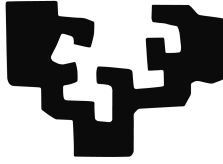


eman ta zabal zazu



EUSKAL HERRIKO UNIBERTSITATEA
Euskal Hizkuntza eta Komunikazioa Saila

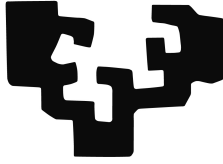
Doktorego-tesia

**EPEC corpora predikatu-mailan
etiketatzeko oinarriak:
EPEC-RolSem, BVI eta e-ROLda**

Ainara Estarrona Ibarloza

2014

eman ta zabal zazu



EUSKAL HERRIKO UNIBERTSITATEA
Euskal Hizkuntza eta Komunikazioa Saila

EPEC corpora predikatu-mailan etiketatzeko oinarriak: EPEC-RolSem, BVI eta e-ROLda

Ainara Estarrona Ibarlozak Izaskun Aldezabal Rotetaren zuzendaritzapean egindako tesiaren txostena, Euskal Herriko Unibertsitatean Euskal Filologian Doktore titulua eskuratzeko aurkeztua.

Donostia, 2014ko uztaila.

A verb is, of course, the conveyance of action. "To do". We run, jump, crawl, love. We abscond, congregate, pontificate. Action is at the core of what it means to be human. By assigning our instinctual motions a specific utterance -the runs and jumps- we give them meaning and significance. They become a part of the greater human experience. While the subject is truly the core of a sentence -without words for you or myself, or that tree, how would I tell you that we should go climb it?- the verb makes what we say come alive, in a sense. Any language is dead without it. Life is dead without it.

Joel Hans (<http://joelhans.com/a-world-without-verbs/>)

Gauzak eta ideiak lengoaiaren bidez adierazten dira eta, lengoaiari buruz gogoeta egin ez duenak ez du deusi buruz egin gogoeta.

Joseba Sarrionandia

Etrekoei

Esker onak

Tesia egitea ibilbide luze eta malkartsua izan ohi da, horregatik, bide horretan lagun izan ditudan guztiei eskerrak eman nahi nizkieke:

- Izaskuni, ezin hobeto zuzentzeaz gain, lankidearen gainera pertsona dagoela sentiarazteagatik beti. Azken finean, ibilbide aldapatsua izan ohi dena, erraz eta atsegin bihurtzeagatik. Ah! Eta idaztean enroilatzeko dudan joera errotik mozteagatik ☺
- Arantzari, zuzendari ofiziala izan ez arren, behar izan den guztietan “ordezko” zuzendari izateko prest egoteagatik.
- IXAkide guztiei, tesi-lan honetan zuek guztiek egindako lana baliatzeko aukera izan dudalako eta talde-lanaren balioa eta indarra erakutsi didazuelako. Animo bidean zaudeten guztiei, besteok lortu badugu, zuek ere lortuko duzue!
- AbarHitzekin etengabeko borrokan aditzak etiketatzen lagundu didazuenei: Izaskun, Maxux eta Izaskun.
- Tesia fintzen eta hobetzen lagundu didazuen guztiei (Larraitx, Maxux, Arantza eta Xabier), zuen zuzenketa eta iruzkinak lan hau ganoraz bukatzeko ezinbestekoak izan direlako. Eta Xabierri, bereziki, “puntalakilo” eta “txakur-belarri” bezalako hitzak erakusteagatik ☺
- Kikeri eta Estherri, makinetan erabat ezjakin den letretako honekin izan duzuen pazientziagatik, eta, batez ere, AbarHitz eta e-ROLDa tresnak garatzeko eta erabiltzeko eskainitako laguntza nekaezinagatik.

Tresnaren izena ere zuek asmatu zenuten, merci! Baina zeinen politak ziren “PatROL” eta “ROL-on”, ezta? ☺

- Manexi, lexikoia informatikoki modu txukunean jartzeko, eta hizkuntzalariek editatu ahal izateko egindako lanagatik; eta nola ez, Atlantikoaren beste aldetik azken mementoko SOS guztiei erantzun ezin hobea emateagatik. Patxaran bat (edo botila bat) nire kontu!
- Krisi-batzordeari, eta, batez ere, Arantxari, Oierri eta Aitorri, txosten hau erdi txukun geratzeko ditxosozko LaTeXekin emandako laguntzagatik; baina ez, LaTeXek ez du norberak nahi duena egiten ☺
- Bulegokideei eta tupper-txokoko lagunei, egunean behin gutxienez, lana ahazteko “kit-kat” mementoak izateko aukera emateagatik.
- Klara, Larraitz eta Arantxari, konpainia eta solasaldi ederrean emandako bazkariengatik. Ibilbide luze honetan lanaz gain, memento pertsonal zailak pasa ditudalako eta oso gertu sentitu zaituztedalako. Muak!
- Henrikeri, euskal filologia edo kazetaritza ikastea erabaki ezinik nenbilenean, euskal filologia egiteko aholkatzeagatik: “Lehenengo ikasi idazten, eta gero, izango duzu denbora kazetaritza egiteko”; eta hasi, eta hizkuntzaren birusak erabat jota, hemen geratu nintzelako.
- Aitorri, hizkuntzalaritza konputazionalaren mundua, niretzat erabat ezezaguna zen mundua, erakusteagatik eta IXA deitzen den “sekta” honetan sartzen animatzeagatik.
- Parrandetako bikotekide “esplosiboari”, behar izan zaitudanean, hurbil egon zarela beti. Azkenean “depositatu” dut!!!
- Familiari, egiten dudana ulertu ez arren, zarata handirik gabe, hor egon zarela beti. Bereziki, hain zenbaki mundukoa zen aitari, gauzak ulertzeko eta behin eta berriz hausnartzeko piztu zidan grinagatik, eta, batez ere, Araban euskaraz ikasteko aukera izan genezan egindako lan eskerga izugarriagatik. Zauden lekuan zaudela, muxu, aita.
- Sorori, djeredjef, yow mi sopi sama dundu te djangal ma guem, damala nop!!!

Guzti-guztiori, berriz ere, eskerrik asko!

Laburtzapenak

- Siglak eta akronimoak

ADESSE:	Alternancias de Diátesis y Esquemas Sintáctico-Semánticos del Español
AMR:	Abstract Meaning Representation
BDT:	Basque Dependency Treebank
BVI:	Basque Verb Index
BWN:	Basque WordNet
EADB:	Euskal Aditzen Datu Basea
EPEC:	Euskararen Prozesamendurako Erreferentzia Corputa
EusWN:	Euskal WordNet
FN:	FrameNet
HAUL:	Hitz Anitzeko Unitate Lexikala
HM:	Hautapen Murriztapena
HP:	Hizkuntzaren Prozesamendua
LIRICS:	Linguistic Infrastructure for Interoperable Resources and Systems
MG:	Murriztapen Gramatika
PB:	PropBank (Proposition Bank)
PB-VN:	PropBank-VerbNet
PDT:	Prague Dependency Treebank
SENSEM:	Sentence Semantics
SRL:	Semantic Role Labeling
UCCA:	Universal Conceptual Cognitive Annotation
UNL:	Universal Networking Language
VerbIndex:	Unified Verb Index
VN:	VerbNet

WN: WordNet

• Laburdurak¹

ABL: Ablatiboa
ABS: Absolutiboa
ABU: Adlatibo bukatuzkoa
ABZ: Adlatibo bide zuzenezkoa
ADI: Aditza
ADIZE: Aditz-izena
ADJ: Adjektiboa
ADL: Aditz laguntzailea
ADM: Aditz mota
ADOIN: Aditz-oina
ADT: Aditz trinkoa
ADV: *Adverbial*
ALA: Adlatiboa
ARR: Arrunta
ASP: Aspektua
AZP: Azpikategoria
ATZ: Atzizkia
BIZ: Biziduntasuna
BURU: Burutua
CAU: *Cause*
DAT: Datiboa
DENB: Denborazkoa
DES: Destinatiboa
DIR: *Directional*
DIS: *Discourse*
EMEN: Emendiozkoa
ERG: Ergatiboa
ERL: Erlazio-morfema
ERR: Erroa
EXT: *Extent*

¹Ingelesezoak zuzenean PropBank-etik (Palmer *et al.*, 2005) hartu ditugu.

GEL:	Genitibo leku denborazkoa
GEN:	Genitiboa
HELB:	Helburuzkoa
INE:	Inesiboa
INS:	Instrumentala
IOR:	Izenordaina
IZB:	Izen berezia
IZE:	Izena
IZGMGB:	Izenordain zehaztugabe mugagabea
KAS:	Kasua
KAT:	Kategoria
KAUS:	Kausala
KONPL:	Konpletiboa
LOC:	<i>Location</i>
M:	Mugatua
MDN:	Modalitatea eta denbora
MG:	Mugagabea
MNR:	<i>Manner</i>
MOD:	<i>Modal / Moduzkoa</i>
MOT:	Motibatiboa
MUG:	Mugatasuna
NEG:	<i>Negation</i>
NUM:	Numeroa
P:	Plurala
PAR:	Partitiboa
PART:	Partizipioa
PLU:	Pluraltasuna
PRP:	<i>Purpose</i>
PNT:	Puntukaria
PRD:	<i>Secondary Predication</i>
PRO:	Prolatiboa
REC:	<i>Reciprocal</i>
SIN:	Simplea
SOZ:	Soziatiboa
TMP:	<i>Temporal</i>

Gaien aurkibidea

Esker onak	vii
Laburtzapenak	ix
Gaien aurkibidea	xiii
Irudien zerrenda	xvii
Taulen zerrenda	xix
1 Proiektuaren nondik norakoak	1
1.1 Sarrera	1
1.2 Motibazioa	7
1.3 Helburuak	14
1.4 Tesi-txostenaren eskema	15
1.5 Argitalpenak	16
1.5.1 Tesiari hertsiki lotutako argitalpenak	16
1.5.2 Bestelako argitalpenak	18
2 Ereduaren hautapena eta deskribapena	21
2.1 Aurrekariak	21
2.1.1 Euskarari dagokionez	25
2.1.2 Beste hizkuntzetan	28
2.1.2.1 VERBNET	29
2.1.2.2 PROPBANK	36
2.1.2.3 FRAMENET	41
2.1.2.4 ADESSE	43
2.1.2.5 SENSEM	47
2.1.3 Rol-zerrenda estandarraren bila: LIRICS	50

2.2	Hautuaren arrazoiak	57
2.3	Informazio semantikoa txertatzeko etiketa: <i>arg_info</i>	58
2.4	Laburbilduz	60
3	PB-VN eredia jarraitzeko erabaki nagusiak	63
3.1	Arg0 eta Arg1 argumentuen inguruan	67
3.2	VNk bi klase edo gehiago esleitzen ditu	71
3.3	ADV rola	72
3.4	Laburbilduz	73
4	Etiketatzeko aurretikoak: baliabideak, aurreprozesuak eta tresna	75
4.1	Baliabideak	75
4.1.1	Euskal Aditzen Datu Basea (EADB)	75
4.1.2	EPEC-DEP (BDT) corpora	85
4.1.3	Levinen (1993) aditzen euskarazko itzulpena	86
4.2	Aurreprozesu automatikoak	89
4.2.1	Levinen (1993) aditzen itzulpena PBrekin gurutzatuta eta corpusarekin lotuta	89
4.2.2	EADBko 100 aditzen informazioaren esleipen automatikoa, postposizio-atzizkietan oinarrituta	93
4.3	AbarHitz: etiketatzeko tresna	95
4.4	Laburbilduz	101
5	Metodologia	103
5.1	Lehen hurbilpena	103
5.2	Metodologiaren oinarrien finkatzea	107
5.2.1	EADBko 100 aditzen zenbait agerpenen eskuzko etiketatzea eta BViren sorrera	109
5.2.2	Metodologiaren ebaluazioa	109
5.2.2.1	Adjuntuak	115
5.2.2.2	Hitz Anitzeko Unitate Lexikalak	115
5.2.2.3	Bokatiboak	116
5.2.2.4	EADB edo PB-VN jarraitu? Sarrerren zehaztapen-lana	116
5.2.3	EADBko 100 aditzen gainontzeko agerpenen etiketatze erdi-automatikoa	122

	5.2.3.1	Corpuseko konbinazio sintaktiko-semantikoak baliatuz	122
	5.2.3.2	BVI baliatuz	124
	5.2.3.3	BVI aberasten	129
	5.2.4	Levinen klaseen baliagarritasunaren azterketa	130
5.3		Behin betiko metodologia	135
5.4		Laburbilduz	138
6		EADBko 100 aditzak etiketatzean azaldutako kasuistika	143
6.1		EADB eta PB-VN ereduaren arteko adostasunak eta desadostasunak	143
	6.1.1	EADBko eta PB-VNko informazioa bat etorri da	144
	6.1.2	EADBko eta PB-VNko informazioa ez da bat etorri	145
	6.1.3	Path rola gehitu	147
6.2		Ingelesarekiko aldeak	148
	6.2.1	Euskaraz posible ez den argumentu bat agertzen da PB-VN-n	148
	6.2.2	Mugimenduzko aditzak: aditz-hizkuntzak vs sate-lite-hizkuntzak	150
	6.2.3	Euskaraz aditz bakar batekin esaten dena, ingelesez ordain desberdinekin itzultzen da	152
	6.2.4	Kausatibo/inkoatibo egiturak lexikalizatzeko bide desberdinak	155
	6.2.5	Ingelesez existitzen ez diren aditzak	160
6.3		Euskarazko aditzen inguruan	160
	6.3.1	Hitz Anitzeko Unitate Lexikaletan parte hartzeko joera handia duten aditzak	161
	6.3.2	EADBko aditzen aberastea: balio semantiko berriak	165
	6.3.3	Aditz aspektualak	172
6.4		Laburbilduz	173
7		Emaitzak	177
7.1		EPEC-RolSem: Datuak eta gaur egungo egoera	177
7.2		Basque Verb Index (BVI)	179
7.3		Laburbilduz	183
8		<i>e-ROL</i>da kontsulta-tresnaren deskribapena	185
8.1		Ezaugarri teknikoak	185

8.2	Informazio linguistikoa eta bilaketak	186
8.2.1	Bilaketa orokorrak	186
8.2.2	Bilaketa zehatzak	192
9	Ondorioak, zabaldutako ikerlerroak eta etorkizuneko lanak	197
9.1	Aditz-kopuruaren eta aditzen maiztasunen arteko harremana	198
9.2	Hitz Anitzeko Unitate Lexikaletan aditzek erakusten duten portaera berezia	199
9.3	Aditzek corpus-motaren arabera erakusten duten portaera desberdina	200
9.4	Esplizitu agertzen ez diren argumentuak eta rolak	200
9.5	Etorkizunerako beste lan batzuk	201
	Bibliografia	205
	Eranskinak	219
A	BVI lexikoia	219
B	EPEC-DEPeko aditzak eta maiztasunak	243

Irudien zerrenda

1.1	(2) adibidearen dependentzia-zuhaitza.	5
2.1	<i>spray-9.7</i> klasearen hierarkia VN-n.	30
2.2	<i>spray-9.7-1</i> azpiklaseko aditz-zerrenda eta loturak beste baliabide semantikoetara.	30
2.3	<i>spray-9.7</i> klaseari esleitzen zaizkion rol semantikoak eta hautapen-murriztapenak.	31
2.4	<i>spray-9.7</i> klaseak onar ditzakeen egitura sintaktikoak.	32
2.5	<i>spray-9.7-1</i> azpiklasearen informazioa VN-n.	33
2.6	<i>tell</i> aditzak dituen PB, VN, FN eta OntoNotes-eko loturak.	43
2.7	<i>comer</i> aditzaren sarrera ADESSEn.	45
2.8	<i>comer</i> aditzaren eskema sintaktiko-semantiko baten informazioa.	46
2.9	<i>comer</i> aditzaren sarrera SENSEMen.	48
2.10	<i>comer</i> aditzaren lehen adieran etiketatutako adibideen lagin bat.	49
2.11	<i>comer</i> aditzaren lehen adierako bi adibideren etiketatzea.	50
2.12	LIRICS eta VNko rol bateratuen hierarkia (Bonial <i>et al.</i> , 2011b).	56
4.1	AbarHitz-en lanean hasteko interfazea.	95
4.2	<i>izan</i> aditza duten fitxategien zerrenda, AbarHitz-en.	96
4.3	Etiketatzeko hasteko leihoa.	97
4.4	Aztergai den aditzari dagozkion PBko ordainen informazio-leihoa.	98
4.5	Balentziaren eremuari dagokion menua.	99
4.6	VNko rolen eremuari dagokion menua.	100
5.1	Behin betiko metodologian proposatutako urratsak.	136
8.1	BVIko sarrera, PB, FN eta EusWNeko estekak, eta EPEC-RoSem corpuseko adibideak <i>jan</i> aditzarentzat.	188
8.2	<i>jan</i> aditzaren adibide baten dependentzia-harremanak eta etiketa semantikoak.	190

8.3	<i>eman</i> aditzaren 1. adiera.	191
8.4	<i>call.01</i> PB-VNko ordaina eskatuta, <i>e-ROldak</i> ematen duen in- formazioa.	192
8.5	<i>eman</i> aditzaren 1. adierako datibodun adibideak.	193
8.6	<i>hil</i> aditzean Arg0 argumentua esplizitu duten adibideak.	194
8.7	<i>ekarri</i> aditzean Arg1 argumentua eta ABSa duten adibideak.	195
8.8	<i>ekarri</i> aditzean <i>Theme</i> rola ERGan duten adibideak.	196
9.1	<i>The long tail</i> fenomenoaren irudikapena.	199

Taulen zerrenda

1.1	<i>gustatu_like.01</i> adiera BVIn.	9
1.2	<i>itxi_close.01</i> adiera BVIn.	10
1.3	<i>sinatu_sign.01</i> adiera BVIn.	11
1.4	Kapitulu bakoitzarekin lotutako argitalpenak.	18
2.1	Rol semantikoak proiektu desberdinetan.	54
2.2	VN-n egindako rol-aldaketak konparaketa egin ondoren (Bonial <i>et al.</i> , 2011a).	54
3.1	<i>forget.01</i> adiera PBn.	63
3.2	<i>hear.01</i> adiera PBn.	65
3.3	<i>listen.01</i> adiera PBn.	65
3.4	<i>come.01</i> adiera PBn.	69
3.5	<i>separate.01</i> adiera PBn.	71
3.6	<i>banandu_1</i> adiera BVIn.	72
4.1	Levin (1993) eta euskal aditzen arteko loturaren hainbat adibide.	88
4.2	Levinen (1993) klaseen eta 2012ko PB-VN klaseen arteko kon- paraketa.	91
5.1	Metodologiaren lehen hurbilpeneko pausoen laburpena.	104
5.2	<i>say.01</i> , <i>tell.01</i> eta <i>call.01</i> adieren sarrerak PB-VN-n	105
5.3	Metodologiaren oinarriak finkatzen.	108
5.4	<i>adierazi</i> aditzaren EADBko informazioa, dagozkion PB-VN or- dainen informazioarekin alderatuta.	111
5.5	<i>izan</i> aditzaren EADBko informazioa, dagozkion PB-VN ordai- nen informazioarekin alderatuta.	111
5.6	<i>etorri</i> aditzaren EADBko informazioa, dagozkion PB-VN or- dainen informazioarekin alderatuta.	112
5.7	Etiketatzailen arteko adostasun-maila adiera aukeratzean.	113

5.8	Cohen's Kappa neurriaren emaitzak adiera kontuan hartuta. . .	113
5.9	Cohen's Kappa neurriaren emaitzak PBko ordaina (adiera) eta balentzia kontuan hartuta.	114
5.10	Cohen's Kappa neurriaren emaitzak PBko ordaina (adiera), balentzia eta rola kontuan hartuta.	114
5.11	Cohen's Kappa neurriaren emaitzak, sarrerak zehaztu ondoren.	118
5.12	Cohen's Kappa neurriaren emaitzak, balentzia ere kontuan hartuta.	119
5.13	Cohen's Kappa neurriaren emaitzak, balentzia eta rola ere kontuan hartuta.	119
5.14	Cohen's Kappa neurriaren emaitzak PropBank-en.	120
5.15	Frantsesezko corpusean aditzak eta argumentuak aukeratzean etiketatzaileen artean egon den adostasunaren <i>F-measure</i> neurria erabilita.	120
5.16	<i>aldatu_alter.01/change.01</i> adieraren konbinazio sintaktiko-semantikoak.	122
5.17	<i>aldatu_change.01/alter.01</i> adieran postposizio-atzizki bakoitza zein rolekin eta zein portzentajetan agertu den.	123
5.18	<i>joan_go.01</i> adiera BVIn.	124
5.19	<i>ikasi_learn.01/study.01</i> adiera BVIn.	125
5.20	<i>izan</i> aditzaren sarrera BVIn.	125
5.21	<i>eskatu</i> aditzaren lehenengo adiera BVIn.	126
5.22	<i>egin</i> aditzaren lehen adiera BVIn.	127
5.23	Corpusa etiketatzean <i>egin</i> aditzarentzat aurkitutako 1. adiera berria.	128
5.24	Corpusa etiketatzean <i>egin</i> aditzarentzat aurkitutako 2. adiera berria.	128
5.25	<i>hasi</i> aditzaren sarrera BVIn.	129
5.26	Levinen (1993) klase bera partekatzen duten etiketatu gabeko eta etiketatutako aditzen zerrenda.	131
5.27	<i>kritikatu</i> adieraren sarrera BVIn.	132
6.1	PB-VNk eta EADBk <i>alter.01/change.01</i> eta <i>aldatu_1</i> adierentzat ematen duten informazioa.	144
6.2	PB-VNk eta EADBk <i>replace.01</i> eta <i>aldatu_2</i> adierentzat ematen duten informazioa.	145
6.3	PB-VNk eta EADBk <i>obtain.01</i> eta <i>lortu_1</i> adierentzat ematen duten informazioa.	146

6.4	<i>say.01</i> adiera PB-VN-n.	147
6.5	<i>come.01</i> adiera PB-VN-n.	148
6.6	<i>etorri_1</i> adiera BVI lexikoian.	149
6.7	<i>go.01</i> adiera PB-VN-n.	149
6.8	<i>joan</i> aditzaren sarrera BVIn.	150
6.9	<i>atera_1</i> adieraren sarrera BVI lexikoian.	152
6.10	<i>hasi</i> aditzaren sarrera BVI lexikoian.	160
6.11	<i>bet.01</i> adiera PBn.	167
6.12	<i>jokaturen</i> hirugarren adiera BVIn.	167
6.13	PB-VNk eta BVIk <i>overlook.02</i> eta <i>eman_3</i> adierentzat ematen duten informazioa.	168
6.14	<i>egonen</i> hirugarren adiera BVIn.	169
6.15	<i>eramanen</i> bosgarren adiera BVIn.	170
7.1	Etiketatzeko prozesuko urratsetan lortutako emaitzak.	179
7.2	“Postposizio-atzizki+rol” pareen agerpen-kopurua EPEC-Rol- Sem corpusean.	181
7.3	<i>ekarri</i> aditzaren sarrera BVIn.	183
7.4	<i>jan</i> aditzaren sarrera BVIn.	183
8.1	<i>jan</i> aditzaren adibideen zerrenda.	189
B.1	EPEC-DEP corpuseko aditzen maiztasunak.	277

Proiektuaren nondik norakoak

1.1 Sarrera

Tesi-lan hau Hizkuntzaren Prozesamenduaren alorrean kokatzen da (hemen-dik aurrera HP), Euskal Herriko Unibertsitateko Informatika Fakultateko IXA taldearen² jardunaren barruan, eta, zehazki, baliabideak sortzeko atalean. Azkenaldian hain modan dagoen *corpus based* edo *corpus driven linguistic* delakoaren barnean sartzten da bete-betean eta zehatzago, corpusetan oinarritutako lexikografian.

Aspaldiko aldarrria izan da HPan lan egiteko erreferentziazko corpus baten beharra (Sagarna, 2007; Urkia, 2008; Aldezabal *et al.*, 2009b). Baina zer da zehazki corpus bat eta zer esan nahi du erreferentziazkoa izateak? Corpus bat testu-bilduma bat da eta gure kasuan, HPan alegia, euskarri elektronikoan dagoena eta hizkuntzaren azterketara eta hizkuntza-teknologiak garatzera bideratuta dagoena. Erreferentziazkoa izateak, berriz, hizkuntzaren (bere osotasunean hartuta) erakusgarri izatea esan nahi du, hau da, hizkuntzaren aldaera esanguratsuenak erakusteko behar besteko tamaina eta kalitatea izatea. Aldaera esanguratsuenak jasotzeko, corpusak orotariko testuez osatua behar du izan; tamainari dagokionez, esan ohi da milioi bat hitzeko corpusa dela azterketetarako gutxienekoa, eta kalitateaz den bezainbestean, testuek, helburuen arabera etiketatuta egon behar dute, eta, horretarako, corpusa alde zuzenetik konputazionalki prestatu beharra dago³.

