161-198 or.
- (Artola, 1993) Artola, X., 1993, *HIZTSUA: Hiztegi-sistema urgazle adimendunaren sorkuntza eta eraikuntza / Conception et construction d’un système intelligent d’aide dictionnaire (SIAD)*, UPV-EHU, Donostia. Doktorego-tesia.
- (Atserias *et al.*, 1998) Atserias J., Castellón I, Civit M., 1998, “Syntactic parsing of unrestricted spanish text”, *Proceedings of 1st international conference on language resources and evaluation (LREC)*, Granada.
- (Azkarate, 1990a) Azkarate M., 1990, “Euskal –garri vs. latinezko –bilis: bi balio desberdin”, *Patxi Altunari omenaldia (Mundaiz-5)*, Deustuko Unibertsitatea, Donostia. 19-39 or.
- (Azkarate, 1990b) Azkarate, M., 1990, *Hitz elkartuak euskaraz*, Mundaiz, Universidad de Deusto, Donostia.
- (Azkue, 1923-1925) Azkue, R.M., 1923-1925, *Morfología Vasca*, La Gran Enciclopedia Vasca (1969), Bilbo.

- (Bateman, 1990) Bateman, J., 1990, "Upper modeling: organizing knowledge for natural language processing", *Proceedings of 5th international workshop on natural language generation*, Pittsburg, AEB.
- (Bauer, 1983) Bauer, L., 1983, *English Word Formation*, Cambridge University Press, Cambridge.
- (Bierwisch, 1989) Bierwisch M., 1989, "The semantics of gradation", *Dimensional adjectives: Grammatical structure and conceptual interpretation*, Berlin: Springer-Verlag. 51-74 or.
- (Bilgin *et al.*, 2004) Bilgin O., Çetinoglu Ö., Oflazer K., 2004, "Morphosemantic relations in and across Wordnets. A study based on Turkish, *Proceedings of the Second Global WordNet conference (GWC)*, Brno, Txekiar Errepublika. 60-66 or.
- (Binot eta Jensen, 1987) Binot J.L., Jensen K., 1987, "A semantic expert using an on-line standard dictionary", *Proceedings of the 10th international joint conference on artificial intelligence (IJCAI)*, Milan, Italia. 709-714 or.
- (Black eta El-Kateb, 2004) Black W.J., El-Kateb S., 2004, "A prototype English-Arabic dictionary based on WordNet", *Proceedings of the Second Global WordNet conference (GWC)*, Brno, Txekiar Errepublika. 67-74 or.
- (Boguraev *et al.*, 1989) Boguraev B., Byrd R.J., Klavans J.L., Neff M.S., 1989, "From structural analysis of lexical resources to semantics in a lexical knowledge base", *Proceedings of the 1st international lexical acquisition workshop*, Detroit.
- (Boguraev eta Briscoe, 1987) Boguraev B., Briscoe T., 1987, "Large lexicons for natural language processing: utilising the grammar coding system of LDOCE", *Computational Linguistics*. 13. liburukia, 3-4 zk., 203-218 or.
- (Boguraev eta Briscoe, 1989) Boguraev, B. and Briscoe, T. (argitaratzaileak), 1989, *Computational Lexicography for Natural Language Processing*, Logman, New York.
- (Boguraev, 1987) Boguraev B., 1987, "The definitional power of words", *Proceedings of the 3rd workshop on theoretical issues in natural language processing (TINLAP)*, New Mexico. 11-15 or.