Hori guztia kontuan hartuta, IXA taldean euskararen prozesamendu-

²<http://ixa.si.ehu.es/Ixa>

³Xehetasunetarako jo bedi Aldezabal *et al.*era (2009b).

rako erreferentziazko corpora eraikitzen ari gara azken urteotan: EPEC-*Euskararen Prozesamendurako Erreferentzia Corpora* (Aduriz *et al.*, 2006). EPEC corpora euskara estandarrean idatzitako 300.000 hitzek osatzen dute. Heren bat UZEI Terminologia eta Lexikografia zentroak⁴ egindako *XX. mendeko euskararen corpus estatistikotik*⁵ hartu da, eta beste bi herenak *Euskaldunon Egunkariatik*⁶.

Tamainari dagokionez oso handia ez bada ere, kalitateari dagokionez oso corpus aberatsa da, etengabe ari garelako maila linguistiko desberdinen araberatik etiketatzen. Oraindaino, corpus osoa, eskuz, hiru mailatan etiketatu dugu:

1. Segmentazio-mailan (EPEC-SEG (Aldezabal *et al.*, 2007b)). Segmentazio-mailan etiketatzea testu hitza dagokion lema, kategoria, azpikategoria, postposizio-atzizki eta numero informazioaz hornitzea da. Hots, testuko hitz ororentzat ahal dela lema eta analisi morfologiko egokia ematea. Etiketatze-maila hau IXA taldean garatutako EULIA (Artola *et al.*, 2004) tresna baliatuta egin zen. Tresna automatikoak testuko hitzaren analisi posible guztiak eskaintzen ditu⁷ eta eskuz erabaki behar da analisi egokia zein den. (1) adibidean⁸ segmentazio-mailan etiketatutako esaldi bat ikus daiteke. Interpretazio posible bat baino gehiago dagoenean, berdez markatu dugu eskuz aukeratutako analisisia.

- (1) *Egiptora Axtarteren hilobia bilatzera zihoala esan zion norbaiti.*
 - a. Egiptora [Egipto⁹]
 - **KAT:** IZE¹⁰; **AZP:** LIB; **PLU:** -; **KAS:** ALA; **NUM:** S; **MUG:** M

⁴<http://www.uzei.com>

⁵Corpus honek 4.658.036 hitz ditu eta guk horietatik 48.000 hartu ditugu EPEC corpora osatzeko, zehazki 1991-1999 bitarteko testuak.

⁶1999ko urtarriletik 2000ko maiatz bitarteko ale guztiak bildu ditugu.

⁷Guk hemen segmentazioko analisiak sinplifikatu egin ditugu. Analisisien anbiguotasunari buruzko xehetasun guztiak ikusteko jo bedi Adurizen (2000) tesi-lanera.

⁸Adibideak, ez badugu kontrakorik adierazten, EPEC corpusetik atereak dira.

⁹Kortxete artean testuko hitzaren lema ematen da.

¹⁰Laburduren adierak ezagutzeko jo bedi Laburtzapenak ataleko Laburdurak azpiatale-
ra.

- b. Axtarteren [Axtarte]
- **KAT:** IZE; **AZP:** IZB; **PLU:** -; **KAS:** GEN; **NUM:** S; **MUG:** M
- c. hilobia [hilobi]
- **KAT:** IZE; **AZP:** ARR; **BIZ:** -; **KAS:** ABS; **NUM:** S; **MUG:** M
- d. bilatzera [bilatu]
- **KAT:** ADI; **AZP:** SIN; **ADM:** ADIZE; **ERL:** HELB; **KAS:** ALA; **MUG:** M
 - **KAT:** ADI; **AZP:** SIN; **KAT:** ATZ; **KAS:** ALA; **NUM:** S; **MUG:** M
- e. zihoala [joan]
- **KAT:** ADT; **ASP:** PNT; **ERL:** MOD/DENB; **MDN:** B1¹¹; **NOR:** HURA
 - **KAT:** ADT; **ASP:** PNT; **ERL:** KONPL; **MDN:** B1; **NOR:** HURA
 - **KAT:** ADT; **ASP:** PNT; **ERL:** KONPL; **MDN:** B5B; **NOR:** HURA
 - **KAT:** ADT; **ASP:** PNT; **ERL:** MOD/DENB; **MDN:** B5B; **NOR:** HURA
- f. esan [esan]
- **KAT:** IZE; **AZP:** ARR
 - **KAT:** IZE; **AZP:** ARR; **KAS:** ABS; **MUG:** MG
 - **KAT:** ADI; **AZP:** SIN; **ADM:** PART; **ASP:** BURU
 - **KAT:** ADI; **AZP:** SIN; **ADM:** ADOIN
- g. zion
- **ERR:** *edun; **KAT:** ADL; **MDN:** B1; **NOR:** HURA; **NORK:** HARK; **NORI:** HARI
 - **ERR:** ukan; **KAT:** ADT; **MDN:** B1; **NOR:** HURA; **NORK:** HARK; **NORI:** HARI; **ASP:** PNT
- h. norbaiti [norbait]
- **KAT:** IOR; **AZP:** IZGMGB; **KAS:** DAT; **MUG:** MG

¹¹B1: Indikatibozko lehenaldia; B5B: Subjuntibozko lehenaldia.

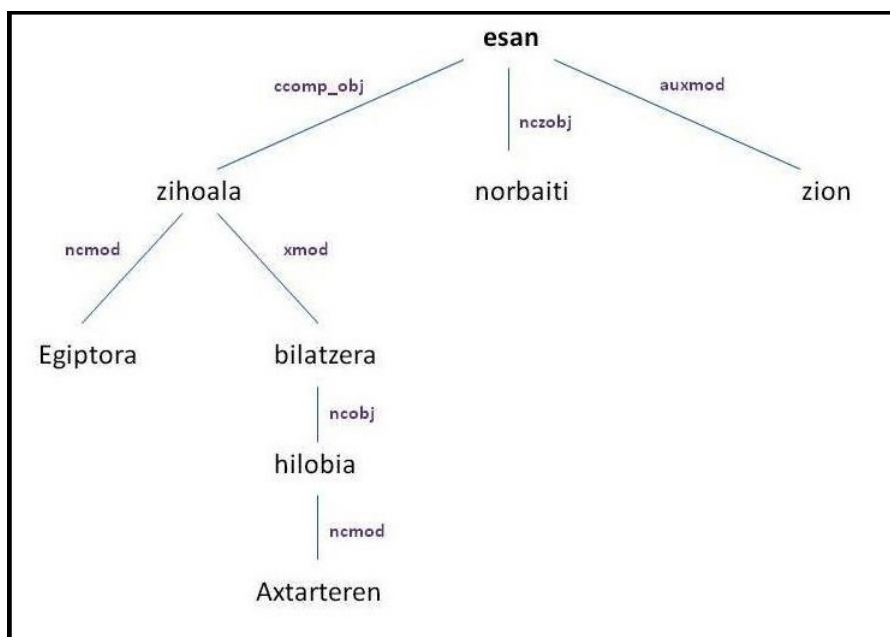
2. Sintaxi-mailan Dependentsia Gramatikaren eredu (Tesnière, 1959) jarraituz (EPEC-DEP (BDT¹²) (Aranzabe, 2008; Aldezabal *et al.*, 2009a), aditzaren mende dauden elementuak etiketatzen dira dependentsia-zuhaitza osatzeko. Aurreko adibide bera (2) jarriko dugu sintaxi-mailako dependentsia-etiketekin, eta, jarraian, adibidearen dependentsia-zuhaitza emango dugu (1.1 irudia), etiketen formatua hobeto ulertzeko:

(2) *Egiptora Axtarteren hilobia bilatzera zihoala esan zion norbaiti.*

ncmod¹³ (ala, zihoala, Egiptora, Egiptora)
ncmod (gen, hilobia, Axtarteren, Axtarteren)
ncobj (abs, bilatzera, hilobia, hilobia, obj)
xmod (helb, zihoala, bilatzera, bilatzera)
ccomp_obj (konpl, esan, zihoala, zihoala)
auxmod (-, esan, zion)
nczobj (dat, esan, norbaiti, norbaiti, zobj)

¹² *Basque Dependency Treebank*.

¹³ **ncmod**: adizlaguna (ez-perpaua); **ncobj**: objektua (ez-perpaua); **xmod**: mendeko perpaua ez-jokatua, adizlaguna edo izenlaguna; **ccomp_obj**: mendeko perpaua osagarri jokatua, objektua; **auxmod**: aditz laguntzailea eta **nczobj**: zehar objektua (ez-perpaua).



1.1 irudia: (2) adibidearen dependentzia-zuhaitza.

1.1 irudian ikusten den bezala, aztergai dugun *esan* aditz nagusiak mendeko perpaus konpletibo bat (*ccomp_obj*), zehar objektu bat (*nczobj*) eta aditz laguntzailea (*auxmod*) ditu bere mendean, eta era berean, mendeko perpaus konpletiboak beste bi elementu gobernatzen ditu: i) adizlagun bat (*ncmod*) eta ii) helburuzko mendeko perpaus ez-jokatu bat (*xmod*). Bukatzeko, helburuzko perpaus ez-jokatuaren mendean objektu zuzena (*ncobj*) eta horren modifikatzailea (*ncmod*) ikusten ditugu.

3. Semantika-mailan (EPEC-Eusemcor), Euskal WordNet-eko (EusWN) adierak erabiliz eta soilik izen arruntei gagozkiela (Pociello, 2007; Pociello *et al.*, 2011). Ikus (3) adibidea:

- (3)
 - *hilobia* [hilobi]
Hitz honi 02768416 *synseta*¹⁴ esleitu zaio eta horren azalpena hau da ingelesezko WordNet-en (WN) (Miller, 1985; Fellbaum, 1998): *a place for the burial of a corpse (especially beneath the ground and marked by a tombstone)*

¹⁴WordNet-eko adierei *synset* esaten zaie.

(EusWNko itzulpena: gorpu bat lurperatzeko lekua (beretziki lur azpian eta hilarri batekin markatuta)).

Hain zuzen gure helburua orain, semantika-mailarekin jarraituz, baina aditzak aztergai hartuta, corpusari beste etiketatze-geruza bat gehitzea da aditzak aztergai hartuta. Zehazki, argumentu/adjuntu hautagai diren elementuen gainean, **predikatu-mailako etiketatzea** egingo dugu, hau da, aditzen adierak, balentziak, rol semantikoak eta, zenbaitetan, hautapen-murriztapenak (HM) etiketatuko ditugu. Predikatu-mailan etiketatutako corpus horri EPEC-RolSem deituko diogu. Ondoren, corpus etiketatua oinarri hartuta, aditzen lexikoa osatuko dugu (tesi-lan honen helburuen azalpen zabalagoa [1.3](#) atalean).

Adibide berarekin jarraituz, (4) adibidean ikusten den bezala, dependenzia-etiketa bakoitzaren ondoren predikatu-mailako etiketa (*arg_info*) ezarriko dugu, adibidean berdez markatuta (etiketa honi buruzko xehetasunak [2.3](#) atalean).

- (4) *Egiptora Axtarteren hilobia bilatzera zihoala esan zion norbaiti.*

ncmod (ala, zihoala, **Egiptora**, Egiptora)
arg_info (go.01, zihoala, Egiptora, Arg2, Destination, helburuko kokapena, -)
ncmod (gen, hilobia, Axtarteren, Axtarteren)
ncobj (abs, bilatzera, **hilobia**, hilobia, obj)
arg_info (look.05, bilatzera, hilobia, Arg1, Theme, gaia, -)
xmod (helb, zihoala, **bilatzera**, bilatzera)
arg_info (go.01, zihoala, bilatzera, ArgM, PRP, -, -)
ccomp_obj (konpl, esan, **zihoala**, zihoala)
arg_info (say.01/tell.01, esan, zihoala, Arg1, Topic, gaia, -konkr)
auxmod (-, esan, zion)
nczobj (dat, esan, **norbaiti**, norbaiti, zobj)
arg_info (say.01/tell.01, esan, norbaiti, Arg2, Recipient, -, -)

[1.2](#) atalean zehatzago ikusiko dugun moduan, mota honetako baliabide semantikoak ezinbestekoak dira HPan ikertzen diren hainbat atazatan.

1.2 Motibazioa

Jakina da perpausaren gunea aditza dela. Aditzak du perpausa gauzatzeko informazio nagusia, eta, beraz, perpausa bere osotasunean ulertzen saiatzeko aditz horrek bere inguruan behar dituen elementuak zeintzuk eta nolakoak diren jakitea beharrezkoa da. Baina gakoa da, hain zuzen, nola erabaki zeintzuk diren elementu horiek eta nola karakterizatu. Horregatik, aditzak beti izan dira aztergai interesgarria hizkuntzalarientzat.

Orain artean, sintaxi hutsa baliatuta aztertu izan dira aditzak. Euskararen kasuan, esaterako, hartzen duten aditz laguntzailearen arabera sailkatu izan dira (Oregi, 1975; Euskaltzaindia, 1987; Elhuyar, 2006). Baina hori nahikoa al da aditzak gordetzen duen informazio guztia aztertzeko? Esaterako, *atera* aditza NOR-NORK sailean sailkatzen da, baina baditugu *atera DA*, *atera DU* eta *atera ZAIO* ere. Zein motatakoa da orduan? Laguntzaile bakoitzeko aditz “desberdin” bat dugu ala egitura erlazionatuak? Argi geratzen da aditz laguntzailean oinarritzen den sailkapen klasiko hau herren geratzen zaigula aditzaren izaera osoa aztertu nahi dugunean.

Bestetik, badira sintaxi hutsa nahikoa ez dela iradokitzen duten hainbat fenomeno, Euskaltzaindiak berak NOR-NORK saileko salbuespen gisa aipatzen dituenak (Euskaltzaindia, 1987: 48-49):

- Objektu berezien gaia: instrumentalean gauzatu arren, objektu edo *Theme* diren elementuak agertzen dira (*-z baliatu*) eta berdin inesiboan gauzatu arren, lekua ez (ez eta denbora ere), baizik gaia (*Theme*) adierazten duten elementuak (*josten ikasi* edo *futbolean jokatu*). Sintaxitik haranzko ezaugarriak dira, alegia.
- NOR sintagma, berez, “sintagma adberbiala” edo da, hau da, objektua ez den beste zerbait da (*bilerak bi ordu iraun du*).
- NOR-NORI-NORK motako aditzak izanik, objektu zuzenik, hots, NOR sintagmarik ez da ageri (*amari deitu diot telefonoz*).
- NOR-NORK motako aditza izanik, ez du objekturik edo NOR sintagmarik ageri (*urak irakin du* edo *Lehendakariak dimititu du*).
- NOR-NORK motako aditza izanik, NORK sintagma ez da agertzen (*euria hasi du*).

Lehenengo laurak, hau da, benetako NOR sintagmarik ez dutenak, **NOR izuna** duten egiturak direla dio Euskaltzaindiak (Euskaltzaindia, 1987: 48), baina besterik ez.

Horiek guztiak ikusita, aditzak erakusten duen konplexutasun guztia azaltzeko sintaxiaren mugak gainditu beharra dagoela argi geratzen da, eta gure hipotesia da semantikak horretan lagundu dezakeela. Baina nolako semantikak?

Galdera horri erantzuteko, gure ustez, ikuspegi aplikatua, hain zuzen, lagungarria ez ezik beharrezkoa da. Tesi-lan honetan hautatutako predikatu-mailako etiketatzeak, eta baita hortik sortutako aditzen lexikoiak ere, laguntza eskerga ematen dute ondoren azalduko ditugun lanetan, eta etiketatze-modua bera ere horrexegatik hautatu dugu neurri handi batean.

Hizkuntzaren prozesamenduaren esparruan murgilduz, aipatutako bi baliabide semantikoak (predikatu-mailan etiketatutako corpusa eta aditzen lexikoiak) hainbat atazatan baliatzen hasita gaude, hala nola, galderak sortzeko sistemetan (Aldabe *et al.*, 2013), itzulpen automatikoan eta desanbiguzio morfosintaktikoan (Wiechetek eta Arriola, 2011).

- Galderak sortzeko sistemetan rol semantikoek eskaintzen duten informazioa oso baliagarria izan da morfologia anbigua den zenbait kasutan. Esaterako, euskaraz, lekuzko zein denborazko osagaiak biak gauzatzen dira inesiboan (INE); beraz, sistemak ezin du jakin osagai horri buruzko galdera sortu behar duenean NON edo NOIZ galdetzaila erabili behar duen (5). Elementu hori corpusean rol semantikoaren informazioarekin etiketatuta badugu, aldiz, sistemak informazio egokia du NON ala NOIZ galdetzaila duen galdera sortzeko (6).

- (5) – *goizean* >NOIZ edo NON?
– *herriko plazan* >NOIZ edo NON?

- (6) – [*goizean*]_{TMP} *eguraldia aldatu da* >Noiz *aldatu da eguraldia*?
– *saria jaso zuen [herriko plazan]*_{LOC} >Non *jaso zuen saria*?

- Itzulpen automatikoan, nahiz eta oraindik gauzak sakonkiago aztertu behar diren, lexikoian dagoen informazioa sistemak ematen duen itzulpenaren postposizio-atzizkiak zuzentzeko baliagarria dela ikusi dugu. Kontuan izan behar dugu gure lexikoian (*Basque Verb Index* (BVI))

euskal aditzei buruzko informazio sintaktiko-semanticoa gordetzeaz gain, euskarazko aditzaren ingelesezko PropBank-eko (PB) (Palmer *et al.*, 2005) ordaina zein den ere zehazten dela, eta hori oso baliagarria dela ingelesa-euskara itzulpen-sistema batentzat¹⁵.

Demagun *I like milk*¹⁶ gisako zerbait euskaratu behar duela taldean garatzen ari garen Matxin-Eneus¹⁷ ingelesa-euskara itzulpen-sistemak. Sintaxia kontuan izanda, hor bada subjektu bat (*I*) eta objektu bat (*milk*). Ingeleseztako *to like* aditza iragankorra da eta gure sistemak erregela morfologiko eta sintaktiko orokorreari esker, badaki euskaraz aditz iragankor baten subjektua ergatiboan gauzatu behar duela eta objektua absolutiboan. Hortaz, **nik esnea gustatzen dut* moduko zerbait itzul dezake eta itzulpen hori, begibistakoa denez, ez da egokia.

BVIn dagoen informazioa (1.1 taula) baliatuta, itzulpen oker hori zuzendu egin dezakegu:

gustatu like.01		
Arg0	experiencer	- (DAT)
Arg1	stimulus	- (ABS/KONP)

1.1 taula: *gustatu_like.01* adiera BVIn.

Ikusten dugu *gustatu* aditza PBko *like.01* aditzarekin lotuta dagoela, bi argumentu dituela, argumentu horiek *Experiencer* eta *Stimulus* rol semanticok betetzen dituztela, eta datiboan (DAT) eta absolutiboan (ABS) edo konpletibo (KONPL) baten bitartez gauzatzen direla, hurrenez hurren. BViko informazio hau eskura izanda, sistemari PB erduan etiketatutako ingelesezko adibidea emanaz gero (7), berak jakingo du: i) ingelesezko *like.01* euskaraz *gustatu* dela eta ii) Arg0 rola duen argumentua datiboan gauzatu behar duela eta Arg1 duen argumentua, berriz, absolutiboan (8). Informazio hori guztia izanda¹⁸, gure sistema gai izango litzateke itzulpen zuzena emateko (9).

¹⁵Kontuan izan behar dugu PBko ordainari esker, beste zenbait baliabidetako informazioa ere badugula, hala nola, WordNet (Miller, 1985; Fellbaum, 1998) edo FrameNet-ena (Baker *et al.*, 1998), 8. kapituluaren aurkeztutako dugun *e-ROLda* tresnan ikusiko dugun bezala.

¹⁶Asmatutako adibidea da hau.

¹⁷Itzulpen-sistema hau Europar Batasuneko FP7 programako PIEF-302038 zenbakidun REA ikerketa-laguntzarekin garatu da.

¹⁸Honetaz gain, bestelako informazio linguistikoa ere izango du sistemak.

- (7) $I_{subj. Arg0} \text{ like}_{01} \text{ milk}_{obj. Arg1}$
- (8) $I_{Arg0} > niri$
 $\text{milk}_{Arg1} > esnea$
- (9) $I \text{ like milk} > Niri \text{ esnea gustatzen zait}$

Iragangaitz/iragankor kasuak ondo itzultzeko ere (10) arazoak ditu itzulpen-sistemak, eta honelakoak zuzentzeko baliagarria da BVIn dagoen informazioa.

- (10) Sarrera: *This gap eventually closes, but slowly.*
 Itzulpena: *Hutsune honek ixten du azkenean, baina astiro.*

Ingeleseko esaldia PropBank-eko rolekin etiketatuta dago¹⁹ eta *this gap* horrek Arg1 etiketa dauka. Sistemari gure lexikoian dugun informazioa gehituta (1.2 taula), Arg1 hori euskaraz absolutiboan (ABS) gauzatu behar dela jakingo du eta, beraz, hasierako itzulpena zuzendu egingo du (11).

itxi_close.01

Arg0	agent	kausa (ERG)
Arg1	patient	gai ukitua (ABS)
Arg2	instrument	- (SOZ/INS)

1.2 taula: *itxi_close.01* adiera BVIn.

- (11) *Hutsune hau ixten da azkenean, baina astiro*

Egitura pasiboekin ere zailtasunak ditu sistemak (12). Rol semantikoak etiketatuta izateak eta BVIn dagoen informazioak horrelako egiturak zuzentzen laguntzen du.

- (12) Sarrera: *Signed by 25 students*
 Itzulpena: *Sinatu 25 ikasleen*

¹⁹Esperimentu hau egiteko PropBank corpuseko lagin bat erabili da; beraz, hemen ematen ditugun ingelesezko adibideak PropBank-etik atareak dira eta euskarazko itzulpenak Matxin-ENEUS sistemak ematen dituenak dira.

Rolen etiketatzeak *Agent* edo *Experiencer* rolak ematen dizkie egitura pasiboetako agenteei eta gure lexikoian *sinatu* aditzaren agentea ergatiboan (ERG) gauzatzen dela zehazten da (1.3 taula); beraz, sistemari informazio hori emanda ondo itzuliko luke (13).

Arg0	agent	- (ERG)
Arg1	destination	- (ABS)
Arg2	theme	- (SOZ/INS)

1.3 taula: *sinatu_sign.01* adiera BVIn.

(13) *Sinatua 25 ikaslek*

- Anbiguitasun morfosintaktikoa ebazteko hainbat esperimentu egin dira taldean aditzen balentziak erabiliz (Wiechetek eta Arriola, 2011). Jakina da euskaraz *-ak* atzizkia, morfologikoki, ABS plurala edo ERG singularra izan daitekeela eta sintaktikoki, berriz, subjektua, objektua edo predikatiboa. Hori arazo bat da euskarazko analisi automatikoan, eta Murriztapen Gramatikan (MG (Karlsson, 2006)) aditzen balentzien informazioa gehituta, horrelako anbiguitasuna murriztea lortzen da. (14) adibidean²⁰ *ezpainak* predikatiboa edo objektua izan daiteke eta analizatzaile automatikoak ezin du jakin zein etiketa jarri, baina *ikusik* aditzaren balentziaren informazioa gehitzen badiogu, *ikusik* objektua eskatzen duela jakingo du eta, beraz, objektu funtzio sintaktikoa esleituko zaio horretarako sortutako erregela (15) erabilita.

²⁰Adibidea Wiechetek eta Arriolaren (2011) lanetik hartua da.

- (14) *Nekez lortu nuen zure ezpainak ikustea.*
- (15) SELECT (@OBJ) IF (0 ABS LINK 0 (@OBJ)) (0 ERG) (*1 Thcase_Abs BARRIER ADI/ADL/ADT LINK 0 Thsyn_Obj);²¹

Bestalde, predikatu-mailan etiketatutako corpora ezinbestekoa da rol semantikoen etiketatzailer automatikoa garatzeko. Corpusak automatikoki etiketatuko dituen edozein tresna trebatu beharra dago, eta horretarako ikaske-ta automatikoa erabiltzen da. Lehenik, tresnak ezkuz etiketatutako corpuseko informazioa erabilia ikasi egiten du, eta, ondoren, gai izango da ikasitakoa beste corpus batean bilatu eta etiketatzeko. IXA taldean hasiak gara rol semantikoak automatikoki etiketatzeko sailkatzailea garatzen eta horretarako ezinbestekoa izan da eskuz etiketatuko EPEC corpora erabiltzea. Euskararako garatzen ari den rol semantikoen sailkatzaileak emaitza onak lortu ditu orain arte egindako esperimentuetan: % 84,3 F₁²² PropBank-eko argumentu zenbakituak sailkatzean eta % 82,9 F₁ VerbNet-eko rolak sailkatzean (Salaberri *et al.*, 2014). Gainera, etiketatzailer automatiko horren laguntza izango dugu orain arte eskuz etiketatu ez den corpusaren % 15 etiketatzeko. Beraz, bi norabideko laguntza daukagu: i) lehenik, sailkatzaile automatikoak eskuz egindako lana erabili du tresna trebatzeko, eta ii) ondoren, eskuzko lana erraztu eta azkartzeko, sailkatzaile automatikoa erabiltzeko aukera izango dugu.

Azaldutako horietaz gain, beste hainbat aplikazio eta atazatan ere lagungarri izan daiteke corpusean eta lexikoian jaso dugun aditzari buruzko informazioa, hala nola:

- Galderak erantzuteko sistemetan. Ingeleserako egin dira jada honelako saiakera arrakastatsu batzuk (Narayanan eta Harabagiu, 2004; Shen eta Lapata, 2007) eta euskarazko sistemak hobetzeko ere aukera bat izan daiteke rol semantikoen informazioa baliatzea.

²¹Hauxe litzateke erregela horren esanahi zehatza: Hautatu objektua baldin eta helburu-hitza absolutiboan eta ergatiboan badago eta @OBJ funtzio-etiketa badu. Halaber, baldin eta, eskuinetara hainbat urratsetara (*1) ezaugarri hauek aurkitzen badira: tema absolutiboan eta objektu funtzio sintaktikoarekin; baldintza horiek egiaztatzeke ezin izango da aditz, aditz laguntzaile edo aditz trinko interpretazioa duen analisirik zeharkatu.

²²Ohiko F puntuazio orekatua (*F1 score*) doitasunaren eta estalduraren arteko batezbesteko harmonikoa da. Esan behar dugu ingeleserako dauden sailkatzaile automatikoen emaitzak % 80 inguruko F₁ lortzen dutela gaur egun.

- Ebentoak eta ebento-kateak erauzteko sistemetan. Ebentoak erauzteko NORK ZER NON eta NOIZ egin duen jakin behar dugu, eta horretarako rol semantikoak etiketatuta egotea ezinbestekoa da. Taldean NewsReader²³ proiektuaren barruan, rol semantikoen informazioa baliatuta, ebento-kateen erauzketa lantzen ari gara gaztelania eta ingeleserako, eta tesi-lan honen emaitza diren baliabideak erabilia, euskararako ere gauza bera egin ahal izango da etorkizunean.
- Korreferentzia ebazteko sistemetan. Ingeleserako hasi dira informazio semantikoak erabiltzen sistemak hobetzeko (Ponzetto eta Strube, 2006) eta euskararentzat ere egin litezke horrelako saiakerak.
- Hizkuntzaren sorkuntzarako sistemetan. Hizkuntza hutsetik sortzeko aditzaren azpikategoriazioa ezinbestekoa da.
- Sintaxi osoaren azterketan aurrera egiteko (argumentu implizituak azalertzeko, adibidez) (Hajicova, 2000; Hajicova *et al.*, 2000).

Bukatzeko, **hizkuntzalaritza teorikoaren bidetik**, predikatu-mailako informazioa gordetzen duten baliabideak (corpus etiketatua eta aditzen lexikoia) oso baliagarriak dira fenomeno linguistikoak aztertu eta hizkuntzen arteko aldeak ikertzeko. Hainbat fenomeno azter ditzakegu corpus etiketatua eta lexikoia baliatuz:

- Aditzen portaera kolokazioetan: aditzek kolokazioetan parte hartzen dutenean, euren adieren, eta ondorioz, inguruan hartzen dituzten elementuen portaera azter dezakegu.
- Argumentu elidituen edo implizituen gaia: aditz batzuek beste batzuek baino erraztasun handiagoa al dute argumentuak elidituta agertzeko?; zein argumentu edo rol semantiko agertzen da gehiagotan elidituta?; ba al da eliditua izateko aukera gehiago duen rolik?; etab. Hori dena aztertzeko, lehenik eta behin, tesi-lan honetan landu dugun moduko lexikoia beharrezkoa da, adiera jakin bateko nahitaezko argumentuak definituta dituen.
- Hizkuntzen arteko aldeak: hizkuntza bakoitzean aditzek dituzten adiera desberdinak; adierek dituzten alternantzia desberdinak; hizkuntza

²³<http://www.newsreader-project.eu/files/2012/12/NewsReader-316404-D4.2.pdf>

batean aditz batekin esaten dena bestean hainbat aditzekin esatea (ikus [6.3.1](#) atala); etab.

Laburbilduz, aditzaren inguruko informazio semantikoa ezinbestekoa da sintaxiaren mugak gaindituz, testuaren ulermen sakonagoa lortzeko. Hori dela eta, aditzari buruzko informazio hau bilduko duten baliabideak garatzea beharrezkoa da HPko hainbat atazarentzat eta sistemarentzat.

1.3 Helburuak

Motibazioari buruzko atalean azaldutako guztia kontuan hartuta, tesi-lan honen helburu nagusiak bi izan dira:

1. **EPEC corpora predikatu-mailan etiketatzea.** Predikatu-mailan etiketatuko dugula esaten dugunean adierazi nahi dugu aditzaren adierak, balentziak, rol semantikoak eta, zenbaitetan, hautapen-murriztapenak (HM) etiketatzea dela helburu. Termino horiek guztiak definitzea eta zehaztea lan zaila da eta oso gai eztabaidatua izan da hizkuntzalaritzaren historian (ikus [2.1](#) atala). Horregatik, azterketa teorikoe-tan galdu gabe, taldean dugun esperientzian eta praxian oinarrituta, aurretik definituta zegoen eredu bat hautatu nahi izan dugu horrek ere dituen zailtasunak kontuan izanda ([3.](#) eta [6.](#) kapituluetan ematen da horren berri).
2. Aditzen informazio sintaktiko-semantikoa bilduko duen **lexikoa eratzea**, hautatutako eredia oinarri hartuta.

Eskuz etiketatutako corpusetan oinarrituta lexikoiak eratzea izan da alor honetan lanean dihardutenen ildo nagusia. Horren eredu dira PropBank ([Palmer et al., 2005](#))²⁴ eta PDT ([Hajic, 1998](#)) corpusak eta corpus horietan oinarrituta sortutako lexikoiak: VerbNet ([Kipper, 2005](#)) eta Vallex ([Hajic et al., 2003](#)), hurrenez hurren. Bestelako proiektuetan ere bide berbera jarraitu da, hala nola, FrameNet ([Baker et al., 1998](#)) proiektuan; gaztelania eta katalanerako proposatutako ADESSE ([García-Miguel eta Albertuz, 2005](#)) eta SENSEM ([Castellón et al., 2006](#); [Vázquez et al., 2006](#)) proiektuetan, eta, PropBank-en eredia jarraitzen duen AnCora ([Aparicio et al., 2008](#); [Taulé](#)

²⁴PropBank PennTreebank2 ([Marcus et al., 1993](#); [Marcus, 1994](#); [Marcus et al., 1994](#)) corpusaren gainean etiketatu zen.

et al., 2008) egitasmoan ere. Guk ere ildo nagusi horri jarraitu diogu gure lanaren helburuak zehazterakoan.

Corpusa etiketatzearen eta lexikoia osatzearen helburu nagusiak erdieste-ko, tarteko bestelako helburuak ere planteatu ditugu, hala nola, **etiketatze-ko metodologia egokia garatzea** (ikusi 5. kapitulua) eta etiketatze-lana aurrera eramateko irizpide argiak bilduko dituen **eskuliburua** (*Aldezabal et al.*, 2010b) osatzea.

1.4 Tesi-txostenaren eskema

2. kapituluan, lehenik eta behin, gure lan-esparruko aurrekariak azalduko ditugu, aditzen eta predikatu-mailan etiketatutako corpusen inguruan bai euskaraz bai beste hizkuntzetan egin diren lanak aztertuz. Jarraian, orain arteko lanak kontuan izanda, corpusa etiketatze-ko eredu eta eredu hori aukeratzeko arrazoiak azalduko ditugu. Kapituluak bukatzeko, EPEC corpusean informazio semantikoa txertatzeko sortu dugun etiketa (*arg_info*) aurkeztuko dugu.

3. kapituluan, hautatutako eredu jarraitzeko irizpide nagusiak azalduko ditugu zehatz.

4. kapituluan, gure lanari ekiteko alde aurretik eskura izan ditugun baliabideak aurkeztuko ditugu. Hasteko, gure lanaren oinarri nagusietako bat izan den Euskal Aditzen Datu Baseaz (EADB) hitz egingo dugu. Ondoren, sintaxi-mailan etiketatutako EPEC-DEP (*Basque Dependency Treebank* (BDT)) corpusaz ariko gara, corpus horren gainean egin baitugu guk gure predikatu-mailako etiketatzea. Jarraian, eta ingelesarekin lotura egiteko, Levinen aditzen (*Levin, 1993*) euskarazko itzulpena nola egin zen aipatuko dugu labur. Aldez aurretiko baliabideak erabiltzeaz gain, baliabide horietako informazioarekin hainbat prozesu automatiko garatu ditugu: alde batetik, Levinen aditzen itzulpena PropBank-ekin gurutzatu eta corpusarekin lotu dugu, eta, bestetik, EADBko 100 aditzen informazioa automatikoki esleitu dugu, postposizio-atzizkietan oinarrituta. Bukatzeko, AbarHitz etiketatze-ko tresna aurkeztuko dugu.

5. kapituluan, etiketatze-ko proposatzen dugun metodologiaz jardungo gara luze. Lanean aritu ahala, metodologia fintzen eta zehazten joan gara. Jarraitutako prozesuaren berri zehatza eman ondoren, corpusa etiketatze-ko behin betiko metodologia aurkeztuko dugu.

6. kapituluan, EADBko 100 aditzak etiketatzean azalduko kasuistikaz

arituko gara. Lehenik eta behin, hautatutako ereduaren eta EADBren arteko adostasunak eta desadostasunak aipatuko ditugu. Bigarrenez, ingelesarekiko aurkitu ditugun aldeez jardungo gara, eta, bukatzeko, euskal aditzen inguruan agertu zaizkigun hainbat fenomeno aztertuko ditugu.

7. kapituluaren, tesi-lan honen emaitza nagusiak (EPEC-RolSem corpusa eta *Basque Verb Index* (BVI) lexikoia) eta egungo egoera aurkeztuko ditugu.

8. kapituluaren, corpusean eta lexikoian bildutako informazio guztia modu errazean kontsultatzeko garatu dugun *e-ROLda* kontsulta-tresna aurkeztuko dugu.

Bukatzeko, 9. kapituluaren, tesi-lanaren ondorio nagusiak bildu, eta etorkizunerako zabaldu ditugun ikerlerroak eta aurrera begira aurreikusten ditugun lanak aipatuko ditugu.

1.5 Argitalpenak

Sarrera-kapitulu honi amaiera emateko, tesigilearen argitalpenen berri eman-go dugu. Bi multzotan banatu ditugu artikuluok. Alde batetik, tesiarekin hertsiki lotutako argitalpenak zerrendatuko ditugu. Bestetik, tesiarekin lotura estua izan ez arren, Hizkuntzaren Prozesamenduaren inguruan idatziriko gainerakoak.

1.5.1 Tesiari hertsiki lotutako argitalpenak

Jarraian, tesi-lan honekin lotura zuzena duten argitalpenen zerrenda aurkeztuko dugu:

- Aldezabal I., Aranzabe M.J., Díaz de Ilarraza A. eta Estarrona A. (2010a). Building the Basque PropBank. Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odjik, Stelios Piperidis, Mike Rosner and Daniel Tapias (arg.), *Proceedings of the Seventh Conference on International Language Resources and Evaluation* (LREC 2010), 1414-1417 or., European Language Resources Association (ELRA), ISBN: 2-9517408-6-7. Valletta, Malta.
- Aldezabal I., Aranzabe M.J., Díaz de Ilarraza A., Estarrona A., Fernández K. eta Uria L. (2010b). *EPEC-RS: EPEC (Euskararen Prozesamendurako Erreferentzia Corpusa) rol semantikoekin etiketatze eskuliburua*. Barne-txostena, UPV / EHU / LSI / TR 02-2010.

- Aldezabal I., Aranzabe M.J., Díaz de Ilarraza A., Estarrona A. eta Uria L. (2010c). EusPropBank: Integrating Semantic Information in the Basque Dependency Treebank. *Lecture Notes in Computer Science (LNCS) nº 6008, Alexander Gelbukh (arg.), Computational Linguistics and Intelligent Text Processing*. 60-73 or. Springer. ISSN: 0302-9743, ISBN-10: 3-642-12115-2 Springer Berlin Heidelberg New York, ISBN-13: 978-3-642-12115-9 Springer Berlin Heidelberg New York. *11th International Conference, CICLing 2010*, Iasi, Errumania.
- Aldezabal I., Aranzabe M.J., Díaz de Ilarraza A. eta Estarrona A. (2011). Preliminary evaluation of EPEC-RolSem, a Basque corpus labelled at predicate level. *SEPLN 2011*. Nº 47 (1-9). ISSN: 1135-5948. Universidad de Huelva.
- Aldezabal I., Aranzabe M.J., Díaz de Ilarraza A. eta Estarrona A. (2013). A methodology for the semiautomatic annotation of EPEC-RolSem, a Basque corpus labeled at predicate level following the PropBank-VerbNet Model. Barne-txostena. UPV/EHU/LSITR01-2013. Donostia.
- Estarrona A., Aldezabal I., Díaz de Ilarraza A. eta Aranzabe M.J. (2014a). EPEC-RolSem: Ingelesezko PropBank-VerbNet eredura etiketatutako euskarazko corpusa. Erabakiak, egokitzapenak eta berezitasunak. Gomez, R. eta Ezeizabarrena, M.J. (arg.), *Eridenen du zerzaz kontenta. Sailkideen omenaldia Henrike Knörr irakasleari (1947-2008)*. EHUren Argitalpen Zerbitzua (argitaratzeke). Bilbo.
- Estarrona A., Aldezabal I., Díaz de Ilarraza A. eta Aranzabe M.J. (2014b). Methodology of construction of the corpus-based Basque Verb Index (BVI) Lexicon. *Language Resources and Evaluation* (bidalia).
- Estarrona, A. (2014). EPEC corpusa predikatu-mailan etiketatzeko oinarriak: EPEC-RolSem, BVI eta e-ROLda. Aduriz, I. eta Urizar, R. (arg.), *Hizkuntzalari Euskaldunen I. Topaketa. Egungo ikerlerroak..* Udako Euskal Unibertsitatea-Buruxkak (argitaratzeke).

1.4 taulan, argitalpen bakoitza zein kapitulurekin lotuta dagoen ikus daiteke:

Kapitulua	Argitalpenak
III	Aldezabal <i>et al.</i> , 2013 Estarrona <i>et al.</i> , 2014a Estarrona, A. 2014
IV	Aldezabal <i>et al.</i> , 2010a, 2010c Estarrona, A. 2014
V	Aldezabal <i>et al.</i> , 2010b, 2010c Aldezabal <i>et al.</i> , 2011 Aldezabal <i>et al.</i> , 2013 Estarrona, A. 2014
VI	Estarrona <i>et al.</i> , 2014a
VII	Estarrona <i>et al.</i> , 2014b Estarrona, A. 2014
VIII	Estarrona <i>et al.</i> , 2014b

1.4 taula: Kapitulu bakoitzarekin lotutako argitalpenak.

1.5.2 Bestelako argitalpenak

Jarraian, Hizkuntzaren Prozesamenduaren alorrean egindako ikerketei dagozkien bestelako argitalpenak zerrendatuko ditugu:

- Aldezabal I., Aranzabe M.J., Arriola J., Díaz de Ilarraza A., Estarrona A., Fernández K., Iruskieta M. eta Uria L. (2007). EPEC (Euskararen Prozesamendurako Erreferentzia Corpusa) dependentziekin etiketatzeke eskuliburua. Barne-txostena. UPV/EHU / LSI / TR 12-2007. Donostia.
- Aldezabal I., Ceberio K., Esparza I., Estarrona A., Etxeberria J., Iruskieta M., Izagirre E. eta Uria L. (2007). EPEC (Euskararen Prozesamendurako Erreferentzia Corpusa) segmentazio-mailan etiketatzeke eskuliburua. Barne-txostena. UPV/EHU / LSI / TR 11-2007.
- Ixa taldea eta Elhuyar Fundazioa. (2007). Testu-corpusak: ezaugarriak, eraketa eta tresnak. M^a Jose Arrieta Sagarna (koord.), *Hizkuntza, komunikazioaren eta teknologiaren garaietan. Herri Arduralaritzaren Euskal Erakundea (HAEE/IVAP)*, 15-28 or. Gasteiz. ISBN 978-84-7777-316-0.

- Aduriz I., Aldezabal I., Aranzabe M.J., Arriola J.M., Ceberio K., Estarrona A., Iruskietza M., Lersundi M., Pociello E., Uria L., Urizar R. eta Aldasoro E. (2008). *Euskarazko postposizio-lokuzioen tratamendu konputazionala*. Barne-txostena. Euskal Herriko Unibersitatea / Universidad del País Vasco, LSI /TR 07-2008. Donostia.
- Agirre E., Aldezabal I., Estarrona A. eta Pociello E. (2008). A methodology for the joint development of the Basque WordNet and Semcor. *Dutch SemCor workshop*. Amsterdam
- Agirre E., Casillas A., Díaz de Ilarraza A., Estarrona A., Fernández K., Gojenola K., Laparra E., Rigau G. eta Soroa A. (2009). The KYOTO Project. *Procesamiento del Lenguaje Natural*, 43, 389-390. ISSN edición impresa: 1135-5948. ISSN edición digital: 1989-7553.
- Aldezabal I., Aranzabe M.J., Díaz de Ilarraza A., Estarrona A., Ezeizaga N. eta Uria L. (2009). Corpusen etiketatze linguistikoa. Ricardo Etxepare, Ricardo Gómez, Joseba A. Lakarra (arg.), *A Festschrift for Bernard Oyharçabal*. Volumen especial del *Anuario del Seminario de Filología Vasca Julio de Urquijo (ASJU)*, XLIII:1-2(2009), 37-50 or., Euskal Herriko Unibertsitateko Argitalpen Zerbitzua/Servicio Editorial de la Universidad del País Vasco. ISBN: 0582-6152.
- Arriola J.M., Aduriz I., Aldezabal A., Aranzabe M.J., Ceberio K., Estarrona A., Iruskietza M., Lersundi M., Pociello E., Uria L. eta Urizar R. (2013). Reusing the CG-2 Grammar for Processing Basque Complex Postpositions. Alberto Díaz, Iñaki Alegria y Julio Villena (arg.), in *Actas del XXIX Congreso de la Sociedad Española del Procesamiento del Lenguaje Natural (SEPLN 2013)*, 20-27 or., Madrid (España). ISBN (edición digital): 978-84-695-8349-4.

Ereduaren hautapena eta deskribapena

Kapitulu honetan, corpora etiketatzeko eredia nola eta zergatik aukeratu dugun azalduko dugu, eta eredu horren deskribapen zehatza egingo dugu. Hasteko, gure lanaren aurrekariak aztertuko ditugu, aurretik zegoena ondo ezagutzeko eta hautua egin ahal izateko informazio guztiaren berri izateko. Ondoren, [2.2](#) atalean, eredia aukeratzeko hainbat arrazoi aipatuko ditugu, eta, bukatzeko, [2.3](#) atalean, informazio semantikoa txertatzeko sortu dugun etiketa azalduko dugu.

2.1 Aurrekariak

Urteetan eztabaidatu izan da aditzak lantzeko garaian adierazgarri diren ezaugarrien inguruan. Hizkuntzalaritza teorikoan eta gramatika sortzailearen hastapenetan batez ere ([Chomsky, 1957](#)), aditzak sintaxiaren barruan aztertzen ziren. Lexikoak ez zuen garrantzi handirik eta ez zen atal autonomotzat jotzen; hitzen gordailu huts jotzen zen eta sintaxiko erregeletan zegoen definituta. Ondorengo lanetan, lexikoa bere lekua hartzen hasi zen, baina betiere sintaxia jomugan ([Chomsky, 1965](#)). Interes behinena sakoneko egituratik azaleko egiturara iristeko mekanismo sintaktikoek hartzen zuten, eta eztabaida edo azterketa guztiak gauzatze-modu horren harian sortzen ziren. Lexikoan aditzen argumentu-egitura definitutzat ematen zen (aditz batek eskatzen dituen argumentuen zehaztapena), eta horren inguruko ez-

tabaidak batez ere kasu gramatikalekin gauzatzen ziren argumentuen rolei buruzkoak ziren.

Testuinguru honetan bi ikuspegi nagusi bereizten dira aditzen rolak teorikoki definitzean:

1. Fillmore-ren kasuen teoriatik (Fillmore, 1968) eratortzen den rolen zerrenda klasikoa, Gruber, Jackendoff eta Grimshaw-en tipologietan (Gruber, 1965; Jackendoff, 1972; Grimshaw, 1979) oinarritzen dena. Rol semantiko hauek ez daude erabat finkatuta, baina oro har zerrenda bateratu hau²⁵ onartzen da:
 - **Agentea** edo **egilea**: predikatuak adierazten duen ekintza nahita eragiten eta gauzatzen duena.
 - **Kausa**: predikatuak adierazten duen ekintzaren eragilea. Oro har, goiko egileari kontrajarria, ekintza nahigabe, berez, gertatzen dela adierazten du.
 - **Gaia**: predikatuak adierazten duen ekintzaren ondorioz ukitua edo eragina gertatzen den entitatea, bai egoeraz aldatzen delako, bai kokapena aldatzen duelako.
 - **Esperimentatzailea**: predikatuak adierazten duen egoera (psikologikoa) jasaten duen entitatea edo egoera psikologiko baten hartzailea.
 - **Onuraduna**: predikatuak adierazten duen ekintzaren onura lortzen duena.
 - **Helburua**: predikatuak adierazten duen ekintza zuzentzen zaion entitatea.
 - **Iturria**: predikatuak adierazten duen mugimenduzko ekintzaren iturburua edo jatorria.
 - **Lokazioa** edo **kokapena**: predikatuak adierazten duen ekintza edo egoera non kokatzen den adierazten duena.

Rolak hierarkikoki antolatzen dira eta hierarkia horri jarraituz esleitzen zaizkio aditz bakoitzari dagozkion rolak. Hierarkia definitzerakoan, bada, desberdintasunak autoreen artean, baina denetan agertzen da ideia

²⁵Ez da zerrenda itxia eta, askotan, beharren arabera sortzen dira rol berriak (adibidez, jabea, tresna, etab.). Definizioak Fernández eta Anularen (1995) eta Gràcia *et al.* en (2000) liburuetatik hartu ditugu.

berdina: *egile* rol semantikoa duen argumentua gauzatuko da subjektu gisa; *egile* rola ez badago, hurrengo hautagaia *experimentatzailea* izango da eta horrela, hierarkiaren rolik baxuenera iritsi arte. Guk hemen, adibide gisa (16), Speas-ek proposatzen duen hierarkia erakusten dugu (Speas, 1990):

(16) egilea < experimentatzailea < gaia < helburua / iturria / lokazioa < modua / denbora

(17) adibidean Fillmore-k berak emandako adibideak ikus ditzakegu (Fillmore, 1968). Lehen esan dugun bezala, aurkeztutako rolen zerrenda ez da itxia eta beharren arabera rol berriak gehi daitezke; kasu honetan, *agentea* (*Agent*) eta *gaia* (*Patient*) rolez gain, *tresna* (*Instrument*) rola ere ikus daiteke (17c) adibidean:

(17) a) [John]_{agent} broke [the window]_{patient}
 b) [A hammer]_{agent} broke [the window]_{patient}
 c) [John]_{agent} broke [the window]_{patient} [with a hammer]_{instr.}

2. Dowtyren (1991) proto-rolak (*proto-agent* eta *proto-patient*). Dowtyrentzat ezinezkoa da predikatuen argumentu guztiak deskribatuko dituen rol semantikoaren zerrenda osatua egitea, eta horregatik proposatzen ditu *proto-rol* deitzen dituen rol semantikoak. Rolak aditzak argumentuei ezartzen dizkien ezaugarrien arabera desberdintzen dira, ez dira berez eta aldez aurretik desberdintzen, ez dago aldez aurretiko zerrendarik, baizik eta aditzak baldintzatzen ditu argumentuei esleituko zaizkien rolak. Gauzak horrela, elkarren kontrakoak diren bi rol baino ez dira behar: *proto-agent* eta *proto-patient*²⁶. Proto-rol hauek “kontzeptu-multzoak” dira, eta “kontzeptu-multzo” horiek ezaugarri semantiko batzuen arabera definitzen dira. Jarraian *proto-agent* eta *proto-patient* rolen ezaugarri horiek aipatuko ditugu (Levin, 2009)²⁷:

²⁶Van Valinek (1993) gauza bera adierazteko *actor* eta *undergoer* terminoak erabiltzen ditu.

²⁷Sareko helbide honetatik hartua (2013-12-12):
<http://www.stanford.edu/~bclevin/lisa09semDETproto.pdf>

- *Proto-agent*
 - Borondatezko parte-hartzea
 - Kontzientzia (edo pertzepzioa)
 - Ebento bat eragitea edo beste parte-hartzaile baten egoera aldatzea
 - Mugimendua (beste parte-hartzaile batekikoa)
- *Proto-patient*
 - Egoera-aldaketa jasaten du
 - Gai gehigarria
 - Kausaren eragina jasaten du
 - Proto-agentearen mugimenduarekiko egonkorra; alegia, ez da mugitzen.