- (Brants, 2000) Brants T., 2000, "TnT – A statistical part-of-speech tagger", *Proceedings of the 6th applied natural language processing conference (ANLP)*, Seattle, Washington.
- (Brill, 1992) Brill E., 1992, "A simple rule-based part of speech tagger", *Proceedings of the 3rd applied natural language processing conference (ANLP)*, Trento, Italia. 152-155 or.
- (Briscoe eta Carroll, 1993) Briscoe E., Carroll J., 1993, "Generalised probabilistic LR parsing of natural language (corpora) with unification-based grammars", *Computational Linguistics*. 19. liburukia, 1 zk., 25-60 or.
- (Brown *et al.*, 1991) Brown P., Lai J., Mercer R., 1991, "Aligning sentences in parallel corpora", *Proceedings of the 29th annual meeting of the association for computational linguistics (ACL)*, Berkeley, California. 169-176 or.
- (Bruce *et al.*, 1992) Bruce R., Wilks Y., Guthrie L., Slator B., Dunning T., 1992, "NounSense - A Disambiguated Noun Taxonomy with a Sense of Humour". *Memorandum in Computer and Cognitive Science (MCCS-92-246)*, Computing Research Laboratory, New Mexico State University.
- (Byrd, 1989) Byrd R., 1989, "Discovering relationship among word senses", *Proceedings of the 5th annual conference of the UW centre for the New OED*, Oxford. 67-79 or.
- (Calzolari, 1986) Calzolari N., 1986, "Structure and access in an automated lexicon and related issues", *Workshop automating the lexicon*, Grosseto.
- (Calzolari, 1990a) Calzolari N., 1990, "Structure and access in an automated lexicon and related issues", *Computational lexicology and lexicography, Special Issue dedicated to Bernard Quemada*, Pisa. 1. liburukia, 139-161.
- (Calzolari, 1990b) Calzolari N., 1990, *Trends in Computational Lexicography and natural language processing*, read at the X reunión anual de la sociedad española del procesamiento del lenguaje natural (SEPLN), Donostia.
- (Canesson eta St. Dizier, 2002) Canesson E., Saint-Dizier P., 2002, "Defining and representing preposition senses: a preliminary analysis", *Proceedings of the*

- workshop word sense disambiguation: recent successes and future directions*, University of Pennsylvania, Philadelphia, PA, USA. 25-31 or.
- (Carlson eta Nirenborg, 1990) Carlson L., Nirenborg S., 1990, *Word modeling for NLP*. Centre for machine translation, Carnegie Mellon University. Technical report CMU-CMT-90-121.
- (Castellón, 1992) Castellón, I., 1992, *Lexicografía Computacional: Adquisición automática de conocimiento léxico*, Universidad de Barcelona. Tesis doctoral.
- (Chaffin *et al.*, 1988) Chaffin R., Hermann D.J., Winston M. E., 1988, “An empirical taxonomy of part-whole relations: Effects of part-whole relation type on relation identification”, *LANGUAGE AND Cognitive Processes*. 3. zenbakia, 17-48 or.
- (Chodorow *et al.*, 1985) Chodorow M.S., Byrd R.J., Heidorn G.E., 1985, “Extracting semantic hierarchies from a large on-line dictionary”, *Proceedings of the 23rd Annual Meeting of Association for Computational Linguistics (ACL)*, Chicago, Illinois. 299-304 or.
- (Cliff, 1959) Cliff N., 1959, “Adverbs as multipliers”, *Psychological Review*. 66, 27-44 or.
- (Copestake, 1990) Copestake A., 1990, “An approach to building the hierarchical element of a lexical knowledge base from a machine readable dictionary”, *Proceedings of the 1st international workshop inheritance in NLP*, Tilburg, Netherlands. 19-29 or.
- (Cruse, 1986) Cruse D.A., 1986, *Lexical Semantics*, Cambridge, England: Cambridge University Press.
- (Díaz de Ilarraza *et al.*, 2000a) Díaz de Ilarraza A., Lersundi M., Mayor A., Sarasola K., 2000, “Etiquetado semiautomático del rasgo semántico de animicidad para su uso en un sistema de traducción automática”, *Procesamiento del Lenguaje Natural (SEPLN)*, Vigo. 24. liburukia, 147-152 or.
- (Díaz de Ilarraza *et. al.*, 2000b) Díaz de Ilarraza A., Lersundi M., Martínez D., Sarasola K., Urizar R., 2000, “Extraction of semantic relations from a Basque monolingual dictionary using Constraint Grammar”, *Proceedings of the 9th*