Predikatuak eskatzen dituen ezaugarri semantikoen arabera, *proto-agent* eta *proto-patient* rolek subjektua eta objektua aukeratuko dituzte. Beraz, rol bakoitzaren ezaugarri semantiko horiek subjektutasunaren edo objektutasunaren ezaugarri gisa azaltzen dira. Kasu batzuetan, subjektuak eta objektuak proto-rolen ezaugarri guztiak (edo ia guztiak) biltzen dituzte (18)²⁸.

(18) *Brutusek Zesar hil zuen*

(18)-an *Brutus* subjektua da eta *proto-agent* rolaren ezaugarri guztiak biltzen ditu, hots, borondatez hartzen du parte hiltzeko ekintzan, egindakoaren kontzientzia du, beste parte-hartzailearen (*Zesar*) egoera aldatzen du eta mugitu egiten da. *Zesar*, aldiz, objektua da eta *proto-patient* rolaren ezaugarri gehienak betetzen ditu. Hala, egoera-aldaketa jasaten duen parte-hartzailea da, kausaren eragina berak jasaten du eta beste parte-hartzailearen (*Brutus*) mugimenduarekiko egonkorra da.

Beste kasu batzuetan, ordea, subjektuak edo objektuak proto-rolen ezaugarrietako bat baino ez du betetzen (19).

(19) *Mikelek Miren ikusi du*

²⁸Adibide hauek ere Levinen lanetik (Levin, 2009) atera ditugu eta euskarara moldatu (<http://www.stanford.edu/~bclevin/lisa09sem-detproto.pdf> (2013-12-12)).

(19)-an subjektuak (*Mikel*) *proto-agent* rolari dagozkion ezaugarri semantikoen artean, pertzepzioarena baino ez duela dirudi. Bestalde, (20) adibidean, objektuak (*akatsa*), *proto-patient* rolaren ezaugarri semantikoen artean, egoera-aldaketa baino ez du erakusten.

(20) *Mikelek egindako akatsa ezabatu du.*

Eztabaidak eztabaida²⁹, eta hizkuntza baten tratamendu automatikoa jomugan, hizkuntza berorren aditzak landuta izatea beharrezkoa da edozein azterketari ekin aurretik. Behar horren aurrean, lehenik euskaraz, eta gero beste hizkuntzetan zer egin den ikusiko dugu.

2.1.1 Euskarari dagokionez

Euskararen kasuan ezer gutxi aztertu izan da lexikoiaeren eta rol semantikoaren inguruan, eta ez da inolako saiakerarik egin corpus bat maila honetan etiketatzeko. Egindako azterketa gehienak aditzen sailkapenaren inguruan egin dira, baina batez ere, hartzen duten aditz laguntzaileari begira. Oregik (1975), esaterako, aditz laguntzaileek adierazten dituzten NOR, NORK eta NORI marketan oinarritzea proposatzen du:

- NOR-(NORI) (*izan, etorri, ibili, joan, egon, etzan...*)
- NOR-(NORI)-NORK (*ukan, ekarri, eraman, erabili, egin, ikusi...*)
- NOR-NORI (*izeki, jarraitu, atriki...*)
- NORK-NOR (HURA izuna) (*izeki, jarraitu, irakin, iraun, jardun, iruditu...*)
- NORK-NORI (HURA izuna)-NOR (*jarraitu, eutsi, ekin, iritzi, eragon, inotsi, erion...*)

Euskaltzaindiaren gramatikan (Euskaltzaindia, 1987) ere, hartzen duten aditz laguntzailearen arabera sailkatzen dituzte aditzak:

²⁹Eztabaida teoriko horien inguruko zehaztasun gehiago nahi izanez gero, jo bedi Aldezabalen tesira (Aldezabal, 2004).

“[...] aditz oinak beste modu batera ere sailka daitezke. Aditz bakoitzak onartzen dituen kasu gramatikaletan (hau da, **nor**, **nork** eta **nori** kasuetan) oinarritzen da beste sailkapen erizpide hau, aditzak erabakitzen baitu perpaus batean **nork** edo **nori** sintagmarik agertu behar duen ala ager daitekeen. Adibidez, **etorr(i)** aditzak ez du **nork** sintagmarik onartzen. **Nori** sintagma, berriz, onartzen du, nahiz ez beharrezkoa izan:

- **berandu etorri da**
- **berandu etorri zaigu**

Era berean, **jakin** aditzak beti **nork** sintagma eskatuko du, baina ez du **nori** sintagmarik onartuko:

- **aitak egia jakin du**
- ***aitak egia jakin dio**

Horrela, **nor** sintagmaren bat beti izango dugunez euskal perpausa osatzerakoan (ikus, dena den, **nor** izunaz dioguna aurrerago), aditz oinen sailkapen nagusia honako hau dela esan dezakegu:

- **NOR motakoak / NOR-NORK motakoak**” ([Euskaltzaindia, 1987: 47-48](#)).

[Etxeparek \(2003\)](#) semantika ere jorratzen du, batetik, aditzak adiera orokorren arabera sailkatzen dituenean:

- Mugimenduzkoak (*joan, etorri, igo, jaitsi...*)
- Existentzia adierazten dutenak (*izan, egon*)
- Eguraldi-aditzak (*berotu, hoztu, ilundu, egin...*)
- Egoera-aldaketa aditzak (*jaio, sortu, bihurtu, jarri...*)

eta, bestetik, aditz-klase bakoitzak hartzen dituen rolei buruz hitz eginen duenean, nahiz eta rol batzuei bakarrik egin erreferentzia (*agent, theme/patient* eta *beneficiary*). Baina, azken buruan, bere helburu nagusia euskararen aditz laguntzailearen aukeraketa-sistema azaltzea da eta ez aditzen egitura sintaktiko-semantikoak azaltzea lexikoia osatzeko asmoz.

Artiagoitiak (2002) rol tematikoei buruz hitz egiten du (*agentea, jasalea, (onuraduna)/kaltetua, bitartekoa, proposizioa...*), baina arreta nagusia sintaxian jartzen du, eta ez du azaltzen aditz batean rol bat edo bestea agertzea zerk baldintzatzen duen; hau da, egitura mailako printzipioak aztertzen dira, baina ez da azaltzen aditz baten lexikoko informazioa zein den.

Hiztegietan ere bada aditzei buruzko informazioa, baina adieren eta horietako bakoitzean aditzak hartzen duen aditz laguntzailearen ingurukoa da, gehienbat; ez da azaltzen aditzaren argumentu-egiturari buruzko informaziorik (argumentu-kopurua), ez eta argumentuek jasotzen dituzten rol semantikoei buruzkorik ere. Ikusi, esaterako, Elhuyar Hiztegiko³⁰ (21) adibidea:

(21) **sartu**

1. da/du ad. **entrar, meter(se), introducir**. *Zinemara sartu eta berehala ikusi zintudan*: entré en el cine y te vi en seguida. *Musikari-talde batean sartu naiz*: me he metido en un grupo de músicos. *Musika belarrietan sartua daukagu*: tenemos la música metida en los oídos. *Erdaratik sartu diren hitzak*: las palabras que han entrado del castellano. *Zerbait buruan sartzen zaizunean*: cuando se te mete algo en la cabeza. *Fraide sartu zen*: se metió fraile. *Politikan sartzen dena*: el que se mete en política. *Belar ondu guztia ganbaran sartu dute*: han metido todo el heno en el desván.
2. da/du ad. **incluir; insertar, adjuntar**. *Prezio horretan ez da aseguruia sartzen*: el seguro no está incluido en el precio.
3. da ad. [eguzkia] **ponerse**. *Eguzkia sartu da*: se ha puesto el sol.
4. da ad. Inform./Telekom. **acceder**. *Sisteman sartu*: acceder al sistema.
5. du ad. **meter, introducir, encajar, hundir**. *Labana sartu dio*: le ha metido el cuchillo. *Uretan sartu zuen*: lo metió en el agua. *Ardatza zuloan sartu*: encajar el eje en el agujero.
6. du ad. [dirua] **meter, invertir**. *Diru asko sartu dute jatetxe berri horretan*: han invertido mucho dinero en ese restaurante.
7. du ad. [dirua] **meter, ingresar**. *Zenbat diru sartu duzu kontuan?*: ¿cuánto dinero has ingresado en la cuenta?

³⁰<http://www.elhuyar.org/hizkuntza-zerbitzuak/EN/Dictionary-search> (2013-12-12).

8. du ad. **plantar**. *Pinua sartu behar dugu*: tenemos que plantar pinos.
9. dio ad. **engañar, meter una mentira/trola**. *Gezur handia sartu zion*: le metió una buena mentira. *Niri ez didak sartuko*: a mí no me engañas.

(21)-ean ikusten den bezala, adiera bakoitzean ematen dituzten adibideak oso argigarriak dira, baina ez dago inolako sailkapen semantikorik.

Badira postposizio-atzizki jakin batzuei buruzko azterketak (Ibarretxe-Antuñano, 2001, 2003), baina gauza zehatzagoak dira (lekuzko postposizio-atzizkiak lantzen dira) eta ez dago aditzen lexikoari buruzko azterketa orokorrik. Ez da aztertu aditzen argumentu horiek nola bereizten dituzten adierak edo bereizten ote dituzten, eta oraindano ez dago euskal aditzen hiztegi kontsultagarririk. Gure ahalegina izan da, dakigula behintzat, alor honetan euskaraz egin den lehen saiakera.

Orain artean, aditzen azpikategorizazioaren eta rol semantikoaren inguruan euskaraz egin den azterketarik orokorra Aldezabalek, Levenen (1993) lanean oinarrituta, bere tesian (Aldezabal, 2004) egindakoa da (aparteko puntua eskaini diogu 4.1.1 atalean), eta horregatik hartu dugu lan hori gure lanaren oinarri gisa. Gure helburua corpusa predikatu-mailan etiketatzea da, eta horretarako aditzaren inguruko informazio sintaktiko-semantiko orokorra behar dugu. Orain arteko lanek, ez eta hiztegiek ere, ez digute informazio hori eskaintzen; beraz, Aldezabalek egindako euskarazko 100 aditzen azterketatik abiatuko gara corpusa etiketatu eta lexikoa osatzeko.

2.1.2 Beste hizkuntza batzuetan

Aditzen azterketaren eta etiketatzearen inguruan hainbat hurbilpen proposatu dira, batez ere ingelesez eta gero baita beste hizkuntza batzuetan ere, hala nola, VerbNet (Kipper, 2005), PropBank (Palmer *et al.*, 2005) eta FrameNet (Baker *et al.*, 1998) ingeleserako (gero beste hizkuntza batzuetara ere hedatu dira baliabide horiek), eta gugandik hurbilago, ADESSE (García-Miguel eta Albertuz, 2005) gaztelaniarako eta SENSEM (Castellón *et al.*, 2006; Vázquez *et al.*, 2006) gaztelania eta katalanerako. Jarraian, baliabide horien aurkezpena egingo dugu, gure lanaren abiapuntua eta testuingurua zein den ulertzeko.

2.1.2.1 VERBNET

VerbNet (VN) proiektuaren egile nagusiak Martha Palmer eta Karin Kipper dira (Kipper *et al.*, 2000; Kipper, 2005; Kipper *et al.*, 2008). VerbNet-en web-orrian³¹ eta sarean eskuragarri dauden gidalerroetan³² aurkituko ditugu ereduaren xehetasunak.

VerbNet ingelesezko aditzen lexikoia da, 5.800 aditz inguru³³ jasotzen ditu eta Levinen klaseen arabera (Levin, 1993) antolatuta dago. Horrela izanda, aditz bakoitzari Levin-en klase semantikoetako bat esleitzen zaio (9.1: *put verbs*, 9.2: *verbs of putting in spatial configuration*, 10.1: *remove verbs...*). Klase horiek zehaztu eta findu egin dituzte klase bereko aditzen arteko koherentzia semantiko-sintaktikoa lortzeko asmoz; adibidez, Levinen 37.1 klasean (*verbs of transfer of a message*) hiru azpisail egin dituzte: i) 37.1.1 (*transfer_message*); ii) 37.1.2 (*inquire*) eta iii) 37.1.3 (*interrogate*). Beraz, oinarrian Levin (1993) dute, baina beren moldaketak ere egin dituzte.

Levinen teoriaren ildotik, aditz-klase bakoitzak hainbat egitura sintaktiko (*frame*) izan ditzake, hau da, argumentu-egituraren azaleko hainbat gauzatze sintaktiko edo diatesi-alternantzia (kontzeptu horren gaineko azalpen gehiago 4.1.1 atalean). VNk egiten duena da aditz-klase bakoitzak izan ditzakeen gauzatze sintaktiko horiek dagokien informazio semantikoarekin osatu. Gaur egun, 270 klase inguru desberdintzen dituzte, horien artean klase nagusiak (*Top Class*) 100 dira³⁴.

VNko klase bakoitzean hainbat osagai zehazten dira:

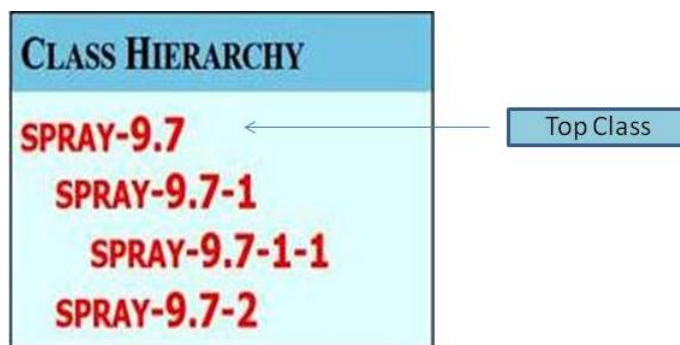
- **Klase-hierarkia** (*Class Hierarchy*): klase horretako aditzen arteko hierarkia zehazten da zuhaitz-egitura baten bidez. Zuhaitzaren goialdean *Top Class* deitzen dutena dago, eta horren azpian klase nagusiaren azpiklaseak adierazten dira (2.1 irudia):

³¹<http://verbs.colorado.edu/~mpalmer/projects/verbnet.html>

³²http://verbs.colorado.edu/verb-index/VerbNet_Guidelines.pdf

³³Etengabe ari direnez VN osatzen, egunetik egunera datuak aldatu egiten dira. Mementoko datu zehatzak ikusi nahi izanez gero, jo bedi VNren web-orrira.

³⁴Lehenengo klase nagusiak 9 zenbakia du eta azkenak 109; 9tik 57ra doazen klaseak Levinen (1993) lanetik hartuak dira zuzenean.



2.1 irudia: *spray-9.7* klasearen hierarkia VN-n.

- **Klase horretako kideak** (*Members*): klase edo azpiklase horretako kide diren aditzen zerrenda agertzen da eta, gainera, aditz bakoitzaren ondoren, FrameNet (Baker *et al.*, 1998), WordNet (Miller, 1985; Fellbaum, 1998) eta OntoNotes (Pradhan *et al.*, 2007) baliabideetara dituzten loturak azaltzen dira (2.2 irudia)³⁵:

MEMBERS		
BASTE (WN 1; G 1)	SEED (WN 4; G 3)	SPLATTER (FN 1; WN 1, 2)
BRUSH (FN 1, 2; WN 6; G 2)	SEW (FN 1; WN 1; G 1)	SPRAY (FN 1, 2; WN 1, 2, 3; G 1)
DRIZZLE (FN 1, 2; WN 2)	SHOWER (FN 1, 2; WN 1, 2, 5; G 1, 2)	SPREAD (FN 1, 2; WN 3, 9, 10; G 2, 3)
HANG (FN 1, 2; WN 2, 12, 14; G 1)	SMEAR (FN 1, 2; WN 3, 2; G 1)	SPRINKLE (FN 1, 2; WN 1, 4; G 1)
PLASTER (FN 1, 2; WN 2, 3, 4, 1, 5, 6)	SMUDGE (WN 1)	SPRITZ (WN 1, 2)
PUMP (FN 1, 2; WN 2, 4, 5; G 2)	SOW (FN 1, 2; WN 1, 3; G 1)	SPURT (WN 1)
RUB (FN 1, 2; WN 1; G 1)	SPATTER (FN 1; WN 1, 3)	SQUIRT (FN 1, 2, 3; WN 1, 2; G 1)
SCATTER (FN 1, 2; WN 3, 4, 6; G 1)	SPLASH (FN 1, 2, 3, 4; WN 3, 6; G 1)	STICK (FN 1, 2, 3, 4, 5; WN 1, 12, 13; G 1, 2)

2.2 irudia: *spray-9.7-1* azpiklaseko aditz-zerrenda eta loturak beste baliabide semantikoetara.

- **Rol semantikoak** (*Roles*): klase horretako aditzek euren argumetuekin duten harreman semantikoa adierazten duten rol semantikoen zerrenda agertzen da (2.3 irudia):

³⁵WordNet-en eta OntoNotes-en ez da aditzen egitura sintaktiko-semantikoari buruzko informaziorik ematen, eta, beraz, ez ditugu lan honetan zehatz azalduko. Batetik, WordNet teoria psikolinguistikoetan oinarrituta dagoen ingelesezko ezagutza-base lexikala da, hau da, kontzeptuen artean hainbat motatako harreman semantikoak (hiperonimia, hiponimia, sinonimia...) ezarritu egiten den sare semantikoa. Bestetik, OntoNotes-en aditzen adierak zehazten dira WNko adieren nolabaiteko multzokatzea eginez, eta adiera bakoitzak PB, VN, FN eta WNrekin dituen loturak zehazten dira.

ROLES	REF
<ul style="list-style-type: none"> • AGENT [+ANIMATE] • THEME • DESTINATION [+LOCATION & -REGION] 	

2.3 irudia: *spray-9.7* klaseari esleitzen zaizkion rol semantikoak eta hautapen-murriztapenak.

- **Hautapen-murriztapenak** (*Selectional Restrictions*): zenbaitetan, rol semantiko bakoitzari hautapen-murriztapenak (*+animate/-animate, +organization/-organization, +communication/-communication, +concrete/-concrete, +location/-location, +region/-region, +animal/-animal...*) zehazten zaizkio, eta argumentuak onartzen dituen rol semantikoak mugatzeko erabiltzen dituzte (ikusi 2.3 irudia eta (22) adibidea):

(22) *eat-39.1-1*: agent (+animate); patient (+comestible & +solid)

- **Egitura sintaktikoak** (*Frames*): aditz-klase bakoitzak onar ditzakeen egitura sintaktikoak zehazten dira (ikusi 2.4 irudia):

FRAMES	
NP V NP PP.DESTINATION	
EXAMPLE	"Jessica loaded boxes into the wagon."
SYNTAX	<u>AGENT</u> <u>V</u> <u>THEME</u> {{+LOC +DEST_CONF}} <u>DESTINATION</u>
SEMANTICS	MOTION(DURING(E), THEME) NOT(PREP(START(E), THEME, DESTINATION)) PREP(END(E), THEME, DESTINATION) CAUSE(AGENT, E)
NP V NP.DESTINATION PP.THEME	
EXAMPLE	"Jessica loaded the wagon with boxes."
SYNTAX	<u>AGENT</u> <u>V</u> <u>DESTINATION</u> {WITH} <u>THEME</u>
SEMANTICS	MOTION(DURING(E), THEME) NOT(LOCATION(START(E), THEME, DESTINATION)) LOCATION(END(E), THEME, DESTINATION) CAUSE(AGENT, E)
NP V NP.THEME	
EXAMPLE	"Jessica squirted water."
SYNTAX	<u>AGENT</u> <u>V</u> <u>THEME</u>
SEMANTICS	MOTION(DURING(E), THEME) NOT(LOCATION(START(E), THEME, ?DESTINATION)) LOCATION(END(E), THEME, ?DESTINATION) CAUSE(AGENT, E)
NP V NP.DESTINATION	
EXAMPLE	"Jessica sprayed the wall."
SYNTAX	<u>AGENT</u> <u>V</u> <u>DESTINATION</u>
SEMANTICS	MOTION(DURING(E), ?THEME) NOT(LOCATION(START(E), ?THEME, DESTINATION)) LOCATION(END(E), ?THEME, DESTINATION) CAUSE(AGENT, E)

2.4 irudia: *spray-9.7* klaseak onar ditzakeen egitura sintaktikoak.

Top Class horretan zehazten diren egitura sintaktikoak eta rol semantikoak klase horretako aditz guztiek partekatzen dituzte, eta, 2.5 irudian ikusten den bezala, klase nagusiaren azpian, azpiklase bakoitzaren osagaiak zehazten dira goian aurkeztu dugun modu berean (aditzak, rolak, hautapen-murriztapenak eta egitura sintaktikoak). Kasu honetan, klase nagusiarekiko aldatzen dena *Theme* rolari hautapen-murriztapenak gehitu dizkiotela da (+*substance*, +*concrete* eta +*plural*):

spray-9.7-1		
<i>Members: 29, Frames: 2</i>		
<i>No Comments</i>		
MEMBERS		
BASTE (WN 1; G 1)	SEED (WN 4; G 3)	SPLATTER (FN 1; WN 1, 2)
BRUSH (FN 1, 2; WN 6; G 2)	SEW (FN 1; WN 1; G 1)	SPRAY (FN 1, 2; WN 1, 2, 3; G 1)
DRIZZLE (FN 1, 2; WN 2)	SHOWER (FN 1, 2; WN 1, 2, 5; G 1, 2)	SPREAD (FN 1, 2; WN 3, 9, 10; G 2, 3)
HANG (FN 1, 2; WN 2, 12, 14; G 1)	SMEAR (FN 1, 2; WN 3, 2; G 1)	SPRINKLE (FN 1, 2; WN 1, 4; G 1)
PLASTER (FN 1, 2; WN 2, 3, 4, 1, 5, 6)	SMUDGE (WN 1)	SPRITZ (WN 1, 2)
PUMP (FN 1, 2; WN 2, 4, 5; G 2)	SOW (FN 1, 2; WN 1, 3; G 1)	SPURT (WN 1)
RUB (FN 1, 2; WN 1; G 1)	SPATTER (FN 1; WN 1, 3)	SQUIRT (FN 1, 2, 3; WN 1, 2; G 1)
SCATTER (FN 1, 2; WN 3, 4, 6; G 1)	SPLASH (FN 1, 2, 3, 4; WN 3, 6; G 1)	STICK (FN 1, 2, 3, 4, 5; WN 1, 12, 13; G 1, 2)
ROLES		
• THEME [+SUBSTANCE [+CONCRETE & +PLURAL]]		
FRAMES		
NP V PP.DESTINATION		
EXAMPLE	"Paint sprayed onto the wall."	
SYNTAX	<u>THEME</u> V {{+LOC +DIR +DEST_CONF}} <u>DESTINATION</u>	
SEMANTICS	MOTION(DURING(E), THEME) NOT(PREP(START(E), THEME, DESTINATION)) PREP(END(E), THEME, DESTINATION))	
NP V NP PP.DESTINATION-CONATIVE		
EXAMPLE	"Jessica squirted water at me."	
SYNTAX	<u>AGENT</u> V <u>THEME</u> (AT) <u>DESTINATION</u>	
SEMANTICS	MOTION(DURING(E), THEME) NOT(LOCATION(START(E), THEME, DESTINATION)) CAUSE(AGENT, E))	

2.5 irudia: *spray-9.7-1* azpiklasearen informazioa VN-n.

VerbNet-en 21 rol semantikoez osatutako zerrenda erabiltzen dute³⁶:

- *Actor*: komunikazio-gertaerak adierazten dituzten klaseetan (esaterako, *chitchat-37.6*, *marry-36.2* eta *meet-36.2*) erabiltzen dira, bi argumentuak (objektu/subjektu) semantikoki simetrikotzat har daitezkeen (sasi-agenteak)³⁷.
- *Agent*: predikatuak adierazten duen gertaera nahita eragiten eta gauzatzen duena, subjektua. Oro har, biziduna edo gizakia, baina makinak edo borondatea erakusten duen edozer ere izan daiteke.
- *Asset*: ordainketak eta antzeko diru-trukaketak edota -aldaketak adierazten dituzten klaseek erabiltzen dute. Rolak *currency* ("dirua") hautapen-murriztapena du.

³⁶Zerrenda hori eta definizioak sareko <http://verbs.colorado.edu/~mpalmer/projects/verbnet.html> helbidetik hartu ditugu (2013-12-12).

³⁷Gaur egun, eta 2.1.3 ataleko 2.2 taulan ikusten den bezala, *Actor* hauek *Agent* eta *Co-Agent* bihurtu dira

- *Attribute*: predikatuak parte hartzen duen eta aldatzen ari den zerbaiten ezaugarria.
- *Beneficiary*: predikatuak adierazten duen gertaeraren onura edo etekina jasotzen duena.
- *Cause*: predikatuak adierazten duen gertaeraren eragilea. Oro har, *Agent* rolari kontrajarria, ekintza nahi gabe, berez, gertatzen dela adierazten du. Normalean aditz psikologikoekin eta gorputzarekin zerikusia duten aditzetan erabiltzen da.
- *Destination*: predikatuak adierazten duen mugimenduaren helburuko kokapena edo mugimenduaren norabidea.
- *Experiencer*: kontziente den edota zerbait esperimendatzen ari den parte-hartzailea. Oro har, aditz psikologikoetan, hautemate-aditzetan, eta gorputzarekin eta ukitzearekin zerikusia duten aditzetan erabiltzen da.
- *Extent*: predikatuak adierazten duen aldaketaren maila edo gradua (*Calibratable-45.6* klasean baino ez da erabiltzen).
- *Instrument*: tresna gisa erabiltzen diren objektuak, aldaketak eragiten dituzten objektuak.
- *Location*: predikatuak adierazten duen gertaera non kokatzen den adierazten duena, kokapen konkretua.
- *Material*: predikatuak adierazten duen aldaketaren hasierako egoera, gertaeran zehar aldatuz doana eta entitate berri bihurtzen dena (*build-26.1* eta *grow-26.2* klaseetan baino ez da erabiltzen).
- *Product*: predikatuak adierazten duen aldaketaren azken egoera, emaitza (*build-26.1* eta *grow-26.2* klaseetan baino ez da erabiltzen).
- *Patient*: predikatuak adierazten duen gertaeran egoera-, kokapen- edo izaera-aldaketa jasaten duena. Horrelako aldaketak adierazten dituzten aditzen objektuentzat erabiltzen da.
- *Predicate*: osagarri predikatiboa duten aditz-klaseetan erabiltzen da.
- *Recipient*: predikatuak adierazten duen gertaeraren helburu biziduna.

- *Source*: predikatuak adierazten duen mugimenduaren abiapuntuak kokapena.
- *Stimulus*: *Experiencer* diren parte-hartzaileetan nolabaiteko erantzun emozionala edo psikologikoa eragiten duena. Hautemate-aditzetan erabiltzen da.
- *Time*: gertaerak edo egoerak irauten duen denbora-tartea adierazten duen parte-hartzailea (*begin-55.1* klasean baino ez da erabiltzen).
- *Theme*: kokapen jakin batean dagoen edo kokapen-aldaketa jasaten duen parte-hartzailea.
- *Topic*: komunikazio-aditzetan hizpide den gaia adierazteko erabiltzen da.

Ikusten dugunez, VerbNet-en ere Gruber, Jackendoff eta Grimshaw-en tipologian oinarritzen den zerrenda erabiltzen dute. Hasiera batean behintzat, rol hauek klaseak errepresentatzeko pentsatuak dira (Levinen klaseak oinarrian, bederen).

Gauzak horrela, bertsio desberdinak ateratzen doaz, etengabe aditz eta klase semantiko berriak gehituz (Kipper *et al.*, 2006, 2008); memento honetan 3.2 bertsioan daude.

Gainera, azkenaldian, 2.1.3 atalean azalduko dugun LIRICS proiektua³⁸ oinarri hartuta, rol semantikoen zerrenda berrikusten ari dira. Azterketa horren emaitzak rol semantikoen zerrenda luzeagoa³⁹ eta rolen hierarkia izan dira. Gure ikerlan honetan, hemen aurkeztu dugun rolen zerrenda erabiliko dugu, hori baita lana hasi genuenean VN-n erabiltzen zutena.

Atal honen hasieran esan dugun bezala, VerbNet, izatez, baliabide lexikoa da eta ez du rol semantikoekin etiketatutako corpusik eskaintzen. Gabezia horri irtenbidea emateko, *SemLink*⁴⁰ baliabidea (Loper *et al.*, 2007) sortu zen. Baliabide honek VNko rol semantikoak eta hurrengo atalean azalduko dugun PropBank (Palmer *et al.*, 2005) corpora lotzen ditu, PBko argumentu zenbakitu bakoitzari dagokion VNko rol semantikoa esleituz.

³⁸<http://lirics.loria.fr>

³⁹Zerrenda berria, eta rol semantikoen definizio guztiak, sarean eskuragarri dagoen *VerbNet Guidelines.pdf*-n ikus daitezke.

⁴⁰<http://verbs.colorado.edu/semlink/>

2.1.2.2 PROPBANK

Paul Kingsbury eta Martha Palmer jo ohi dira **PropBank** (PB) proiektuaren aitzindari gisa (Kingsbury eta Palmer, 2002; Palmer *et al.*, 2005). PBren web-orrian aurkituko ditugu proiektuaren zehaztasunak⁴¹.

Guk, jarraian, proiektuaren nondik norakoak aipatuko ditugu labur:

- Ingelesezko *Wall Street Journal* corpora semantikoki etiketatzeko proiektua da.
- Oinarrian *Penn Treebank*-eko zuhaitz sintaktikoak ditu.
- Aditz-adiera bakoitzari rol-zerrenda desberdina egokitzen diote.
- Hipotesia:
 - Levenen ezagutza lexikalaren teoriaren ildotik, aditzaren esanahia (klase semantikoa) zuzenean dago lotuta delako aditzaren izaera sintaktikoarekin, hau da, izaera sintaktiko bera duten aditzek osagai semantiko berberak izango dituzte:

“Further examination of the nature of lexical knowledge confirms that various aspects of the syntactic behavior of verbs are tied to their meaning. Moreover, verbs that fall into classes according to shared behavior would be expected to show shared meaning components” (Levin, 1993: 5)
 - Demagun, 3 aditz iragankor ditugula:
 - * *to break*
 - * *to burn*
 - * *to touch*

Hirurek dituzte subjektu bat eta objektu bat, baina horietako bik bakarrik onartzen dute kausatibo/inkoatibo (DA-DU) alternantzia (23a eta 23b):

- (23) a. *to break* ('puskatu')
- *John broke the window*⁴².
 - *The window broke.*

⁴¹<http://verbs.colorado.edu/~mpalmer/projects/ace.html>

⁴²Hemen ematen ditugun ingelesezko adibideak guk asmatuak dira, Levenen (1993) lanean oinarrituta.

- b. *to burn* ('erre')
 - *The man burnt the house.*
 - *The house burnt.*
- c. *to touch* ('ukitu')
 - *Carol touched the dog.*
 - **The dog touched.*

Kausatibo/inkoatibo alternantzia onartzen duten bi aditzek (*to break* eta *to burn*) predikatu-mota bera adierazten dute: egoera-aldaketa. Egitura kausatiboan egoera-aldaketaren kausa den subjektua eta egoeraz aldatzen den objektua (pazientea) ditugu eta, inkoatiboan, berriz, subjektua da egoera-aldaketa jasaten duen pazientea. Hirugarren aditzak ez du alternantzia hori onartzen eta ez du egoera-aldaketa adierazten, subjektuaren eta objektuaren arteko kontaktua baizik. Beraz, izaera sintaktikoa kontuan hartuta, aditzak klase semantikoetan sailkatzen saiatzen dira.

- Etiketatzean, elkarren artean independente diren 2 maila bereizten dituzte:
 - Argumentu zenbakituak edo balentzia (0tik 4ra: Arg0, Arg1, ..., Arg4), eta adjuntuak (adjuntuek ez dute zenbakirik; besterik gabe modifikatzailearen M letra (ArgM)).
 - Rol semantikoak: oso rol espezifikoa dira, aditz banakoari lotuak (*buyer, thing bought, speaker...*); horregatik ez dago rolen zerrenda estandar itxi bat. Beharren arabera erabiltzen dituzte. Hauxe da euren filosofia:
 - * Corpus bat etiketatzeko erabilgarriak diren rolak sortu dituzte.
 - * Aditz baten rolen kasuistika corpusean oinarrituta dago.

Bi maila independente egoteak hau esanahi du:

- Argumentu batek rol semantiko desberdinak har ditzake (adibidez, aditzaren edo aditz-adieraren arabera, Arg1 *Theme* edo *Topic* izan daiteke; eta orobat Arg0, *Agent* edo *Theme*⁴³).

⁴³Hori ez da beti horrela betetzen PropBank-en. Horri buruzko xehetasun gehiago 3.1 atalean.

- Era berean, rol semantiko bera argumentu desberdinetan ager daiteke (*Theme* batzuetan Arg1 da eta beste batzuetan Arg0).

Funtzio sintaktikoekin parekatuz gero, eta inakusatibitatearen teoria jarraituta, horrelaxe agertuko lirateke argumentu zenbakituak adibideotan:

1. Aditz iragankorrek: **Pellok liburua** ekarri du

Arg0 = SUBJ

Arg1 = OBJ

2. Aditz inakusatiboak: **Pello** etorri da

Arg1 = SUBJ

3. Aditz inergatiboak: **Pellok** dimititu du

Arg0 = SUBJ

Jarraian, PropBank-eko adibide bat (24) emango dugu etiketatzeko modua argiago ikusteko. Hona *to buy* aditzaren agerpen etiketatu bat:

(24) *John*_{Arg0} *bought* *his mother*_{Arg2} *a dozen of roses*_{Arg1}

Arg0: John

REL: bought

Arg2: his mother

Arg1: a dozen roses

Aditz-sarrerak eta esaldi etiketatuak beheragoko bi helbideetan ikus daitezke: bata zaharra eta bestea gaur egungoa. Izan ere, PropBank-ek hainbat bertsio izan ditu eta datu-basea etengabe eguneratuz doaz. IXA taldean erosi zen datu-basea lehenengo bertsioari dagokio (ikusi 4.2.1 atala). Izatez, badira desberdintasunak bien artean, eta horregatik komeni da bi bertsioak begi-bistan izatea:

- Zaharra: <http://verbs.colorado.edu/propbank/framesets-old/>
- Gaur egungoa: <http://verbs.colorado.edu/propbank/framesets-english/>

Horretaz gain, esana dugun bezala, PropBank-en aditz-adiera⁴⁴ bakoitzak bere gauzatze sintaktikoak edo *frameak* ditu. Adibidez: *tell.01* aditz-adierak (25) adibideko *frameak* ditu:

(25) rolset *tell.01*:

Arg0 speaker
Arg1 utterance
Arg2 hearer

- **ditransitive**⁴⁵

The score tell you what the characters are thinking and feeling.

Arg0: The score
REL: tell
Arg2: you
Arg1: what the characters are thinking and feeling

- **odd ditransitive**⁴⁶

*What this tells us *trace*⁴⁷ is that US trade law is working.*

Arg0: this
REL: tells
Arg2: us
Arg1: *trace* is that US trade law is working

⁴⁴Aditz-adiera deitzen diogu PropBank-eko sarreretan agertzen diren 01, 02, 03... horiei; PropBank-en *rolset* esaten diote aditz-adierari.

⁴⁵Ditrantsitiboa, hau da, bi objektu (zuzena eta zeharkakoa) dituen.

⁴⁶Ditrantsitibo berezia, arrunta ez dena.

⁴⁷*Penn Treebank*ean *trace* (arrasto) izeneko kategoria hutsak etiketatzen dira. Askotan, kategoria huts horiek zuhaitzeko beste osagaiekin indexatuta egoten dira.

- **prepositional arg2**⁴⁸

John told a story to Mary.

Arg0: John
REL: told
Arg1: a story
Arg2: to Mary

- **fronted**⁴⁹

*["Only Nixon could go to China,"]-1 he told a group of Americans *trace*-1*

Arg0: he
REL: told
Arg2: a group of Americans
Arg1: *trace*

2.4 atalean azaldu dugun bezala, *SemLink* proiektuan PropBank Ver-Net-i lotua agertzen zaigu (Loper *et al.*, 2007). PropBank-eko argumentu zenbakituen ondoan, Ver-Net-eko rola zein den ikus dezakegu, bai eta PB-ko aditz-adiera bakoitzari esleitzen zaion VNko klase semantikoa (*vncls*) ere. Honela, (26) adibidean ikusten den bezala, predikatuaren alderdi semantikoa gehitzea eta, beraz, sarrera osatuagoa izatea lortzen da:

- (26) buy.01, purchase, source; vncls: 13.5.1
- Arg0:** *buyer* (vnrole: 13.5.1-Agent)
- Arg1:** *thing bought* (vnrole: 13.5.1-Theme)
- Arg2:** *seller* (vnrole: 13.5.1-Source)
- Arg3:** *price paid* (vnrole: 13.5.1-Asset)
- Arg4:** *benefactive* (vnrole: 13.5.1-Beneficiary)

⁴⁸Arg2 preposizio-sintagma denean.

⁴⁹Arg1 aditzaren aurretik agertzen denean.

IXA taldean PropBank datu-basea erosi zen garaian, bi baliabide horiek ez zeuden lotuta, eta taldean lotu ziren informazio guztia eskura izateko⁵⁰.

Birgogoratzen dugu gaur egun PBk, bai eta VNk ere, beste bertsio bat darabilte, eta konprobatu dugu aldeak egon badirela (ikusi 4.2.1 atala).

Sareko *PBguidelines.pdf*-n⁵¹ aurki daitezke PropBank ereduarekin etiketatzeko xehetasun gehiago (Babko-Malaya, 2005).

Esan behar da PB ereduak beste hainbat hizkuntza tratatzeko erabili dela: hindia (Bhatt *et al.*, 2009), txinera (Xue eta Palmer, 2003; Xue, 2008; Xue eta Palmer, 2009), koreera (Palmer *et al.*, 2006), arabiera (Palmer *et al.*, 2008), gaztelania (Aparicio, 2007; Taulé *et al.*, 2006), katalana (Civit *et al.*, 2005; Taulé *et al.*, 2006), frantsesa (Gardent eta Cerisara, 2010; Van Der Plas *et al.*, 2010), Brasilgo portugesa (Duran eta Aluísio, 2012) eta nederlandera (Monachesi *et al.*, 2007).

2.1.2.3 FRAMENET

PropBank eta **FrameNet** (FN) (Baker *et al.*, 1998; Fillmore eta Baker, 2001)⁵² askotan konparatu izan dira, bataren eta bestearen arteko hautua egiteko. Gauza da bien abiapuntua eta, beraz, ikuspegia, desberdina dela. PropBank aditzean eta aditz-klaseetan oinarritzen da, eta FrameNet, berriz, marko edo *frame*etan. FrameNet-en markoak edo *frame*ak eszenario semantikoak dira (ez egitura sintaktikoak), eta marko bakoitzean aditz bat baino gehiago bil daitezke.

FrameNet Fillmore-ren *Markoen Semantika* teorian (Fillmore, 1976) oinarrituta dago. Markoak egoera, objektu edo gertaera-mota batzuen erre-presentazio semantikoak dira, eta kontzeptuen arteko egitura koherente bat osatzeko balio dute. Markoetan jasotzen den ezagutza munduari buruz dugun ezagutza orokorra da. Hitz baten esanahi semantikoa zehazteko erabiltzen den egitura da marko semantikoa.

Markoetan, alde batetik, unitate lexikalak (*Lexical Unit*) ditugu eta, bestetik, marko-elementuak (*Frame Element*). Unitate lexikalak, hitz bat bere adierarekin lotzen duten elementuak dira, marko semantikoak aktibatzen dituzten elementuak, hain zuzen ere. Marko-elementuak, aldiz, unitate lexikalei esleitzen zaizkien rol semantikoak dira; horietako batzuk *core*ak dira eta,

⁵⁰PBko eta VNko datu-baseak lotuta, ikusgai ditugu biotan ageri diren aditzen zerrendak, dagokien informazioarekin (rolset-zenbakia edo aditz-adiera, argumentuak, rolak...).

⁵¹<http://verbs.colorado.edu/~mpalmer/projects/ace/PBguidelines.pdf>

⁵²<http://framenet.icsi.berkeley.edu>

beraz, oinarrizkoak direla ulertu behar da, eta beste batzuk, berriz, *non-coreak* dira, eta horrek esan nahi du adjuntuak direla, baina ez oinarrizkoak edo ezinbestekoak.

Ikus dezagun hemen ere *to buy* aditza. *Buy* unitate lexikala *commerce_buy* marko semantikoan sartuta dago. (27) adibidean ikus daiteke markoaren definizioa:

- (27) These are words describing a basic commercial transaction involving a Buyer and a Seller exchanging Money and Goods, taking the perspective of the Buyer. The words vary individually in the patterns of frame element realization they allow. For example, the typical pattern for the verb BUY: Buyer buys Goods from Seller for Money.

Abby bought a car from Robin for \$5,000.

Jarraian, marko-elementuak zehazten eta azaltzen dira, *core* elementuak lehenik eta *non-coreak* ondoren:

- a) Oinarrizko elementuak edo *coreak*: *buyer, goods*.
- b) Elementu adjuntuak edo *non-coreak*: *manner, means, money, period of iteration, place, purpose, purpose of goods, rate, reason, recipient, seller, time, unit*.

Ikusten dugunez, FrameNet-en kasuan marko bakoitzerako rol semantikoak definitzen dira; hau da, PropBank-en antzera, ez dute rolen zerrenda itxirik erabiltzen, eta marko semantiko bakoitzari bere rol konkretuak definitzen dizkiote.

Eredu horrekin hainbat hizkuntza (txinera (You eta Liu, 2005), Brasilgo portugesa (Salomão, 2009), alemana (Boas, 2009), gaztelania (Subirats eta Petruck, 2003), japoniera (Ohara, 2012), suediera (Friberg eta Toporowska, 2012)) tratatu dira, baina oraindik zenbaitetan marko gutxi batzuk baino ez dira landu.

2.4 eta 2.1.2.2 ataletan aipatu dugun *SemLink* proiektuan, FrameNet baliabidea ere gehitu dute, hiru baliabideen arteko (PB, VN eta FN) loturak sortuz (Palmer, 2009). VerbNet azaldu dugunean, WordNet eta OntoNotes

ere lotuta zeudela ikusi dugu, eta, baliabideen arteko lotura horiekin jarraituz, *Unified Verb Index* esker⁵³, batera ikus dezakegu PB, VN, FN eta OntoNotes baliabideek eskaintzen duten informazio guztia. Sistema hori aditzen aurkibide bat da. Aurkibidean aditz bat bilatuta, aipatutako baliabideetara dituen estekak ikus daitezke. Esaterako, *tell* aditza bilatuta, 2.6 irudian agertzen den informazioa ikus dezakegu:

telework	(PROP BANK)
telex	INSTR_COMMUNICATION-37.4, (PROP BANK), (FN COMMUNICATION_MEANS), (FN SENDING), (FN CONTACTING)
tell	TELL-37.2-1, TRANSFER_MESG-37.1.1-1-1, (PROP BANK), (FN REQUEST), (FN REPORTING), (FN TELLING), (GROUPING)
tell off	(FN JUDGMENT_DIRECT_ADDRESS)
temp	(FN BEING_EMPLOYED)
temper	(PROP BANK)
tempt	AMUSE-31.1, FORCE-59, (PROP BANK), (FN ATTEMPT_SUASION), (GROUPING)
tend	(PROP BANK), (FN LIKELIHOOD), (GROUPING)

2.6 irudia: *tell* aditzak dituen PB, VN, FN eta OntoNotes-eko loturak.

2.6 irudian ikusten den bezala, *tell* aditzak bi lotura ditu VNrekin: *tell-37.2-1* azpiklasera lehena, eta *transfer_mesg-37.1.1-1-1* azpiklasera bigarrena. PBko lotura ere ikusten dugu, PBko *tell* aditzera eramaten gaituena. Era berean, FNko hiru markotara dago lotuta *tell* aditza: i) *request*, ii) *reporting*, eta iii) *telling*. Bukatzeko, OntoNotes-eko adiera-multzokatzearekin duen lotura ikus dezakegu (*grouping*). Esteka horietako bakoitzean klik egin da, sistemak dagokion baliabidearen webgunera eramaten gaitu.

2.1.2.4 ADESSE

ADESSE⁵⁴ (*Alternancias de diátesis y esquemas sintáctico-semánticos del español*) proiektua gaztelaniarako garatu zen (García-Miguel eta Albertuz, 2005). Proiektu horretan milioi bat eta erdi hitz dituen gaztelaniazko *Arthus*⁵⁵ (*Archivo de Textos Hispánicos de la Universidad de Santiago*) corpusa etiketatu da aditzen informazio sintaktiko-semantikoarekin, eta informazio hori

⁵³<http://verbs.colorado.edu/verb-index/>

⁵⁴<http://adesse.uvigo.es/>

⁵⁵<http://adesse.uvigo.es/data/corpus.php>

guztia datu-base batean gorde da. 3.400 aditz daude landuta eta 4.000 aditz-adiera baino gehiago. Aditz-adiera bakoitza sarrera bat da datu-basean eta sarrera horiek unitate lexikoak dira. Unitate lexiko bakoitzari klase semantiko bat eta argumentu-multzo bat esleitzen zaizkio.

Argumentuak informazio sintaktiko-semantikoarekin etiketatzen dituzte:

- Funtzio sintaktikoa (subj., obj. zuzena, zeharkako obj. etab.)
- Katetoria sintaktikoa (infinitihozko klausula, klausula mugatua etab.)
- Hautapen-murriztapenak (biziduna, konkretua, abstraktua etab.)
- Rol semantikoa (zenbakiak, eta aditz bakoitzari edo klase semantiko bakoitzari lotutako etiketak erabiliz)
- Gune lexikoa (gauzatze lexiko ohikoenak, aditz-izen konbinazio ohikoenak ateratzeko)

Horretaz guztiaz gain, aditz bakoitzari honako informazio semantiko hau ezartzen zaio:

- Aditz-adiera
- Aditzaren klase semantikoa edo prozesu-mota

Datu-basean aditz bati buruzko informazioa bilatzen dugunean, 2.7 irudian erakusten den informazioa ikus dezakegu.

2.7 irudian ikusten den bezala, *comer* aditzaren klasea edo prozesu-mota “irenstea edo ingestioa” da (*Ingestión*). Jarraian, aditzak dituen argumentuak zehazten dira A1, A2... etiketekin, bai eta argumentuek betetzen dituzten rolen etiketekin ere: i) aditzari lotutako rolak (*comedor*, *comida*) eta ii) klaseari lotutako rolak (*ingestor*, *ingesta*, *poseedor*, *origen*). Argumentu bakoitza corpusean zein portzentajetan agertu den zehazten da eskuinean, eta ondoren, aditzaren eskema sintaktiko-semantiko desberdinak (argumentu semantikoak eta funtzio sintaktikoak) erakusten dira, bakoitza zenbat adibide-tan agertu den zehaztuz. Bukatzeko, corpuseko adibideak ere ikus daitezke.

[–] Clasificación semántica y potencial valencial

Tipo de proceso: **Ingestión**

Argumentos:		Frecuencia
A1 INGR	Comedor INGESTOR	291 (98 %)
A2 ING	Comida INGESTA	170 (57.2 %)
A3 POSR	Poseedor	7 (2.4 %)
A4 ORI	Origen	2 (0.7 %)

perfil combinatorio >>

[–] Realizaciones valenciales (Esquemas sintáctico-semánticos):

Voz	Argumentos semánticos y Funciones sintácticas	N_ejemplos
COMER _{act}	A1:INGR = SUJ	121 >
COMER _{act}	A1:INGR = SUJ A2:ING = ODIR	96 >
COMER _{act}	A1:INGR = SUJ A2:ING = ODIR A3:POSR = OIND	5 >
COMER _{act}	A1:INGR = SUJ A4:ORI =de OBL	2 >
COMERSE _{med}	A1:INGR = SUJ A2:ING = ODIR	60 >
COMERSE _{med}	A1:INGR = SUJ A2:ING = ODIR A3:POSR = OIND	2 >
COMERSE _{impasiva}	A2:ING = SUJ	6 >
COMERSE _{impers}	A1:INGR = SUJ	4 >
COMERSE _{impers}	A1:INGR = SUJ A2:ING = ODIR	1 >

[–] Ejemplos:
solo 10, escogidos aleatoriamente, de un total de 297)

Verbo	Texto
COMER	Inf. B - ...pero no son las más. ¿En general qué hacemos? Vamos mucho a comer . Enc. - Ajá. No se nota. ¿eh?, están flacos.
COMER	me cabrea y no es que yo esté tan rematado, aunque la Rusca ha empezado a comerme más abajo. . Eso es que hoy ha pasado algo. »

2.7 irudia: *comer* aditzaren sarrera ADESEn.

Gainera, eskema sintaktiko-semantiko bakoitzaren eskuinean dagoen zenbakiaren ondoko ikurrean klik eginez gero, eskema horretako argumentuen informazio gehigarria ikus daiteke (2.8 irudia):

Perfil combinatorio (resumen cuantitativo de propiedades de argumentos)

COMER - Ingerir alimentos sólidos

Esquema:
 COMER_{act} = **A1:INGR** **A2:ING** **A3:POSR** >
 = **SUJ** = **ODIR** = **OIND**

ejemplos anotados

Argumento: (Frec. explícito)	A1 Comedor (Ingestor)	A2 Comida (Ingesta)	A3 (Poseedor)
Función sintáctica:	suj 5	ODIR 5	OIND 5
Clíticos objeto:			Ac/Dat 4 Dat 1
Preposiciones:	N/A o nulo 3	N/A o nulo 5	
Categoría sintáctica:	FN 3	FN 5	
Tipo semántico:	Animado 5	Concreto 4 Abstracto 1	Animado 5
Realizaciones léxicas frecuentes: (sólo no animados)		coco 2 hígado 1 tarro 1 coraje 1	

2.8 irudia: *comer* aditzaren eskema sintaktiko-semantic baten informazioa.