- EURALEX International Congress*, Universität Stuttgart, Stuttgart, 2. liburukia 641-650 or.
- (Dolan *et al.*, 1993) Dolan W., Vanderwende L., Richardson S., 1993, “Automatically deriving structured knowledge bases from on-line dictionaries”, *Proceedings of the 1st conference of the Pacific Association for Computational Linguistics (PACLING)*, Simon Fraser University, Vancouver, Canada.
- (Dorr eta Habash, 2002) Dorr B., Habash N., 2002, “Interlingua Approximation: A generation-heavy approach”, *AMTA-2002 Interlingua Reliability Workshop*, Tiburon, California, USA.
- (Dorr, 1993) Dorr B., 1993, *Machine Translation: A View from the Lexicon*, Cambridge, Massachusetts, MIT Press.
- (EAGLES, 1998) EAGLES, 1998, *Preliminary recommendations on subcategorisation*, <http://www.ilc.cnr.it/EAGLES96/rep2/rep2.html> (2004-11-30ean atzitua).
- (EDR, 1995) EDR, 1995, *Electronic Dictionary Technical Guide* (2nd edition), TR-O45, Japan Electronic Dictionary Research Institute, Tokyo, Japan.
- (Elhuyar, 1996) Elhuyar, 1996, *Elhuyar Hiztegia*, Elhuyar K.E., Usurbil.
- (Euskaltzaindia, 1985) Euskaltzaindia 1985, *Euskal Gramatika Lehen Urratsak I*, Bilbo.
- (Euskaltzaindia, 1987a) Euskaltzaindia, 1987, *Euskal Gramatika Lehen Urratsak I (eranskina)*, Bilbo.
- (Euskaltzaindia, 1987b) Euskaltzaindia, 1987, *Hitz-elkarketa/1*, Bilbo.
- (Euskaltzaindia, 1991) Euskaltzaindia, 1991, *Hitz-elkarketa/2*, Bilbo.
- (Euskaltzaindia, 1992) Euskaltzaindia, 1992, *Hitz-elkarketa/4*, Bilbo.
- (Euskaltzaindia, 1993) Euskaltzaindia, 1993, *Euskal Gramatika Laburra: Perpaus Bakuna*, Bilbo.
- (Fellbaum eta Miller, 1990) Fellbaum C., Miller G.A., 1990, “Folk psychology or semantic entailment?. A reply to Rips and Conrad”, *The psychological review*. 97, 565-570 or.

- (Fellbaum eta Miller, 2003) Fellbaum C., Miller G.A., 2003, "Morphosemantic Links in WordNet", *Traitment Automatique de Langues*. 44. liburukia, 2. zenbakia, 69-80 or.
- (Fellbaum, 1999) Fellbaum C., 1999, "Semantics via Conceptual and Lexical Relations", *Bradth an Depth of Semantic Lexicons*, Kluwer Academic Publishers, Netherlands. 247-262 or.
- (Fillmore, 1968) Fillmore C.J., 1968, "The case for case", *Universals in Linguistic Theory*, New York. 1-88 or.
- (Fontenelle *et al.*, 1994) Fontenelle T., Adiaens G., De Braekeleer G., 1994, "The Lexical Unit in the Metal MT System", *Machine Translation*, Kluwer Academic Publishers. 9. liburukia, 1-19 or.
- (Fox *et al.*, 1988) Fox E., Nutter T., Ahlswede T., Evens M., Marcowitz J., 1988, "Building a large thesaurus for information retrieval", *Proceedings of teh 2nd conference on applied natural language processing (ANLP)*, Austin, Texas. 101-108 or.
- (Francis eta Kucera, 1982) Francis W., Kucera H., 1982, *Frequency analysis of English usage*, Houghton Mifflin Company, Boston, Massachusetts.
- (Gomez et al, 1994) Gomez F., Hull R., Segami C., 1994, "Acquiring knowledge from encyclopedic texts", *Proceedings of the 4th conference on Applied Natural Language Processing (ANLP)*, Stuttgart, Germany. 84-90 or.
- (Gràcia *et al.*, 2000) Gràcia L., Fullana O., Turon L., Suñer A., Cabré M.T., Lorente M., Bayà M.R., Bernal E., Varela S., Haouet L., Martín J., Azkarate M., Odriozola J.C., Zabala I., Perez E., 2000, *Configuración morfológica y estructura argumental: léxico y diccionario*, Euskal Herriko Unibertsitatea, Zarautz.
- (Grefenstette, 1993) Grefenstette G., 1993, "Evaluation techniques for automatic semantic extraction: comparing syntactic and window based approaches", *Proceedings of SIGLEX Workshop on Acquisition of lexical knowledge from text*, Columbus, OH.