2.8 irudian, 2.7 irudiko hirugarren eskema sintaktiko-semanticari dagokion informazioa ikus daiteke. Egitura sintaktiko-semantic horrek hiru argumentu ditu: A1, A2 eta A3. Bakoitzaren rol semantic zehazten da: A1, *comedor*, *ingestor* (INGR); A2, *comida*, *ingesta* (ING); eta A3, *poseedor* (POSR). Argumentuek betetzen duten funtzio sintaktikoa markatzen da: A1, *SUJ* (subjektua); A2, *ODIR* (objektu zuzena); eta A3, *OIND* (zeharkako objektua). Kategoría sintaktikoaren informazioa ematen da: A1 eta A2, *FN* (sintagma nominala). Hautapen-murriztapenak zehazten dira: A1, *animado* (biziduna); A2, *concreto* (konkretua), kasu batean *abstracto* (abstraktua); eta A3, *animado*. Bukatzeko, ohikoen diren gauzatze lexikoak ere aipatzen dira: A2 argumentuaren kasuan, adibidez, *coco*, *hígado*, *tarro*, *coraje*. Eskuinean agertzen diren zenbakiek delako elementua zenbat adibidetan agertu den zehazten dute.

2.1.2.5 SENSEM

SENSEM⁵⁶ (*Sentence Semantics: Creación de una Base de Datos de Semántica Oracional*) proiektuan gaztelaniazko 850.000 hitzez osatutako corpusa etiketatu dute informazio sintaktiko-semantikoarekin (Castellón *et al.*, 2006; Vázquez *et al.*, 2006), eta, corpus horretako informazioa oinarri hartuta, aditzen lexikoa ere osatu dute. Corpusean gaztelaniaz maiztasun handia duten 250 aditzak bildu dituzte⁵⁷. Gaztelaniazko corpusaren kazetaritza-testuen itzulpenak bilduta, katalanezko corpusa ere osatu dute, eta gaztelaniarako etiketatutako informazio sintaktiko-semantiko berbera mantendu dute (beharrezkoa izan denean zenbait moldaketa eginda). Katalanezko corpus honek 600.000 hitz inguru ditu, eta webgunean kontsultagarri dago katalanezko aditzen lexikoiarekin batera.

Proiektu honetan, aditzaren mende dauden elementuei edo predikatuaren parte-hartzaileei informazio hau ezartzen zaie:

- Rol semantikoa
- Kategoria sintaktikoa
- Funtzio sintaktikoa
- Argumentu-izaera

Honetaz gain, aditzari eta perpausari buruzko informazioa ere ematen da:

- Aditzak
 - aditz-adierak
- Perpausaren semantika
 - aspektualitatea eta aspektua
 - modalitatea
 - polaritatea
 - joskeren semantika

⁵⁶<http://grial.uab.es/projectes/SenSem.php>

⁵⁷25.000 esaldi (700.000 hitz) kazetaritza-testuetatik hartuak, eta 5.000 esaldi (150.000 hitz) literatura-testuetatik.

Esan bezala, etiketatutako informazio hori oinarri hartuta, aditzen lexikoa osatu dute, eta corpusaren⁵⁸ eta lexikoiaren⁵⁹ gainean bilaketa zehatzak egiteko kontsulta-tresna ere garatu dute.

Aurreko adibidearekin jarraituz, ikus dezagun 2.9 irudian *comer* aditzak SENSEMeko lexikoian duen informazioa.



The screenshot shows the GRIAL website interface. At the top, there are navigation links: Inicio, Corpus, Léxico, Subcategorizaciones, and Descargas. The main header features the GRIAL logo and the text 'Léxico verbal del corpus'. Below this, there is a search section titled 'Criterio de búsqueda' with a 'Buscar por verbo' form. The form includes a dropdown for 'Idioma' set to 'Español' and a text input for 'Verbo' set to 'comer'. A 'Buscar' button is present. To the right of the search form, the results for 'come:' are displayed in a table.

Sentido	Definición	Roles	Clase aspectual	Frecuencia
1	<u>Ingerir los alimentos una vez masticados.</u>	Agente, Tema	Proceso/Evento	103/120
2	<u>Cuando se trata de un juego, hacer que el jugador contrario se quede sin una de sus piezas.</u>	Agente, Tema, Destino	Evento	0/120
3	<u>Dicho de algún producto o efecto, hacer que algo pierda intensidad o color.</u>	Iniciador, Tema afectado	Proceso	0/120
4	<u>En un escrito o un oral, no escribir o decir algo en su totalidad, dejarse algo.</u>	Iniciador, Tema	Evento	0/120
5	<u>Empequeñecer los espacios o figuras en un efecto óptico.</u>	Tema, Tema	Estado	0/120
6	<u>Tomar la comida del mediodía</u>	Agente	Proceso	16/120
7	<u>loc.: comerse un marrón</u>			0/120
8	<u>Producir rabia o gran malestar algo (envidia, celos, etc.) a alguien.</u>	Causa, Tema afectado	Evento	1/120
9	<u>loc. comerse un rosco</u>			0/120

2.9 irudia: *comer* aditzaren sarrera SENSEMen.

2.9 irudian ikusten denez, hasteko, 2. zutabea *comer* aditzaren adierak zehazten dira definizioen bidez. Gero, 3. zutabea adiera bakoitzak dituen rol semantikoak ematen dira. Ondorengo zutabea, aspektua zehazten da, eta, bukatzeko, adiera bakoitza corpusean zein maiztasunekin agertu den esaten da.

Webgunean, 2.10 irudian ikusten den bezala, aditzaren corpuseko adibide etiketatua ere kontsulta daitezke.

⁵⁸<http://grial.uab.es/sensem/corpus/main>

⁵⁹<http://grial.uab.es/sensem/lexico?idioma=es>

28572	También tengo miedo de encontrar, debajo de mis dibujos, escorpiones que se comen las termitas cuyos agujeros utilizo". (anotación) (más info.)
28573	Tenía que comer , tenía que vivir. (anotación) (más info.)
28574	Vila ha realizado un exhaustiva investigación basándose en documentos de distintos archivos para reconstruir qué y cómo se ha comido en Girona durante siglos. (anotación) (más info.)
28575	En el restaurante donde comí el primer día, el camarero me señaló una mesa cercana y me dijo: esos son paisanos suyos. (anotación) (más info.)
28576	Burgos se comió la broma y escribió el citado artículo. (anotación) (más info.)
28577	Me dicen que en sus restaurantes se come bien. (anotación) (más info.)
28578	Era una familia humilde, pero comían como reyes. (anotación) (más info.)
28579	Sin caer nunca en la gula, gustaba del buen comer: "Viene ahora el Hijo del Hombre, que come y bebe, y dicen:" i He aquí un tragón y bebedor! (anotación) (más info.)

2.10 irudia: *comer* aditzaren lehen adieran etiketatutako adibideen lagin bat.

Adibideetako *anotación* estekan klik eginda, adibide bakoitzaren etiketatzea ikus daiteke (2.11 irudia):

28576 Burgos se **comió** la broma y escribió el citado artículo. (anotación) (más info.)

Evento Perfectivo
 Modalidad asertiva - Polaridad positiva
 Topicalización sujeto lógico - Construcción de dativo

Burgos se **comió** la broma y escribió el cit

Sujeto V Objeto directo
 Sintagma nominal (nombre propio) Metafórico Sintagma nominal (nombre común)
 Agente Tema

28577 Me dicen que en sus restaurantes se **come** bien. (anotación) (más info.)

Proceso Habitual
 Modalidad asertiva - Polaridad positiva
 Destopicalización sujeto lógico - Impersonal - Pronominal

Me dicen que en sus restaurantes se **come** bien .

Circunstancial V Circunstancial
 Preposición + sintagma Sintagma adverbial

28578 Era una familia humilde, pero **comían** como reyes. (anotación) (más info.)

28579 Sin caer nunca en la gula, gustaba del buen comer: "Viene ahora el Hijo del Hombre, que **come** y bebe, y dicen:" i He aquí un tragón y bebedor! (anotación) (más info.)

2.11 irudia: *comer* aditzaren lehen adierako bi adibideren etiketatzea.

2.1.3 Rol-zerrenda estandarraren bila: LIRICS

Rol semantikoak definitzeko eta etiketatzeko proiektuen artean, **LIRICS** proiektua⁶⁰ aipatu behar dugu (Bunt eta Romary, 2002). Proiektu horren helburua corpusak etiketatzeko ISO estandarrak adostea eta proposatzea da, estandar horiek erabilia, hizkuntza askotan eta askotarikotan erabilgarri izango diren baliabideak sortzeko asmoz.

LIRICSen 29 rol semantiko proposatu dira (Schiffrin eta Bunt, 2007; Bunt *et al.*, 2007), eta rol-zerrenda hori ebaluatzeko lau hizkuntzetako (ingelesa, nederlandera, italiara eta gaztelania) 500 esaldi etiketatu dira (Petukhova eta Bunt, 2008). Rol-zerrenda hori VerbNet-eko zerrendaren oso antzekoa da, baina, rolen definizioari dagokionez, badira desberdintasun nabarmenak. VNko rolak ez dira kontzeptu semantikoak, eta egitura sintaktiko eta lexikalen bidez definitzen dira (ikusi beheko aipua), hau da, egitura sintaktikoa

⁶⁰Linguistic InfRastructure for Interoperable resourCes and Systems (<http://lirics.loria.fr>)

kontuan hartzen da rola definitzerakoan. Hori dela eta, ez dira argi geratzen rolen arteko desberdintasun semantikoak. LIRICSen, aldiz, rolak kategoria semantiko gisa definitzen dira, eta ez zaio rol horien gauzatze sintaktikoari arreta jartzen. Rolak parte-hartzaileen eta ebentoen arteko erlazioak definitzen dituzten kontzeptuak dira⁶¹:

“VN makes use of (1) commonly used, coarse-grained roles like those of LIRICS (e.g. Agent), (2) roles that are specific to certain classes of events, which are intended to convey key semantic components of some verb classes (e.g. Topic, which is restricted to verbs of communication) (3) roles that are in part syntactically motivated (e.g. Predicate, which is used for classes with predicative complements) (4) roles that are distinguished by internal properties of the participant (e.g. [+animate]) [1], [2]. In contrast, LIRICS uses only roles that are not restricted to specific verb classes, not linked to particular syntactic structures, and not related to internal properties of participants” (Bonial *et al.*, 2011b: 484).

Esan daiteke, beraz, rolen definizioari dagokionez FrameNet eredutik hurbilago daudela, semantikan baino ez direlako oinarritzen, baina rolen zerrendari dagokionez, aldiz, VN eredutik gertuago (ikusi 2.1 taula).

2.1 taulan orain arte aipatu ditugun baliabideen arteko rol-zerrenden konparaketa ikus daiteke. Taula [Petukhova eta Buntten \(2008\)](#) lanetik hartu dugu.

VerbNet	PropBank	FrameNet	LIRICS
Agent	Arg0, Arg1	Agent, Speaker, Cognizer, Communicator, Ingestor, Deformer, etc.	Agent
Actor	Arg0	Avenger, Communicator, Item, Participants, Partners, Wrongdoer	Agent
Actor1	Arg0	Arguer1, Avenger, Communicator, Interlocutor1, Participant1, etc.	Agent

(jarraipena hurrengo orrialdean)

⁶¹Alde horretatik [Aldezabalen \(2004\)](#) roletatik hurbilago daudela esan behar dugu.

VerbNet	PropBank	FrameNet	LIRICS
Actor 2	Arg1, Arg2	Addressee, Arguer2, Injured party, Participant2, Partner2	Partner
Attribute	Arg1, Arg2	Attribute, Dimension, Extent, Feature, etc.	Attribute
Beneficiary	Arg1, Arg2, Arg3, Arg4	Audience, Beneficiary, Benefitted party, Goal, Purpose, Reason, Studio	Beneficiary
Cause	Arg0, Arg1, Arg2, Arg3	Addressee, Agent, Cause, Communicator, etc.	Cause
Destination	Arg1, Arg2, Arg5	Addressee, Body part, Context, Goal, etc.	Final_Location
Experiencer	Arg0, Arg1	Cognizer, Experiencer, Perceiver, etc.	Pivot
Extent	Arg2	Difference, Size change	Amount, Distance
Instrument	Arg2	Agent, Fastener, Heating instrument, Hot cold source, etc.	Instrument
Location	Arg1, Arg2, Arg3, Arg4, Arg5	Action, Area, Fixed location, etc.	Location
Material	Arg1, Arg2, Arg3	Components, Ingredients, Initial entity, Original, Resource, Undergoer	Source
Patient	Arg0, Arg1, Arg2	Addressee, Affliction, Dryee, Employee, Entity, Executed, etc.	Patient
Patient1	Arg0, Arg1	Concept1, Connector, Fastener, Item, Item 1, Part 1, Whole patient	Pivot
Patient2	Arg2, Arg3	Concept 2, Containing object, Item 2, Part 2	Patient

(jarraipena hurrengo orrialdean)

VerbNet	PropBank	FrameNet	LIRICS
Predicate	Arg1, Arg2	Action, Category, Containing event, etc.	-
Product	Arg1, Arg2, Arg4	Category, Copy, Created entity, etc.	Result
Proposition	Arg1, Arg2	Act, Action, Assailant, Attribute, etc.	-
Recipient	Arg1, Arg2, Arg3	Addressee, Audience, Authorities, Recipient	Goal
Stimulus	Arg1	Emotion, Emotional state, Phenomenon, Text	Theme
Theme	Arg0, Arg1, Arg2	Accused, Action, Co-participant, Co-resident, Content, Cotheme, etc.	Theme
Theme1	Arg0, Arg1	Cause, Container, Phenomenon 1, Profiled item, Theme	Pivot
Theme2	Arg1, Arg2, Arg3	Containing object, Contents, Cotheme, etc.	Theme
Time	ArgM_TMP	Time	Time
Topic	Arg1, Arg2	Act, behavior, Communication, Content, etc.	Theme
Asset	Arg1, Arg3	Asset, Category, Measurement, Result, Value	Amount
Value	Arg1	Measurement, Result, Value, Asset, Category	Amount
Source	Arg2, Arg3	Role, Victim, Patient, Source, Paht start, etc.	Initial_Location
-	-	Setting, Containing event	Setting
-	-	Means	Means
-	ArgM_MNR	Manner	Manner

(jarraipena hurrengo orrialdean)

VerbNet	PropBank	FrameNet	LIRICS
-	ArgM_PRP	Purpose	Purpose

2.1 taula: Rol semantikoak proiektu desberdinetan.

Baliabide bakoitzak erabiltzen dituen rola ikusita, argi dago bakoitzak bere zehaztasun-maila (*granularity*) daukala. PBk erabiliko lituzke rolik orokorrenak argumentu zenbakituak roltzat hartuta, baina esan beharra daukagu gu ez gaudela horrekin ados, guretzat argumentu zenbakitu horiek ez baitute inolako informazio semantikorik eskaintzen. Ondoren, LIRICSeko rol-zerrenda legoke. VNk bi maila lituzke (Bonial *et al.*, 2011b), lehenak LIRICSekin bat egingo luke eta bigarrena zehatzagoa litzateke⁶²; eta zehaztasun-mailarik finena FNk izango luke.

Azken urteotan, LIRICSeko eta VNko rol-zerrendak antzekoak zirela ikusita, bien arteko konparaketa egiteari ekin zaio (Bonial *et al.*, 2011a). Helburua hizkuntzaren prozesamenduko hainbat aplikaziotan erabilgarri izango den rol-zerrenda estandarra osatzea da. VNko rola LIRICSen aztertzen ari den ISO estandarrera hurbildu nahi dira, baliabideen arteko mapaketak errazago egin ahal izateko. 2.2 taulan, konparaketa egin ondoren VN-n hartutako erabakiak eta VNko 3.2 bertsioan dagoeneko adierazten direnak ageri dira:

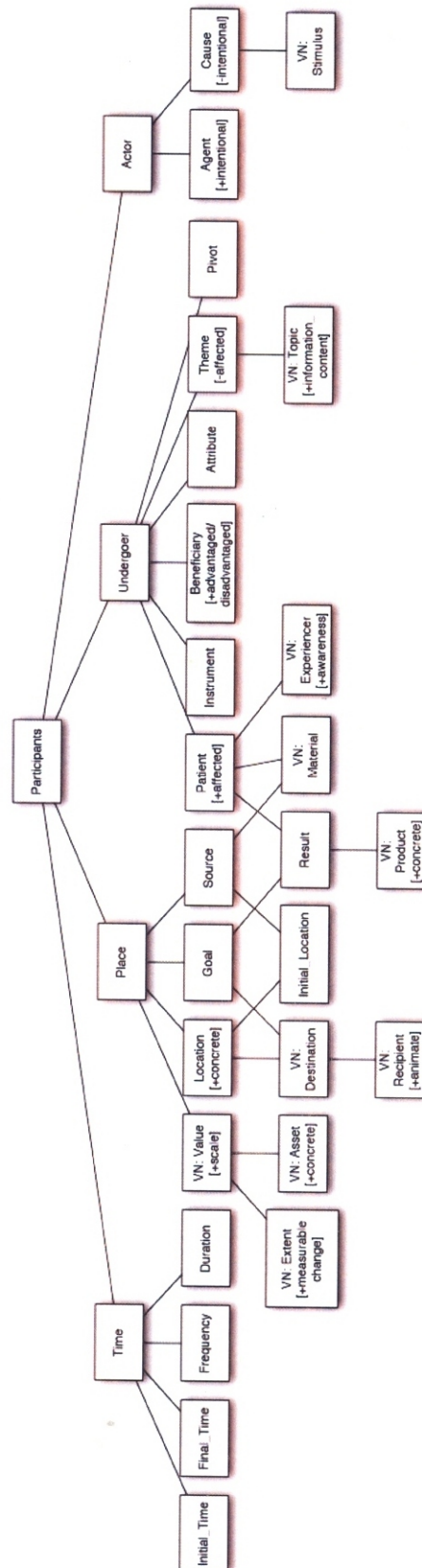
VN	LIRICS	VN 3.2
Actor1	Agent	Agent
Actor2	Partner	Co-agent
Patient1	Patient	Patient
Patient2	Partner	Co-patient
Theme1	Pivot	Theme
Theme2	Theme	Co-theme
Theme1 ⁶³	Pivot	Pivot
Theme2	Theme	Theme

2.2 taula: VN-n egindako rol-aldaketak konparaketa egin ondoren (Bonial *et al.*, 2011a).

⁶² *Theme* rola orokorra litzake eta gero, klase jakin batzuetako aditzei *Topic* rola jaritzen zaie (esaterako, *verbs of communication* klasekoei), baina biak dira funtsean gauza bera, hots, *gaia*. Orobat gertatzen da *Goal/Recipient* eta *Result/Product* rol-pareekin eta horregatik esaten dute VNk bi zehaztasun-maila dituela.

⁶³ *Theme1* eta *Theme2* aldatu egiten dira aditz-klasearen arabera.

VerbNet-eko eta LIRICseko rolak berdindu ondoren, rolen arteko hierarkia bat proposatzen dute [Bonial *et al.*ek \(2011a\)](#). Baliabide bakoitzak duen zehaztasun-maila ikusita, baliabide guztien rolak hierarkikoki antola daitezkeela ikusi da. [2.12](#) irudian ikus daiteke hierarkia hori:



2.12 irudia: LIRICS eta VNko rol bateratuen hierarkia (Bonial *et al.*, 2011b).

HPko aplikazio bakoitzak zehaztasun-maila desberdina behar izaten duela kontuan hartuta, oso baliagarria da rol semantikoen hierarkia bat izatea. Horrela, zehaztasun-maila desberdina duten baliabide semantikoen arteko loturak eta HPko aplikazioen arteko elkarreragina ahalbidetzen delako.

Rol semantikoen etiketatzearen inguruan egin diren proiektuak aztertu ondoren, gure lanerako zein eredu hautatu dugun eta zergatik azalduko dugu.

2.2 Hautuaren arrazoiak

Gure abiapuntua sintaktikoki eta dependentzia-ereduari jarraiki etiketatuta-ko corpusa da. Corpusa sintaktikoki etiketatuta izateak argumentu/adjuntu hautagaiak zeintzuk diren alde aurretik jakitea ahalbideratzen digu eta, beraz, abiapuntu egokia da hautagai horiei informazio semantikoa gehitzeko.

Bestalde, [Aldezabalek \(2004\)](#) [Levinen \(1993\)](#) klaseetan oinarrituta egin-dako euskarazko 100 aditzen azterketa dugu, eta hor aditzak egitura sintaktiko-semantikoaren arabera antolatzen dira. Beraz, semantika bai, baina sintaxia ere aintzat hartuko duen eredu behar dugu.

2.1 atalean, gaztelaniarako eta katalanerako proposatu diren ereduak (ADESSE eta SENSEM) azaldu ditugu, baina eredu horiek ez dira beste hizkuntza batzuetarako masiboki erabili. Horrenbestez, alde batera utzi ditugu eta ingeleserako garatutako baliabideetan oinarritu gara, informazioa berrerabiltzeko eta hizkuntzak konparatzeko aukera handiagoa emango digutelakoan.

Lehen ikusi dugun bezala, FrameNet-en abiapuntua ez da aditza, baizik eta markoa edo *framea*, hau da, abiapuntua semantikoa da erabat eta ez du sintaxia kontuan hartzen informazio semantikoa adierazteko eta interpretatzeko; beraz, ez dirudi gure abiapuntutik hurbil dagoenik.

LIRICSen kasuan ez dugu corpus etiketaturik (500 esaldi baino ez dira etiketatu hizkuntza bakoitzean). Proiektu horren helburua ez da corpus etiketatua eta lexikoa osatzea, baizik eta rolen zerrenda estandarra definitzea (ISO estandarren arabera). Oraindik ere aztertze-prozesuan dagoen proiektua da eta, beraz, hortik ateratzen diren ondorioak kontuan izan beharrekoak badira ere, ezin gara oraingoz eredu horretan oinarritu gure lana aurrera eramateko.

Azkenik, PropBank eta VerbNet ditugu. PB sintaktikoki etiketatuta-ko corpus batetik abiatzen da; VNk Levinen klaseetan eta, beraz, egitura sintaktiko-semantikoan oinarrituta antolatzen ditu aditzak, [Aldezabalek](#)

(2004) egiten duenaren antzera. Badirudi, beraz, egitasmo horien helburuak eta gureak bat datozela, eta eredu egokia izan daitekeela gure lana garatzeko.

Horrenbestez, hiru arrazoi nagusi aipatu nahi ditugu PB-VN ereduaren aldeko hautua azaltzeko:

- Abiapuntutzat sintaktikoki etiketatutako corpora du.
- Beste hizkuntza batzuetarako ere erabili izan da: hindia, txinera, ko-reera, arabiera, gaztelania, katalana, frantsesa, Brasilgo portugesa eta nederlandera.
- IXA taldean aldez aurretik egindako lanen antzeko ildoak proposatzen dira aditzak tratatzeko.

Ohar bedi, beraz, PropBank eta VerbNet, biak, izango ditugula oinarritzat. Orain arte beti aipatu izan da PropBank eredu bakarrik, eta horretan oinarrituta egin dira bestelako hizkuntzetarako PropBank-ak ere. Guk geuk ere hasiera batean, PropBank eredu jarraituko genuela esaten genuen, baina ikusi genuen PropBank eta VerbNet osagarriak direla elkarren artean. VerbNet-ek aditzetan rol semantikoak orokortzeko aukera ematen du, eta PropBank-ek, aldiz, hobeto jasotzen ditu argumentuen egitura sintaktikoen arteko desberdintasunak (Merlo eta Van der Plas, 2009). VerbNet-en erabiltzen den rol semantikoen multzoa, hizkuntzalaritzan, orokorrean, erabili izan denaren antzekoa da (eta, bide batez, Aldezabalek (2004) erabiltzen duenaren antzekoa ere), eta PropBank-ek, ordea, balentzia berberak (Arg0, Arg1, Arg2...) aditz guztientzat eta aditz bakoitzeko rol espezifikoak (*buyer, thing bought, sayer, utterance, hearer...*) erabiltzen ditu. PropBank lotuago dago funtzio gramatikalekin (subj., obj., zobj....) eta aproposagoa da ikasketa automatikorako, baina VerbNet-ek, erabiltzen dituen rol semantiko orokorrei esker, informazio semantiko baliagarriagoa ematen digu. Beraz, uste dugu biak erabilia lortuko dugula hasieran aipatu dugun helburua, hots, aditzaren informazio sintaktiko-semantiko osoa baliabide bakar batean biltzea.

2.3 Informazio semantikoa txertatzeko etiketa: *arg_info*

Eredua hautatuta, etiketatzea nola formalizatu izan da hurrengo pausoa. Formalizazio horretarako *arg_info* etiketa sortu dugu, honako bost eremu hauek dituela:

- **VNrs** (PB-VN aditza eta *rolset* zenbakia): PB-VNko aditza eta horri dagokion *rolset* zenbakiarentzako eremua. Askotan, lehen eremu honetan PB-VNko ordain bat baino gehiago egongo da, esaterako: *tell.01/say.01*; edo *separate.01/divide.02*. Ingelesezko ordain desberdin (edo ordain multzo desberdin) guztiek euskarazko adiera bakarra izango dutela adierazten dugu horrela. Adibidez: *esan_1 = tell.01/say.01*; *esan_2 = call.01*; *onartu_1 = accept.01/allow.01/approve.01*.
- **BAL** (balentzia): argumentu edo adjuntuentzako eremua. Balioak ondokoak dira: Arg0, Arg1, Arg2, Arg3, Arg4, ArgMod (ArgM). Aztergai den unitateari horietako zein dagokion zehazten da.
- **VNrol** (VerbNet-eko rola): VNri dagokion rola. Balioak, printzipioz, 2.4 atalean aipatu ditugunak izango dira⁶⁴.
- **EADBrol** (Euskal Aditzen Datu Baseko rola): EADBko rola ezartzeko eremua. Balioak ondorengo 21 rolak dira: *gaia*, *gai ukitua*, *gai sortua*, *egoera*, *kokapena*, *denbora*, *helburuko kokapena*, *helburuko egoera*, *abiapuntuko kokapena*, *bidea*, *abiapuntua*, *helburua*, *esperimentatzaileria*, *kausa*, *iturria*, *edukitzailea*, *edukia*, *ezaugarria*, *jarduera*, *neurria*, *modua*⁶⁵.
- **HM** (Hautapen-murriztapenak): Hautapen-murriztapenak jartzeko eremua. Balioak, printzipioz, ondorengoak dira: [+biz], [-biz], [+konkr], [-konkr], [+giz], [-giz].

Nahiz eta rolen etiketatzearen oinarria sintaktikoki etiketatutako corpusa izan, *arg_info* etiketa independenteki erabili ahal izatea nahi dugu. Horregatik, bigarren eta hirugarren eremuak gehitu ditugu, hau da, lantzen ari garen aditza (burua, alegia) eta aztertzen ari garen argumentua/adjuntua, hurrenez hurren. Bi balio horiek automatikoki kopiatuko dira etiketatze sintaktikotik, lan errepikakorra ekidinez. Horrela, halaxe geratuko zaigu *arg_info* etiketa:

- **arg_info**: (VNrs, aditza, hitz aztergaia, BAL, VNrol, EADBrol, HM)

Esaterako, *zer esan nian nik hiri?* esaldiko *nik* elementuari (28) adibidean ikusten den *arg_info* etiketa legokioke:

⁶⁴Adjuntuen kasuan, eremu honetan jarri ditugu PBk definitzen dituen adjuntuen (ArgM) balioak (MNR, TMP, LOC, ADV...), VN-n adjuntuen rolak ez baitira zehazten.

⁶⁵EADBN agertzen diren rol semantiko horien definizioak 4.1.1 atalean ikus daitezke.

- (28) **arg_info**: (tell.01/say.01, esan, nik, Arg0, Agent, esperimentatzaileria, [+giz])

Ikus ditzagun, (29) adibidean, sintaktikoki etiketatuta datorkigun esaldi baten *arg_info* guztiak; etiketa hau aditzaren mende dagoen dependenzia-etiketa bakoitzaren ondoren erantsiko da, adibidean berdez nabarmendu ditugu.

- (29) *Egiptora Axtarteren hilobia bilatzera zihoala esan zion norbaiti.*

ncmod (ala, zihoala, **Egiptora**, Egiptora)
arg_info (go.01, zihoala, Egiptora, Arg2, Destination, helburuko kokapena, -)
ncmod (gen, hilobia, Axtarteren, Axtarteren)
ncobj (abs, bilatzera, **hilobia**, hilobia, obj)
arg_info (look.05, bilatzera, hilobia, Arg1, Theme, gaia, -)
xmod (helb, zihoala, **bilatzera**, bilatzera)
arg_info (go.01, zihoala, bilatzera, ArgM, PRP, -, -)
ccomp_obj (konpl, esan, **zihoala**, zihoala)
arg_info (say.01/tell.01, esan, zihoala, Arg1, Topic, gaia, -konkr)
auxmod (-, esan, zion)
nczobj (dat, esan, **norbaiti**, norbaiti, zobj)
arg_info (say.01/tell.01, esan, norbaiti, Arg2, Recipient, -, -)

2.4 Laburbilduz

Kapitulu honetan EPEC corpora predikatu-mailan etiketatzeko eredua nola eta zergatik aukeratu dugun azaldu dugu.

Lehenik eta behin, gai honen inguruko **aurrekariak** aztertu eta landu izan diren bi ikuspegi nagusiak azaldu ditugu: i) Fillmore-ren kasuen teoriarik (Fillmore, 1968) oinarritzen den rol semantikoaren zerrenda klasikoa (*agentea, kausa, gaia, onuraduna, helburua, lokazioa...*), eta ii) Dowtyren (1991) protorolen teoriatik eratortzen den ikerlerroa, non bi rol nagusi baino ez diren definitzen (*proto-agent* eta *proto-patient*).

Ondoren, rol semantikoen inguruan euskaraz egin den lana aztertu dugu, eta orain artean ezer gutxi egin dela ikusi dugu. Aditzen inguruan egingako azterketa gehienak aditz laguntzailean oinarritu izan dira eta horren arabera sailkapenak proposatu izan dituzte. Gure helburua corpusa predikatu-mailan etiketatzea eta aditzen lexikoa sortzea da; beraz, bakarrik aditz laguntzailea kontuan hartzen duten sailkapen horiek ez dira nahikoa guretzat. Guk aditzaren informazio sintaktiko-semantiko orokorra behar dugu, eta horren inguruko azterketa bakarra Aldezabalen tesian (Aldezabal, 2004) aurkitu dugu; beraz, han egindako euskarazko 100 aditzen azterketatik abiatu gara corpusa etiketatzeko eta lexikoa osatzeko.

Bukatzeko, gure helburuak erdiesteko eredu egokiaren bila, beste hizkuntzetan corpus etiketatutako eta horietan oinarritutako lexikoiak sortzeko egin diren proiektuak aztertu ditugu. Ingeleseko PB corpusa (beste hizkuntza askotarako ere erabili izan dena) (Kingsbury eta Palmer, 2002; Palmer *et al.*, 2005) eta VN lexikoa (Kipper *et al.*, 2000; Kipper, 2005) aztertu ditugu zehatz. Marko semantikoetan oinarritzen den FrameNet proiektua (Baker *et al.*, 1998; Fillmore eta Baker, 2001) ere aztertu dugu. Gaztelaniaz eta katalanez egin diren corpusak eta lexikoiak deskribatu ditugu: ADESS (García-Miguel eta Albertuz, 2005) eta SENSEM (Castellón *et al.*, 2006; Vázquez *et al.*, 2006), eta, azkenik, azken urteetan rol semantikoen zerrenda estandartzeko sortutako LIRICS proiektua (Bunt eta Romary, 2002) ere aurkeztu dugu.

HParen alorrean aditzen inguruan egin diren proiektuak aztertu ondoren, gure helburuak aurrera eramateko egokiena **PropBank-VerbNet eredu**a dela erabaki dugu. Alde batetik, beste hizkuntza askotan erabili izan da masiboki, eta horrek informazioa berrerabiltzeko eta hizkuntzak konparatzeko aukera ematen digu. Bestetik, sintaktikoki etiketatutako corpus batean oinarritzen da, eta hori abiapuntu egokia da etiketatze semantikoa egiteko, argumentu/adjuntu hautagaiak alde aurretik markatuta ditugulako. PropBank eta VerbNet biak erabiltzea erabaki dugu, informazio osagarria ematen dutelako. VerbNet-ek erabiltzen dituen rol semantiko orokorreari esker, informazio semantiko baliagarriagoa ematen digu; PropBank-ek, aldiz, hobeto jasotzen ditu argumentuen egitura sintaktikoen arteko desberdintasunak. Beraz, biak batera erabilia lortzen dugu aditzaren informazio sintaktiko-semantiko orokorra jasotzea.

Alde batera utzi ditugu, beraz, ADESEE eta SENSEM proiektuak ez direlako beste hizkuntza batzuetarako erabili, bai eta FrameNet proiektua ere, ez duelako sintaxia kontuan hartzen eta bere abiapuntua ez delako aditza,

baizik eta marko semantikoa (*framea*).

LIRICS proiektuaren helburuek ez dute gure helburuekin bat egiten. LIRICSen asmo nagusia HPko hainbat aplikaziotan erabilgarri izango den rol-zerrenda estandarra sortzea da. Beraz, ez dugu kontuan hartu eredia aukatzean. Hala ere, esan behar dugu rol semantikoak bateratzeko egin duten saiakera oso baliagarria dela baliabideen arteko mapaketak errazteko, eta egin dituzten proposamenak VerbNet-en jasota daudela.

Kapitulua bukatzeko, eta eredia hautatu ondoren, etiketatzea formalizatzekeo sortu dugun *arg_info etiketa* aurkeztu dugu. Etiketa horretan aditzaren informazio sintaktiko-semantiko osoa jasotzen da eta independenteki erabili ahal izateko pentsatua da. Bertan PBko, VNko eta EADBko informazioa biltzen da.

PB-VN eredua jarraitzeko erabaki nagusiak

Eredua aukeratu eta aztertu arren, beti da beharrezkoa hizkuntza jakinera egokitzeko irizpideen gida; izan ere, ereduak ia inoiz ez dira zuzenean inolako moldaketarik gabe aplikatzen. Hasteko, esan behar dugu **gure irizpide orokorra PropBank jarraitzea** izan dela, hots, PropBank-ek egiten duen argumentu/adjuntu bereizketa eta balentzia onartzea. Ondoren, argumentu horiei VNk ematen dizkien rolak esleitu dizkiegu, bien informazioa batuz.

Jarraian, etiketatzeko irizpide nagusi hori nola aplikatu dugun azalduko dugu hainbat adibide erakutsita.

Demagun *ahaztu* aditza lantzen ari garela eta (30) adibidea etiketatu behar dugula.

(30) *Frank de Boerrek nekez ahaztuko du Eurokopako finalerdian bi penalti gelditu zizkion atezaina.*

Ahaztu aditzaren ingelesezko ordaina *forget* izan daiteke eta PBn *forget* bilatuz gero, 3.1 taulan ikusten den informazioa aurkitzen dugu:

forget.01

Arg0	forgetter
Arg1	thing forgotten

3.1 taula: *forget.01* adiera PBn.

Beraz, *forget* aditzak adiera bakarra du ingelesez eta bi argumentu: i) ahazten duen pertsona eta ii) ahazten den gauza.

Etiketatzeko sintaktikoari esker, (30) adibidean *ahaztu* aditz nagusiaren mendean hiru elementu ditugula ikusten dugu:

1. *Frank de Boerrek (ahaztuko du)*
2. *nekez (ahaztuko du)*
3. *Eurokopako finalerdian bi penalti gelditu zizkion atezaina (ahaztuko du)*

Hiru elementu horietatik bik PBko argumentuekin bat egiten dute:

- *Frank de Boerrek >forgetter*
- *Eurokopako finalerdian bi penalti gelditu zizkion atezaina >thing forgotten*

Beraz, bi elementu horiek PBk esaten duenaren arabera etiketatuko ditugu:

- Arg0: *Frank de Boerrek*
- Arg1: *Eurokopako finalerdian bi penalti gelditu zizkion atezaina*

(30)-eko hirugarren elementua (*nekez*) ez dago definitua *forget* aditzaren sarreran, eta, beraz, modifikatzailea (ArgM) dela ondoriozta dezakegu. Modifikatzaileak etiketatzeko etiketen zerrenda bat ematen dute PBko gidarlerroetan (horien laburdurak Laburtzapenak kapituluko Laburdurak atalean ingelesez eman ditugunak dira). Kasu honetan, *nekez* moduzko adberbioa da, eta, beraz, ArgM-MNR bezala etiketatuko dugu:

- Arg0: *Frank de Boerrek*
- Arg1: *Eurokopako finalerdian bi penalti gelditu zizkion atezaina*
- ArgM-MNR: *nekez*

Esan bezala, eta PBk egiten duenaren antzera, argumentu hauei VNk *forget* aditzarentzat ematen dituen rolak esleituko dizkiegu:

- Agent
- Theme

Beraz, PBko eta VNko, hau da, balentziaren eta rol semantikoen informazioa bilduta, horrela geratuko litzateke (30) adibidearen etiketatzea:

- *Frank de Boerrek*: Arg0, *Agent*
- *nekez*: ArgM, MNR
- *Eurokopako finalerdian bi penalti gelditu zizkion atezaina*: Arg1, *Theme*

Adibideekin jarraituz, eman dezagun *entzun* aditza landu behar dugula eta (31) adibidea aurkitzen dugula:

- (31) *Kurt-ek harridura handiz entzuten zuen Rollin komisarioa kondatzen ari zena.*

Euskarazko *entzun* aditzaren ingelesezko ordainak *hear* eta *listen* izan daitezke, eta, aditz horiek PBn bilatuz gero, 3.2 eta 3.3 tauletan ikusten dugun informazioa topatzen dugu:

hear.01

Arg0	hearer
Arg1	utterance, sound
Arg2	speaker, source of sound

3.2 taula: *hear.01* adiera PBn.

listen.01

Arg0	listener
Arg1	sound or speaker

3.3 taula: *listen.01* adiera PBn.

(31) adibidean hiru elementu ikusten ditugu *entzun* aditzaren mendean:

1. *Kurt-ek (entzuten zuen)*
2. *harridura handiz (entzuten zuen)*

3. *Rollin komisarioa kondatzen ari zena (entzuten zuen)*

PBn hiru argumentu ikusi ditugu *hear* aditzean, baina (31) adibidean horietatik bi baino ez dira agertzen:

- *Kurt-ek >hearer/listener*
- *Rollin komisarioa kondatzen ari zena >utterance, sound/sound or speaker*

Hirugarren elementua (*harridura handiz*), (30) adibidean bezala, modifikatzailea da, baina kasu honetan ez da adberbioa, baizik eta modua adierazten duen postposizio-sintagma. Beraz, PBko gidalerroei jarraiki, ArgM-MNR etiketa jarriko diogu eta horrela geratuko da etiketatzea:

- Arg0: *Kurt-ek*
- Arg1: *Rollin komisarioa kondatzen ari zena*
- ArgM-MNR: *harridura handiz*

PBko informazioari VNkoa gehituta, horrela geratuko litzateke (31) adibidearen etiketatzea:

- *Kurt-ek*: Arg0, *Agent*
- *harridura handiz*: ArgM, MNR
- *Rollin komisarioa kondatzen ari zena*: Arg1, *Theme*

(31) adibidean ez da PBko *hear.01* sarreran definitzen den Arg2 (*speaker, source of sound*) argumentua agertzen, baina (32) adibidean ikusten den bezala, posible da euskaraz argumentu hori azaltzea, eta kasu honetan datiboan gauzatzen da:

- (32) *Geroztik Ruper Ordorikari entzun nion beste kanta batean, ondorio berberera eraman ninduen beste zerbait.*

(32)-an bada *speaker* bat (*Ruper Ordorika*). PBko *hear.01* sarreran zehazten den bezala, elementu horrek Arg2 etiketa eraman behar du eta VNk esleitzen dion rola *Source* da.

- *Ruper Ordorikari*: Arg2, *Source*

Gainontzean, beste hiru elementu daude *entzun* aditzaren mendean:

- *Geroztik* (*entzun nion*)
- *beste kanta batean* (*entzun nion*)
- *ondorio berberera eraman ninduen beste zerbait* (*entzun nion*)

Aurreko adibideetan azaldu bezala, eta PB-VN ereduari jarraiki, honela etiketatuko genuke adibide osoa:

- *Geroztik*: ArgM, TMP
- *Ruper Ordorikari*: Arg2, *Source*
- *beste kanta batean*: ArgM, LOC
- *ondorio berberera eraman ninduen beste zerbait*: Arg1, *Theme*

(32) adibidean ikusten dugu (31)-ean ikusten genuen Arg0_ *Agent* argumentua ez dela esplizitu agertzen, eta, beraz, ez dugu etiketatzen.

Etiketatzeko irizpide orokorra nola jarraitzen dugun azalduta, ereduan hainbat hutsune eta eztabaidagai diren alderdi ere aurkitu ditugula esan behar dugu. Horien aurrean guk hartutako erabakiak azalduko ditugu ondoren.

3.1 Arg0 eta Arg1 argumentuen inguruan

2.1.2.2 atalean aipatu dugunez, PropBank-ek bi maila independente bereizten ditu: argumentuak eta rolak. Baina euren etiketatzea aztertuta, ikusten dugu Arg1 beti dela *Theme* eta Arg0 *Agent*. Ez dirudi arrazoi linguistiko garbirik dagoenik hori horrela izateko. Ikusi, esaterako, [Kingsbury eta Palmerrek \(2003\)](#) ematen duten argudioa:

“Arg0 is very consistently assigned an Agent-type meaning, while Arg1 has a Patient or Theme meaning almost as consistently. There are, of course, many verbs in English for which the Patient, the entity undergoing the action of the verb, always appears in subject position. For

these verbs no agent is possible. In order to maintain the consistency of Arg1 as Patient these verbs have no Arg0. A canonical example is *fall* as seen in Figure 1:

fall.01 sense: move downward
 roles:
 Arg1: thing falling
 Arg2: extent, distance fallen
 Arg3: start point
 Arg4: end point

Figure 1” (Kingsbury eta Palmer, 2003: 3).

Antzeko zerbait aipatzen dute urte batzuk beranduago Loper *et al.*ek (2007) ere, hau da, konsistentzia mantentzeko hartutako erabakia izan dela:

“There is no guarantee that an argument label will be used consistently across different verbs. [...]. Generally, the arguments are simply listed in the order of their prominence for each verb. However, an explicit effort was made when PropBank was created to use Arg0 for arguments that fulfill Dowty’s criteria for ‘prototypical agent’, and Arg1 for arguments that fulfill the criteria for ‘prototypical patient’” (Loper *et al.*, 2007: 2-3).

Hala ere, eta konsistentziaren izenean hartutako erabakia dela aitortzen badute ere, hainbat inkonsistentzia topatu ditugu gidalerroetan idatzi dutenaren eta etiketatu dutenaren artean. Adibidez, hona zer dioten Babko-Malaya *et al.*ek (2006):

“In *John and Mary come* the NP *John and Mary* is a constituent in Treebank and it is also marked as Arg0 in PropBank” (Babko-Malaya *et al.*, 2006: 76).

Baina PropBank-en zuzenean begiratuta, 3.4 taulan agertzen dena ikusten dugu, hau da, etortzen den elementua (*entity in motion*) Arg1 da eta ez Arg0 Babko-Malaya *et al.*en (2006) goiko aipuan esaten den bezala.

come.01

Arg1	entity in motion (theme)
Arg2	extent
Arg3	start point
Arg4	end point

3.4 taula: *come.01* adiera PBn.

Inkonsistentzia horiek ikusita, gure erabakia bi mailen arteko independentzia mantentzea izan da, PropBank-eko gidalerroetan esaten dutenari jarraiki, eta, beraz, ez ditugu automatikoki Arg0 *Agent* gisa eta Arg1 *Theme* gisa etiketatu.

Gure ustez, eta (33) adibidean ikusten den bezala, kokapen-aldaketa adierazten duten aditz iragangaitzen subjektua aldi berean da ekintza burutzen duen entitatea eta ekintza hori jasotzen duena (Vázquez *et al.*, 2000: 183).

(33) *Mikel*_{Arg1} *eskolara etorri da gaur.*

Mikel aldi berean da *etorri* aditzak adierazten duen kokapen-aldaketa burutzen duena eta kokapen-aldaketa jasaten duena.

Beraz, aditz horien subjektuak Arg0 balentziarekin etiketatuko ditugu. Erabaki hori PropBank-en gidalerroetan bertan agertzen den arauarekin bat dator (*Choosing Arg0 versus Arg1* atalean):

“Whereas for many verbs, the choice between Arg0 or Arg1 does not present any difficulties, there is a class of intransitive verbs (known as verbs of variable behavior), where the argument can be tagged as either Arg0 or Arg1. (...) Arguments which are interpreted as agents should always be marked as Arg0, independent of whether they are also the ones which undergo the action. (...) In general, if an argument satisfies two roles, the highest ranked argument label should be selected, where Arg0 » Arg1 » Arg2 »...” (Babko-Malaya, 2005: 4).

Honela, *etorri* gisako aditz inakusatibo baten kasuan, zeinetan aldaera iragangaitza baino ez den posible, ekintza burutzen duena eta jasotzen duena entitate bera direla ulertzen dugu, eta, beraz, Arg0 *Theme* etiketak jarri dizkiogu. Gure kasuan, Arg0 izango dela erabakita, *Agent* rola ere jarri genezakeen, baina gure ustez (Taulé *et al.*ekin (2011) bat eginez), *Agent* rol semantikoa egokiagoa da kanpora zuzendutako ekintza gauzatzen duen subjektuarentzat:

“Arg0 is associated to the external causer argument of the verb” (Taulé *et al.*, 2011: 19).

PropBank-en, aldiz, mota horretako aditzen subjektuei Arg1 balentzia jartzen diete, eta, ondorioz, *Theme* etiketa.

Bestalde, *apurtu* bezalako aditz kausatibo/inkoatiboetan, *Theme* rolari beti jarriko diogu Arg1 balentzia, iruditzen zaigulako esaldian agertu ez arren, ekintzaren eragilea den *kausa* (Arg0) beti existitzen dela (34).

(34) *Leihoa*_{Arg1} *apurtu da* [zerbaiten eraginez.]

Esan beharra dago PropBank eredua erabili duten beste hizkuntzetan Arg0_ *Agent*, Arg1_ *Theme* irizpidea jarraitu dutela.

PropBank eredua zuzenean jarraitu ez duten proiektuen artean ADESSSE (García-Miguel eta Albertuz, 2005) eta SENSEM (Vázquez *et al.*, 2006) ditugu. ADESSSE corpusean, PropBank-en antzera argumentu zenbakituak erabiltzen dituzte, baina beren webgunean⁶⁶ azaltzen duten bezala, etiketa-tzea ez dute erabat PropBank-ek bezala egiten:

“En proyectos como PropBank o AnCora, en los que se recurre también a una numeración correlativa para anotar los argumentos de los verbos, se aprecia una tendencia más o menos generalizada a asignar el índice A0 a agentes, causantes y experimentadores, y a asignar el índice A1 a pacientes y temas (i.e. entidades que sufren un cambio de estado o que son afectados por la acción). Es decir, A0 se utiliza para anotar los sujetos de verbos transitivos y los sujetos de verbos inergativos; y A1 se utiliza para anotar los objetos de verbos transitivos y los sujetos de verbos inacusativos [...]. La estrategia aplicada en ADESSSE es ligeramente diferente. En ADESSSE se reserva el A0 para el primer argumento de eventos causativos, de forma que se puedan apreciar más fácilmente las correspondencias entre verbos causativos y verbos no causativos [...]. Por su parte, A1 se utiliza fundamentalmente para anotar experimentadores (e.g. el perceptor en **ver**) y agentes de eventos no causativos (e.g. el agente en **trabajar**), así como para entidades que son afectadas por la acción verbal de un evento causativo (e.g. el afectado en **romper**)”⁶⁷.

SENSEM corpusean, aldiz, ez dute arazo hori, ez baitituzte argumentu zenbakituak edo balentziak erabiltzen.

⁶⁶<http://adesse.uvigo.es/Docu/Rolesgeneralizados> (2012-12-10)

⁶⁷Ikusten den bezala, ADESSSEn ez diote PBk egiten duenari jarraitzen, baina beraiek eta guk hartutako erabakiak ere desberdinak dira.

3.2 VNk bi klase edo gehiago esleitzen ditu

Zenbaitetan, [Levinek \(1993\)](#) aditzak klase batean baino gehiagotan sailkatzen ditu. Horregatik, VN-n aditz batek Levinen klase semantiko bat baino gehiago izan dezake eta, gainera, VNk berak sortutako klaseak eta azpiklaseak ere gehi dakizkioke. Klase eta azpiklase bakoitzak bere rolak ditu, eta, ondorioz, VN-n aditzaren argumentu berari rol bat baino gehiago esleitzen zaizkio askotan.

3.5 taulan *separate.01* aditz-adiera ikus daiteke kasua hobeto ulertzeko:

separate.01	
vncls: 23.1-2, 10.1, 29.10	
Arg0	entity causing separation (vnrole: 23.1-2-Agent, 10.1-Agent, 29.10-Agent)
Arg1	thing being divided (vnrole: 23.1-2-Patient, 10.1-Theme, 29.10-Theme)
Arg2	source, separated from (vnrole: 23.1-2-Patient, 10.1-Source)

3.5 taula: *separate.01* adiera PBn.