- (Grishman *et al.*, 1994) Grishman R., Macleod C., Reysers A., 1994, "Complex syntax: building a computational lexicon", *Proceedings of the 15th Annual meeting of the Association for Computational Linguistics (COLING)*, Kyoto, Japan. 268-272 or.
- (Grishman eta Sundheim, 1996) Grishman R., Sundheim B., 1996, "Message understanding conference-6: A brief history", *Proceedings of the 16th Annual Meeting of the Association for Computational Linguistics (COLING)*, Copenhagen, Denmark. 466-470 or.
- (Gruber, 1993) Gruber T.R., 1993, "Towards principles for the design of ontologies for knowledge sharing", *Proceedings of the international workshop on formal ontology*, Padova, Italy. Also as Technical report KSL 93-04, Stanford University, USA.
- (Guarino, 1997) Guarino N., 1997, "Semantic matching: formal ontological distinctions for information organization, extraction and integration", *Information extraction*, Springer, Berlin, Germany. 139-170 or.
- (Guilbert, 1975) Guilbert L., 1975, *La créativité lexicale*, Larousse Université, Paris.
- (Habash eta Dorr, 2003) Habash N., Dorr B., 2003, "A categorial variation database for English", *Proceedings of NAACL/HLT*, Edmonton, Canada.
- (Habash, 2002) Habash N., 2002, "Generation-Heavy Hybrid Machine Translation", *INLG-02*, New York.
- (Harabagiu *et al.*, 1999) Harabagiu S.M., Miller G., Moldovan D., 1999, "WordNet 2-A Morphologically and Semantically enhanced resource", *Proceedings of SIGLEX*, University of Maryland. 1-8 or.
- (Harabagiu, 1997) Harabagiu S.M., 1997, *WordNet-Baed Inference of Textual Context, Cohesion and Coherence*, University of Southern California. Doktorego-tesia.
- (Hobbs *et al.*, 1993) Hobbs J., Appelt D., Bear J., Israel D., Kameyama M., Tyson M., 1993, "Fastus: A system for extracting information from text", *Proceedings of the ARPA workshop on human language technology*, Princeton. 133-137 or.

- (Hovy eta Knight, 1993) Hovy E., Knight K., 1993, "Motivating shared knowledge resources: an example from the pangloss collaboration", *Proceedings of the IJCAI workshop on shared knowledge*, Chambery, France.
- (Hutchins eta Somers, 1992) Hutchins W.J., Somers H.L., 1992, *An introduction to machine translation*, Academic Press, London.
- (Ide eta Veronis, 1993) Ide N., Veronis J., 1993, Extracting knowledge bases from machine-readable dictionaries: Have we wasted our time?, *Proceedings of the International Conference on Building and Sharing of Very Large-Scale Knowledge Bases*, Japan. 257-266 or.
- (Iris *et al.*, 1988) Iris M.A., Litowitz B.E., Evens M., 1988, "Problems of the part-whole relation", *Relational models of the lexicon*, New York: Cambridge University Press. 261-288 or.
- (Ithurri, 1979) Ithurri J., 1979, *Grammaire Basque*, Hordago, Donostia.
- (Izagirre, 2004) Izagirre E., 2004, *Eratorpenaren azterketa hiztegi eta corpusetan oinarrituta*, Hiztek Masterrerako egindako txostena.
- (Jackendoff, 1983) Jackendoff R.S., 1983, *Semantics and Cognition*, Cambridge, Massachusetts, The MIT Press.
- (Jackendoff, 1990) Jackendoff R.S., 1990, *Semantic Structures*, Cambridge, Massachusetts, The MIT Press.
- (Jensen eta Binot, 1988) Jensen K., Binot J., 1988, "Dictionary text entries as a source of knowledge for syntactic and other disambiguations", *Proceedings of the second conference on applied natural language processing (ANLP)*, Austin, Texas. 152-159 or.
- (Jensen eta Nilsson, 2003) Jensen P.A., Nilsson J.F., 2003, "Ontology-Based Semantic for Prepositions", *The linguistic Dimension of prepositions and their use in computational linguistics formalisms and applications*, IRIT, Toulouse, France. 193-204 or.
- (Karlsson *et al.*, 1995) Karlsson F., Voutilaiknen A., Heikkila J., Anttila A., 1995, *Constraint Grammar: Language-independent System for Parsing Unrestricted Text*, Moun-ton de Gruyter, Berlin / New York.

- (Klavans eta Tzoukermann, 1996) Klavans J.L., Tzoukermann E., 1996, "Dictionaries and Corpora: Combining Corpus and Machine-Readable Dictionary for building lexicons", *Journal of Machine Translation*. 10. liburukia (3-4), 185-218 or.
- (Knight eta Luk, 1994) Knight K., Luk S., 1994, "Building a large-scale knowledge base for machine translation", *Proceedings of the 12th American Association for artificial intelligence (AAAI)*, Seattle, Washington, USA. 1. liburukia, 773-778 or.
- (Lafitte, 1944) Lafitte P., 1944, *Grammaire Basque (Navarro-Labourdin littéraire)*. 1979: Donostia, Elkar.
- (Lascurain *et al.*, 2003) Lascurain V., Agirre E., Lersundi M., Popelínsky L., 2003, "Disambiguation of case suffixes in Basque", *TALN workshop on Automatic Treatment of minority and small languages*, Batz-Sur-Mer.
- (Lenat eta Guha, 1990) Lenat D., Guha R., 1990, *Building large knowledge-based systems: representation and inference in the Cyc Project*, Addison Weley, Reading, Massachusetts.
- (Lenat, 1995) Lenat D., 1995, "Steps to sharing knowledge", *Toward Very Large Knowledge Bases*, IOS Press.
- (Lersundi eta Agirre, 2003) Lersundi M., Agirre E., 2003, "Semantic interpretations of postpositions and prepositions: a multilingual inventory for Basque, English and Spanish", *Workshop on the linguistic dimensions of prepositions and their use in computational linguistics formalisms and applications*, Tolouse, France. 56-65 or.
- (Levi, 1978) Levi J.N., 1978, *The syntax and semantics of complex nominals*, New York: Academy Press.
- (Lewandowski, 1992) Lewandowski T., 1992, *Diccionario de Lingüística*, Ediciones Cátedra, Madrid. Itzultzaileak: M^a.L. García-Denche Navarro eta E. Bernárdez.
- (Li eta Abe, 1995) Li H., Abe N., 1995, "Generalizing case frames using a thesaurus and the MDL principle", *Proceedings of recent advances on natural language processing (RANLP)*. 239-248 or. Also in *Computational Linguistics*, 24. liburukia (2), (1998), 217-244 or.

- (Light, 1996) Light M., 1996, "Morphological Cues for Lexical Semantics", *Proceedings of the 34th Conference of Association for Computational Linguistics (ACL)*, Santa Cruz, California. 25-31 or.
- (Litkowsky, 1978) Litkowsky K.C., 1978, "Models of the semantic structure of dictionaries", *American journal of computational linguistics*. 81 mikrofita, 25-74.
- (Litkowsky, 1980) Litkowsky K.C., 1980, "Requirements of text processing lexicons", *Proceedings of the 18th Annual conference of the Asociation of Computational linguistics (ACL)*, Philadelphia, Pennsylvania. 153-154 or.
- (Lyons, 1977) Lyons J., 1977, *Semantics*, New York, Cambridge University Press.
- (Madsen *et al.*, 2001) Madsen B.N., Pedersen B.S., Thomsen H.E., 2001, "Semantic relations for Onto-Query", *Proceedings of the First International OntoQuery Workshop*, Department of business communication and information science, University of Southern Denmark – Kolding.
- (Marcus *et al.*, 1993) Marcus M., Santorini B., Marcinkiewicz M., 1993, "Building a large annotated corpus of English: The Penn Treebank", *Computational linguistics*, 19. liburukia, 2., 313-330 or.
- (Marcus *et al.*, 1994) Marcus M., Kim G., Marcinkiewicz M., MacIntyre R., Bies A., Ferguson M, Katza K., Schasberger B., 1994, "The Penn Treebank: Annotating predicate argument structure", *Proceedings of ARPA Workshop on Human language technology*, San Francisco. 10-13 or.
- (Markowitz *et al.*, 1986) Markowitz J., Ahlswede T., Evens M., 1986, "Semantically significant patterns in dictionary definitions", *Proceedings of the 24th Annual meeting ACL*, New York. 112-119 or.
- (Michiels eta Noël, 1982) Michiels A., Noël J., 1982, "Approaches to thesaurus production", *Proceedings of the Ninth International Conference on Computational Linguistics*, Amsterdam, North-Holland. 227-232 or.
- (Miller *et al.*, 1990) Miller G.A., Beckwith R., Fellbaum C., Gross D., Miller K., 1990, *Five Papers on WordNet*, Cognitibe Secience Lab, Princeton University. CSL-43.

- (Miller *et al.*, 1993) Miller G., Leacock C., Randee T., Bunker R., 1993, "A semantic concordance", *Proceedings of the 3rd DARPA workshop on human language technology*, Painsboro, New Jersey. 303-308 or.
- (Miller, 1990) Miller G.A., 1990, "Nouns in WordNet: A Lexical Inheritance System", *ler, et. al., Five Papers on WordNet*, Cognitive Science Lab, Princeton University. CSL-43.
- (Miller, 1998a) Miller G.A., 1998, "Nouns in WordNet", *WordNet, an electronic Lexical Database*, Massachusetts Institute of Technology. 23-46 or.
- (Miller, 1998b) Miller G.A., 1998, "Modifiers in WordNet", *WordNet, an electronic Lexical Database*, Massachusetts Institute of Technology. 47-68 or.
- (Mitxelena, 1977) Mitxelena L., 1977, *Fonética Histórica Vasca*, Donostia, Diputación de Guipúzcoa.
- (Mitxelena, 1986) Mitxelena L., 1986, "Eratorbidea eta deklinabidea", *Euskera XXXI*. 535-541 or.
- (Mitxelena eta Sarasola, 1987-) Mitxelena L., Sarasola I., 1987- , *Diccionario General Vasco-Orotariko Euskal Hiztegia*, Desclée de Brower-Mensajero, Bilbao.
- (Montemagni eta Vanderwende, 1992) Montemagni S., Vanderwende L., 1992, "Structural patterns vs. string patterns for extracting semantic information from dictionaries", *Proceedings of COLING*. 546-552 or.
- (Morris, 1998) Morris M., 1998, *Morris Student dictionary*, Klaudio Harluxet Fundazioa, Donostia.
- (Ng eta Lee, 1996) Ng H., Lee H., 1996, "Integrating multiple knowledge sources to disambiguate word sense: an exemplar-based approach", *Proceedings of the 34th annual meeting of the association for computational linguistics (ACL)*, Santa Cruz, California. 40-47 or.
- (Nirenburg eta Raskin, 1987) Nirenburg S., Raskin V., 1987, "The subworld concept lexicon and the lexicon management system", *International journal of computational linguistics*. 13. liburukia (3-4), 276-289 or.

- (Onyshkevych eta Nirenburg, 1994) Onyshkevych B., Nirenburg S., 1994, *The lexicon in the scheme of KBMT things*, Technical Report MCCS-94-277, Computing Research Laboratory, New Mexico State University.
- (Peñas *et al.*, 2000) Peñas A., Chugur I., Gonzalo J., Verdejo F., 2000, “Incorporación de adjetivos al WordNet Español”, *24ª Conferencia de la sociedad Española del Procesamiento del Lenguaje Natural (SEPLN)*,Vigo.
- (Penthedourakis eta Vanderwende, 1993) Penthedourakis J., Vanderwende L., 1993, *Automatically Identifying Morphological Relations in Machine-Readable Dictionaries*, Microsoft internal report MSR-TR-93-06.
- (Pereira *et al.*, 1993) Pereira F., Thishby N., Lee L., 1993, “Distributional clustering of English words”, *Proceedings of the 31st Annual meeting of the association for computational linguistics (ACL)*, Columbus, Ohio. 183-190 or.
- (Pérez Gaztelu *et al.*, 2004) Pérez Gaztelu E., Zabala I., Gràcia Ll. (argitaratzaileak), 2004, *Las Fronteras de la composición en lenguas románicas y en vasco*, Universidad de Deusto, San Sebastián.
- (Quirk eta Greenbaum, 1979) Quirk R., Greenbaum S., 1979, *A University Grammar of English*, Longman Group Limited, London (ninth impression (corrected)).
- (Resnik, 1993) Resnik P., 1993, *Selection and Information: a class-based approach to lexical relationships*, University of Pennsylvania. Doktorego-tesia.
- (Resnik, 1995) Resnik P., 1995, “Disambiguating noun groupings with respect to WordNet senses”, *Proceedings of the 3rd workshop on very large corpora*, MIT.
- (Ribas, 1995) Ribas F., *On acquiring appropriate selectional restrictions from corpora using a semantic taxonomy*, Universitat Politècnica de Catalunya, Barcelona. Doktorego-tesia.
- (Richardson, 1997) Richardson S.D., 1997, *Determining similarity and inferring relations in a Lexical Knowledge Base*, New York. Tesis doctoral.
- (Rigau, 1998) Rigau G., 1998, *Automatic Acquisition of Lexical Knowledge form MRDs*, Universitat Politècnica de Catalunya, Barcelona. Doktorego-tesia.

- (Riloff eta Shoen, 1995) Riloff E., Shoen J., “Automatically acquiring conceptual patterns without an annotated corpus”, *Proceedings of the 3rd workshop on very large corpora*, Boston, MA. 148-161 or.
- (Robert, 1990) Robert Paul (argitaratzailea), 1990, *Le Petit Robert*, Le Robert, Paris.
- (Sanchez, 1991) Sanchez A., 1991, “Informatización de diccionarios convencionales: un sistema de consulta para el “Diccionario Ideológico de la lengua española” de J. Casares, *Proceedings of the 7th annual meeting de la Sociedad Española del Procesamiento del Lenguaje natural (SEPLN)*, Valencia.
- (Santana *et al.*, 2004) Santana O., Pérez J.R., Carreras F.J., Rodríguez G., 2004, “Relaciones morfológicas prefijales del español”, *Procesamiento del lenguaje natural*. 32. zenbakia, 9-36 or.
- (Sarasola, 1996) Sarasola, I. 1996, *Euskal Hiztegia*, Gipuzkoako Kutxa, Donostia.
- (Smith eta Maxwell, 1980) Smith R.N., Maxwell E., 1980, “An English dictionary for computerised syntactic and semantic processing systems”, *Proceedings of the 36th International Conference on Computational Linguistics*. 303-322 or.
- (Somers, 1987) Somers H.L., 1987, *Valency and Case in Computational Linguistics*, Edinburgh University Press.
- (Sparck-Jones, 1964) Sparck-Jones K., 1964, *Synonymy and semantic classification*, University of Cambridge, England. Doktorego-tesia.
- (Sparck-Jones, 1986) Sparck-Jones K., 1986, *Synonymy and semantic classification*, Edinburgh, Edinburgh University Press
- (Sperberg-McQueen eta Burnard, 1994) Sperberg-McQueen C.M., Burnard L., 1994, *Guidelines for Electronic Text Encoding and Interchange (P3)*, Chicago and Oxford: Text Encoding Initiative. Testu berria: <http://etext.virginia.edu/teip4/>
- (Stockwell *et al.*, 1973) Stockwell R., Schachter P., Partee B.H., 1973, *The Major Syntactic Structures of English*, New York: Holt, Rinehart & Winston
- (Talmy, 1985) Talmy L., 1985, “Lexicalization patterns: Semantic structure in lexical forms, *Language typology and syntactic description*, Cambridge, England: Cambridge University Press. 3. liburuika, 57-149 or.

- (Tapanainen, 1996) Tapanainen P., 1996, *The Constraint Grammar Parser CG-2*, University of Helsinki Publications. 27. zenbakia.
- (Urkia, 1997) Urkia M., 1997, *Euskal Morfologiaren tratamendu informatikorantz*, Gasteiz, Euskal Herriko Unibertsitatea. Doktorego-tesia.
- (Utsuro *et al.*, 1993) Utsuro T., Matsumoto Y., Nagao M., 1993, “Verbal case frame acquisition from bilingual corpora”, *Proceedings of International joint conference of artificial intelligence (IJCAI)*, Chambery, France.
- (Van den Hurk eta Meijs, 1986) Van den Hurk I., Meijs W., 1986, “The dictionary as a corpus: analyzing LDOCE’s definition-language”, *Corpus linguistics II*. 99-125 or.
- (Vanderwende, 1995) Vanderwende L., 1995, *The analysis of noun sequences using semantic information extracted from on-line dictionaries*, Georgetown University, Washington, DC. Doktorego-tesia.
- (Viegas *et al.*, 1996) Viegas E., Onyshkevych B., Raskin V., Nirenburg S., 1996, Submit to Submitted via Submission: On Lexical Rules on Large-Scale Lexicon Acquisition”, *Proceedings of the 34th Conference of Association for Computational Linguistics (ACL)*, Santa Cruz, California. 32-39 or.
- (Villasante, 1974) Villasante, L., 1974, *Palabras vascas compuestas y derivadas*, Edit. franciscana Aránzazu, Oñate.
- (Vossen, 1998) Vossen P., 1998, “Introduction to EuroWordNet”, *Computers and the Humanities, special issue on EuroWordNet*, Kluwer Academic Publishers, Dordrecht. 32. zenbakia, 73-89 or.
- (Walker eta Amsler, 1986) Walker D., Amsler R., “The use of machine-readable dictionaries in sublanguage analysis”, *Analysing language in restricted domains*, Lawrence Erlbaum, Hillsdale, NJ. 69-84 or.
- (Wilks *et al.*, 1989) Wilks Y., Fass D., Guo C., McDonald J., Plate T., Slator B., 1989, “A Tractable Machine Dictionary as a Resource for Computational Semantics”, *Computational Lexicography for Natural Language Processing*, Longman, London. 193-228 or.

- (Wilks *et al.*, 1990) Wilks Y., Fass D., Guo C., McDonald J., Plate T., Slator B., 1990, "A tractable machine dictionary as a resource for computational semantics", *Computational Lexicography for natural language processing*, Longman, New York. 9. kapituluu, 193-228 or.
- (Wilks *et al.*, 1996) Wilks Y., Slator B., Guthrie L., 1996, *Electric words: dictionaries, computers and meanings*, The MIT Press, Cambridge, MA.
- (Winston *et al.*, 1987) Winston M.E., Chaffin R., Hermann D.J., 1987, "A taxonomy of part-whole relations", *Cognitive Science*. 11. zenbakia, 417-444 or.
- (Yarowsky, 1992) Yarowsky D., 1992, "Word sense disambiguation using statistical models of Rogets categories traiend on large corpora", *Proceedings of the 14th International conference on computational linguistics (COLING)*, Nantes, France. 454-460 or.
- (Yokoi, 1995) Yokoi T., 1995, "The impact of the EDR electronic dictionary on very lange knowledge bases", *Toward very lange knowledge bases*, IOS Press.

Gaia: tesi-defentsa egunean aulki bat hutsik geratu da.

Zein gogorra den heriok aita
derrepentean ostea
gaur nork alaitu ahalko digu
tesi-defentsa ostea
nahiz ta itxuraz une horrek dun
zoriontsuen ospea
hil denaren albistea (bis)
oraindik ere kostatzen zaigu
joan dela sinistea
zorion pixkat badugu baina
a ze zorion tristea (bis)

Koman zegola jakin nuen-ta
malkoz bete momentuan
hotzikara bat sumatu nuen
nire bihotz kixkurtuan
bizitza honi agur azkarra
egun madarikatuan
denok ginen tristetu han (bis)
ze hutsunea aurkezpen hontan
sumatu den segituan
gure malkoak eseri dira
zuk behar zenun lekuan (bis)

Eskuzabala, bromazalea
alaia une ilunean
liburuzale amorratua
letzen gau ta egunean
hortxe daukazu zure aulkia
nahiz ta egon urrunean
denok estuasunean (bis)
arreta jarri berri baitugu
aulkiaren hutsunean
ez ohartzeko utzi diguna
bihotzaren barreanean (bis)

Doinua: Ligoiz bizitza ez da izaten