3.5 taulan ikusten denez, *separate.01* aditz-adierari 3 klase esleitzen zaizkio⁶⁸ VN-n:

- **23.1-2:** Levinen lanean, 23 klasea *verbs of separating and disassembling* izenekoa da, eta horren barruan 23.1 klasea dago, *separate verbs*. VN-n ere 23 eta 23.1 klaseak berdin izendatzen dira, baina *separate* aditza *separate-23.1* klase nagusiaren 23.1-2 azpiklasean sartu dute.
- **10.1:** Levinen sailkapenean, 10 klasea *verbs of removing* da eta horren barruan dago 10.1 *remove verbs* klasea, VN-n ere berdin mantentzen dena.
- **29.10:** Levinen lanean, 29 klasean *verbs with predicative complements*, hau da, osagarri predikatiboak dituzten aditzak sartzen dira, baina 29.10 klasea ez da agertzen; beraz, *classify-29.10* klasea VN-n sortutakoa da.

Rolak klaseak errepresentatzeko erabiltzen direnez, klase bakoitzak rol desberdinak hartzen ditu; hala izanik, zein rolekin etiketatuko dugu guk

⁶⁸Hiru klaseak OntoNotes ([Pradhan et al., 2007](#)) baliabidearen adiera berarekin lotzen dira *Unified Verb Index*-en: *Sense number 1: (cause to) divide or segregate into parts, sections or smaller groups.*

gure *banandu* aditza? Horrelako kasuetan, EADBrekin bat datozen rolak aukeratzea erabaki dugu. Kasu honetan, *banandu* aditzaren lehen adierak (egoera-aldaketa), hau da, PB-VNko *separate.01* adierarekin bat datorrenak, *kausa* eta bi *gai ukitu* ditu EADBn:

- **DA-1:** gai ukitua (ABS), gai ukitua (ABL)
- **DU-3:** kausa (ERG), gai ukitua (ABS), gai ukitua (ABL)

Beraz, PB-VNko *Agent*, *Patient1*, *Patient2* aukeratuko ditugu, hau da, 23.1-2 klaseari dagozkion rolak (3.6 taula).

Arg0	agent	kausa (ERG)
Arg1	patient1	gai ukitua (ABS)
Arg2	patient2	gai ukitua (ABL)

3.6 taula: *banandu_1* adiera BVIn.

3.3 ADV rola

PB-VNko rolen zerrendan bada ADV izendatzen duten rol bat, baina askotan ez dago oso argi zein den haren erabilera. Guk erabaki dugu rol hau erabiliko dugula adberbio bat anbigua denean; hau da, adberbio kategoriakoa izanda, ez badugu argi denborazkoa (TMP: *atzo*, *berandu*, *gaur...*), moduzkoa (MNR: *nekez*, *horrela*, *ondo...*) edo lekuzkoa (LOC: *hemen*, *han*, *hurbil...*) den, ADV jarriko diogu adjuntu horri (35).

- (35) a. *Postu horretan ez du, nonbait, Miguel Angel Lotinak nahi zuena eskaini Alex Fernandezek.*

nonbait: ArgM, ADV

- b. *Houdaren familiak asko jaten du/Asko gustatzen zaio irakurtzea*⁶⁹.

⁶⁹Badakigu alde batetik, *asko jan* moduko adibideetan *asko* hori determinatzailea dela (*janari asko*), izenordain gisa erabilia, objektu inespezifiko modura, eta, bestetik, nahiz eta corpusean *asko gustatu* modukoak oso maiz agertu, horrelakoetan *oso* erabiltzea gomendatzen dela eta *asko* erabilera okerra dela. Guk corpusean agertzen denari erantzuna eman behar diogu, eta gure dependentzia-analisan modifikatzaile-sintagma gisa etiketatu dira biak eta ez determinatzaile-sintagma gisa. Beraz, adberbio gisa etiketatu ditugu guk ere.

asko: ArgM, ADV

c. *Nire txokorantz abiatu nintzen, bapatean argi bat ikusi nuen.*

bapatean: ArgM, ADV

3.4 Laburbilduz

Kapitulu honetan PB-VN eredia erabiltzeko modua dela-eta hartu ditugun erabaki nagusiak azaldu ditugu.

Gure irizpide nagusia PBk eta VNk esaten dutenari jarraitzea da, eta kapituluaren hasieran, etiketatzean irizpide hori nola gauzatzen den azaldu dugu adibide gutxi batzuen bitartez.

Jarraian, eredia aplikatzean hartu ditugun hiru erabaki nagusiak aurkeztu ditugu:

- Arg0/Arg1 argumentuen inguruko erabakia.
- VN-n argumentu berarentzat bi rol (edo gehiago) esleitzen direnean hartutako erabakia.
- Adjuntuen ADV rola etiketatzeko erabakia.

Lehenik eta behin, **Arg0/Arg1 argumentuen inguruan** aritu gara. PBk Arg0 argumentua *agenteari* (*Agent*) esleitzen dio beti, eta Arg1, berriaz, *gaiari* (*Theme*), baina ikusi dugu euren gidalerroen kontra doala hori, eta, gainera, ez dagoela arrazoi linguistikorik horrela jokatzeko. Gure erabakia gidalerroetan esaten dutenari jarraitzea da, hau da, argumentu batek bi rol bete baditzake, gorago dagoen argumentu-etiketa jarriko zaio, eta goren dagoen argumentu-etiketa Arg0 da. Horregatik, aditz inakusatiboan (esaterako *etorri*) subjektuari, aldi berean ekintza egiten duen eta jasotzen duen subjektuari, Arg0_Theme etiketak jarri dizkiogu.

Bigarrenez, **VNk aditz bati bi klase** (edo gehiago), eta, ondorioz, bi rol (edo gehiago) esleitzen dizkionean hartu dugun erabakiaz aritu gara. Horrelako kasuetan, EADBN dagoen rola begiratu, eta horretatik hurbilen dagoena aukeratzea erabaki dugu.

Bukatzeko, **adjuntuen ADV rolaz** jardun dugu. ADV rola adberbio kategoria duten elementuei jarri behar zaiela pentsa dezakegu, baina askotan adberbio horiek denborazkoak, moduzkoak edo lekuzkoak diren ere zehatz

dezakegu, TMP, MNR eta LOC etiketen bidez. Beraz, adberbio kategoria duten eta denborazkoak, moduzkoak edo lekuzkoak ez diren elementuei edo zalantzak sortzen dituztenei ADV etiketa esleituko diegu.

Etiketatzeko aurretikoak: baliabideak, aurreprozesuak eta tresna

Etiketatzeko-prozesuari ekiteko, taldean dagoeneko garatutako zenbait baliabide izan ditugu eskura; kapitulu honetan horiek azalduko ditugu (4.1). Baita etiketatzen laguntzeko garatu ditugun prozesu automatikoak ere (Aldezabal *et al.*, 2010c, a) (4.2). Bukatzeko, etiketatzeko erabili dugun tresna deskribatuko dugu laburki (4.3).

4.1 Baliabideak

Hasteko, tesi-lan honetan erabili ditugun alde aurretiko baliabideak azalduko ditugu, gure lanaren oinarri izan den Aldezabalen (2004) lanean arreta berezia jarritz.

4.1.1 Euskal Aditzen Datu Basea (EADB)

Euskal Aditzen Datu Basea (EADB) Aldezabalek bere tesi-lanean (Aldezabal, 2004) landutako 100 euskal aditzen datu-basea da. EADBren diseinua eta edukia haren tesian azaltzen dira sakonki. Aldezabalek bere lana Levinen (1993) eta Vázquez *et al.*en (2000) teoria eta irizpideak kontuan hartuz egina du oro har, eta, horretaz gain, corpuseko adibideak aztergai izanda. Levinen (1993) teoria egokituz eta Vázquez *et al.*en (2000) lanaren eskarmentua baliatuz, euskal aditzak lantzeko proposamen bat egiten du.

76 Etiketatzeko aurretikoak: baliabideak, aurreprozesuak eta tresna

Jarraian, proposamen horren oinarriak azalduko ditugu labur; sakontzeko, jo bedi haren lanera.

Aditz baten rol semantikoek aditzaren “esanahia” bereizten dute, esanahi hori predikatu-mota orokor gisa ulertuta, hots, aditzak sail orokorragoetan sailkatzeko bidea emango duten predikatu orokor gisa. Beraz, esanahia ez da hiztegiatan agertzen den modu xehean ulertu behar, baizik eta balio kognitibo orokor moduan (mugimenduzko balio kognitiboa, egoera-aldaketazkoa, sortzekoa...). Izan ere:

“Aditz guztien erabilerak aztertzerakoan (eta ez aditz bakarrarenak edota erabilera bakarra hainbat aditzetan) argi ageri da aditz bakoitzak egitura usu bat (egitura prototipiko bat) edo batzuk erakutsi ohi dituela, aditz guztietan errepikatzen diren balio semantiko orokorren bat adierazteko. Balio semantiko hauek ez dira adierak, hiztegiatan agertzen diren modukoak, baizik eta kategoria kognitibo oinarrikoak, predikatu-mota orokorrak, sailak osatzeko baliagarri izan daitezkeena” (Aldezabal, 2008: 38).

Ikusten den bezala, egitura usu bat edo batzuk ager ditzake aditzak balio bera adierazteko. Hain zuzen, aditz batean balio bera adierazten duten egitura sintaktiko desberdin guztiak elkarren artean **alternantzia** dira, eta alternantzia osatzen duten aldaeretako bakoitzak balio semantiko horren alderdi bat fokalizatzen du⁷⁰. Aldiz, balio desberdina adierazten duten egiturek elkarren artean ez dituzte alternantziak osatzen, baizik eta delako aditzari dagozkion egitura sintaktiko desberdinak dira, besterik gabe.

Adibidez, *esan* aditzak bi balio semantiko ditu: i) adierazpenezko jarduera eta ii) ezaugarri-esleitzea. Lehenengo balio semantikoan bi egitura sintaktiko markatzen dira ((36) eta (37) adibideak) eta horiek, elkarren artean, alternantzia osatzen dute⁷¹:

1. Adierazpenezko jarduera

1.1 DU-1: *esperimentatzailea* [+giz]_ERG; *gaia* [-konkr]_ABS

(36) *Ez iezaidazu hori esan*

1.2 DU-2: *esperimentatzailea* [+giz]_ERG; *gaia* [-konkr]_ELA_KONP

⁷⁰“[...] concepto de alternancia de diátesis, que definimos como la realización sintagmática de oposiciones semánticas de carácter general” (Vázquez *et al.*, 2000: 73).

⁷¹Alternantziei buruzko xehetasunetarako jo bedi Aldezabalen (2004) tesira.

(37) *Beste bizitzaz hitz egiterakoan, han ere bere amorantea izango dela esango dio*

Beste balio semantikoak egitura sintaktiko bakarra du (38):

2. Ezaugarri-esleitzea

2.1 DIO-3: *abiapuntua* [+giz] _ERG; *helburua* _DAT; *ezaugarria* _ABS

(38) *Honi ere amapontekoa esaten zion gure amak*

Egitura horiek orokorki izendatzeko (direla alternantzia, direla egitura sintaktikoak) **balio sintaktiko-semantiko** (*bss*) kontzeptua darabil Aldezabalek (2004). Aditz baten azpikategorizazio-egiturak, berez, egitura sintaktiko-semantikoak dira, eta egitura horiek aditz baten balio sintaktiko-semantikoak definitzen dituzte, beraien artean alternantzia osatu ala ez. Alternantzia baten aldaera diren egitura sintaktiko-semantikoek osagai semantiko, hau da, rol semantiko berak izango dituzte. Elkarren artean alternantzia osatzen ez dutenek, berriz, ez (Aldezabal, 2004: 169). Hala, *esanek* 3 *bss* ditu; horietako bi adiera bati dagozkio eta bestea beste bati.

Balio kognitibo orokor bakoitza definitzeko beharrezkoak diren rolak erabiltzen ditu Aldezabalek, eta, ondoren, rol horiek modu koherente batean konbinatuta, predikatu-mota orokorrak planteatzen ditu.

EADBn 21 rol semantiko definitu dira:

- **Gaia:** jarduerazko predikatuen objektua.
- **Helburuko kokapena:** kokapen-aldaketa adierazten duten predikatuetan, bukaerako lekua edo kokagunea.
- **Esperimentatzailea:** jarduerazko predikatuen subjektua; ez du nahitaez gizakia izan behar ez eta ekintza nahita egina izan behar.
- **Jarduera:** jarduerazko predikatuetan jarduera dena.
- **Gai ukitua:** egoera- edo kokapen-aldaketa adierazten duten predikatuetan aldaketa jasaten duena.
- **Helburuko egoera:** egoera-aldaketa adierazten duten predikatuetan, bukaeran norbait edo zerbait dagoen modua.

78 Etiketatzeko aurretikoak: baliabideak, aurreprozesuak eta tresna

- **Kausa:** egoera- eta kokapen-aldaketako predikatuen eragilea, jarduerazko predikatuen subjektuaren (esperimentatzailea) bereizgarri.
- **Neurria:** entitate baten tamaina adierazten duena.
- **Gai sortua:** sortzeko predikatuen emaitza den gaia.
- **Abiapuntuko kokapena:** kokapen-aldaketa adierazten duten predikatuetan, hasierako lekua edo kokagunea.
- **Iturria:** sortzeko predikatuetako sortzailea.
- **Modua:** jarduera adierazten duten predikatuetan, jarduteko era.
- **Egoera:** entitate bat egoteko modua edo era.
- **Bidea:** kokapen-aldaketa adierazten duten predikatuetan, egiten den ibilbidea edo tartea.
- **Edukitzailea:** edukitzea adierazten duten predikatuetan, edukia daukana.
- **Kokapena:** entitate bat dagoen lekua.
- **Abiapuntua:** aldaketa edo elkarraldatzea adierazten duten predikatuen hasierako egoera edo puntua.
- **Edukia:** edukitzea adierazten duten predikatuetako gaia.
- **Denbora:** jarduerak irauten duen tartea adierazten duena.
- **Helburua:** aldaketa edo elkarraldatzea adierazten duten predikatuen bukaerako egoera edo puntua.
- **Ezaugarria:** deskribapena adierazten duten predikatuetan, gaia deskribatzen duena.

Eta 21 rol semantiko horiek konbinatuta, 13 predikatu-mota orokor proposatzen dira:

- Entitate baten egoera-aldaketa: kausa + gai ukitua

- Entitate baten kokapen-aldaketa: kausa + gai ukitua + abiapuntuko kokapena + helburuko kokapena
- Entitate baten aldaketa: kausa + gai ukitua + abiapuntua/bidea + helburua
- Entitate baten sortzea: gai sortua + iturria
- Entitate baten jarduera: esperimentatzailea + gaia
- Entitate baten elkarraldatzea: abiapuntua + gaia + helburua
- Entitate bat edukitzea: edukitzailea + edukia
- Entitate bati ezaugarria esleitzea: abiapuntua + ezaugarria + helburua
- Entitate baten izatea, existitzea: gaia
- Entitate baten kokapena: gaia + kokapena
- Entitate baten egoera: gaia + egoera
- Entitate baten deskribapena: gaia + ezaugarria
- Ustezko adierazpena: gaia

Ikusten dugunez, predikatu-mota bakoitzak, delako predikatu-motakoa izateko, gutxieneko osagai semantiko bat/batzuk izan behar du/ditu. Adibideetan osagai semantiko horiek azpimarratuta eman ditugu:

- Egoera-aldaketetan *gai ukitua* izatea da nahitaezkoa (*hautsi, hazi, hil, isildu, kezkatu, haserretu, egokitu, aldatu, pasatu, igo, jaitsi...*):

(39) *Triste al hago [HURA] **hil** zelako?- galdetzen dio Juttak.*

- Kokapen-aldaketetan, *gai ukitua* eta osagai lokatiboren bat (*kokatu, zintzilikatu, sartu, iritsi, ekarri, eraman, etorri, joan, jarri...*):

(40) *Non **jarri** da argazkilaria, gozotegiaren barruan ala kanpoan?*

Non izango litzateke osagai lokatiboa, eta *argazkilaria*, aldiz, *gai ukitua*

80 Etiketatzeko aurretikoak: baliabideak, aurreprozesuak eta tresna

- Sortzezkotetan, *gai sortua* eta *iturria* (*egin, etorri, ekarri, atera, jarri...*):

(41) *Horregatik hilekoa duen emakumeari maionesak ez zaizkio **ateratzen**.*

Kasu honetan, datiboan gauzatzen den *hilekoa duen emakumeari* osagaia *iturria* litzateke, eta absolutiboan agertzen den *maionesak*, berriz, *gai sortua*.

- Edukitzeetan, *edukitzailea* eta *edukia* (*izan, ekarri, etorri, jaso, eraman, hartu...*):

(42) a. *Problema gutxiago **izan** zuen zentsura frankistak.*
b. *Bigarren liburu honetan **badatoz**, gainera, aurrekoaren zuzenketak.*

(42a)-n *edukitzailea* ergatiboan gauzatzen da (*zentsura frankistak*) eta (42b)-n, berriz, inesiboan (*bigarren liburu honetan*). *Edukia* osagaia bietan gauzatzen da absolutiboan (*problema gutxiago* eta *aurrekoaren zuzenketak*).

- Jardueretan, azkenik, *esperimentatzailea* eta, normalean, beste osagaien bat (*erabili, abestu, afaldu, ehizatu, grabatu, jan, ikusi, asmatu, oroitu, ahaztu...*):

(43) *Eskolan **bazkaltzen** du [HARK] eta gehien gustatzen zaion janaria saltxitxak dira.*

Hala izanik, entitate baten aldaketa, entitate baten egoera- edo kokapen-aldaketa ere bada (*gai ukitu* bat dagoelako). Gertatzen dena da aldaketatzat jo den predikatu-motan, aditzak ez duela berez egoera edo kokapenaren informazioa kodetuta; informazio hori sintagmek hartzen duten izenaren baitan dago. Esaterako:

(44) • *Karga negatibo, edo elektroiak, gorputz batetik bestera **aldatzen** direnean korrante elektrikoa sortzen dela esaten dugu.*

gorputza = leku fisikoa >kokapen-aldaketa

- *Izan ere, autonomi edo probintzia-mailara **aldatu** nahi baditugu, zatikaturiko inkestak ez dira lehen bezain adierazgarriak.*

probintzia maila = ezaugarria >egoera-aldaketa

Aditzek batzuetan predikatu orokor bakarra adierazten dute, eta horri loturiko alternantziak izan ditzakete; beste batzuetan, alternantziarik ere ez dago. Ondorioz, aditz batek predikatu orokor edo adiera bakarra duenean, mota semantiko horretakotzat jotzen da (adibidez, *jan*: jarduerazko aditza). Ordea, aditz batzuek predikatu orokor bat baino gehiago adieraz dezakete, eta horiei loturiko alternantziak ere izan ditzakete. Horrelakoe-tan sortzen dira predikatu orokor eta alternantzia guztiak elkarren artean bereizteko zailtasunak (eta beharrak, aldi berean); eta aditz batzuk mota semantiko batekotasun jotzen badira ere (adibidez, *eman*: elkarraldatzeko aditza; *ekarri*: kokapen-aldaketazkoa), bestelako predikatu-motetan ere parte hartzen dute (esaterako, *eman*: jarduerazko predikatu-motan (45); *ekarri*: sortzezko (46) eta edutezko (47) predikatu-motetan).

(45) *Urte osoan horrela egotea gustatuko litzaidake; hau lortzeko denboraldi osoa eman dugu borrokan.*

(46) *Estropadako azken luze biek ez zuten aldaketarik ekarri sailkapenean.*

(47) *Baturantz idatzitako editorial edo berri laburrak dakartzate, baxe-nabartar kutsu gehiagorekin Ekaitzak, lapurtera gehiagorekin Enbatak eta H.H.ak.*

Hala, hauexek dira aztertutako 100 aditzetan agertu diren rol-konbinazioak (predikatu-mota bera ez delako beti egitura sintaktiko berarekin gauzatzen). Maiztasun handienekoak, hau da, aditz gehiagok erakusten dituztenak, horiz nabarmenduta ageri dira:

1. Egoera-aldaketa:

- **kausa+gai ukitua**

(48) **hil**: *Bi tiroz hil zuten [haiek] [hura]*

82 Etiketatzeko aurretikoak: baliabideak, aurreprozesuak eta tresna

- kausa + gai ukitua + gai ukitua

(49) **banandu:** *Penintsula Iberiarra inguru erdi-isolatua da espezie asko eta askoren banaketari dagokionez, ehundaka edo milaka metrotako horma batek - Pirinioak hain zuzen - “banantzen” baitu [Penintsula Iberiarra] Europaren gaintzeko lurraldetik.*

- gai ukitua

(50) **irakin:** *Zein tenperaturatan irakiten du? [hark]*

- kausa + gai ukitua + helburuko egoera

(51) **egokitu:** *Lehen gazteek beraiek egokitzen zituzten euren arauak unean uneko egoerara.*

2. Kokapen-aldaketa:

- kausa + gai ukitua + abiapuntuko kokapena / bidea + helburuko kokapena

(52) **ekarri:** *Laurak hozkailutik jela koakorrak ekarri zituen [hona]*

- gai ukitua + abiapuntuko kokapena / bidea + helburuko kokapena

(53) **joan:** *Bizkaia, Araba, Nafarroa eta Iparraldetik joaten da jendea Gipuzkoako sagardotegietara.*

- kausa + gai ukitua + helburuko kokapena

(54) **jarri:** *Pailazok guztien erdian jartzen dute Eli.*

- gai ukitua + helburuko kokapena

(55) **bisitatu:** *Oposizioko buru Ariel Sharonek meskitak⁷² bisitatu zituen.*

- kausa + gai ukitua + abiapuntuko kokapena

⁷²“ABSk, izaki bizidun horrek “bisitatzen” duen gizakia edota lekua adierazten du, nolabait *helburuko kokapen* modura ikus litekeena. Halakoxeztat joz gero, [...] ekintzaren egile den hori lekuz aldatzen den entitate gisa ikus dezakegu; gai desplazatu gisa, alegia” (Aldezabal, 2004: C eranskina, 5).

(56) **hartu:** Hitz berriak Batutik hartzen dituzte [haiek]

3. Aldaketa:

- Kausa+gai ukitua+abiapuntua/bidea+helburua

(57) **aldatu:** Oñatiko ur-hoditeria Urretxuko ur-biltegitik saihesbidera aldatzeko Proiektua.

4. Entitate baten sortzea:

- iturria+gai sortua

(58) **isuri:** Bertatik argi izpi indartsu batzuk isurtzen dira.

- iturri+gai sortua+kokapena

(59) **eragin:** Errefuxiatuak zihoazen konboi bati eraso eta saraskia eragin du NATOk [han]

5. Entitate baten jarduera:

- esperimentatzailea+gaia

(60) **abestu:** Gaztetxe bakoitzeko partaide batek bertso bat abestu eta triste bat kontatu zuen.

- esperimentatzailea+esperimentatzailea

(61) **besarkatu:** Karelean jarriak zituen nebak ukondoak, emakumeak besarkatu egin zuen [neba] eta musu eman zioten elkarri.

- esperimentatzailea+jarduera

(62) **aritu:** Futbolean aritu ziren gehienak.

- esperimentatzailea+modua

(63) **jokatu:** Osasunako zuzendaritza taldeak ez duela zuzen jokatu salatu zuten Indar Gorri taldekoek.

- esperimentatzailea+jarduera+denbora

(64) **eman:** Kondek 20 minutu inguru eman zituen deklaratzeko.

84 Etiketatzeko aurretikoak: baliabideak, aurreprozesuak eta tresna

- esperimentatzailea+jarduera+helburua

(65) **deitu**: Floren Aoziz⁷³ ezin zuen deklaratzera deitu Garzonenek.

- esperimentatzailea+gaia+gaia

(66) **aldatu**: Kentuckyko zaldi haragi potoak, Kentuckyko harrizko upeletan zaharturiko bourbonaz aldatu zituen [hark]

6. Entitate baten elkarraldatzea:

- abiapuntua+gaia+helburua

(67) **eman**: Foru Erkidegoko Parlamentuak terrorismoaren biktimei Nafarroako domina ematea salatu dute.

7. Entitate bati ezaugarria esleitzea:

- abiapuntua+ezaugarria+helburua

(68) **esan**: Honi ere amapontekoa esan zion gure amak.

8. Entitate bat edukitzea:

- edukitzailea+edukia

(69) **izan**: Eibarrek ere bere aukerak izan zituen.

- edukitzailea+edukia+kokapena

(70) **ekarri**: Idazlearen lan guzien izenen zerrenda, bukaeran dakar [liburuak?]

9. Entitate baten izatea, existitzea:

- gaia

(71) **existitu**: Horrek frogatzen du egungo marko politikoan euskaldunon eskubideak ez direla existitzen.

⁷³“[...] *helburua* ABSz (eta ez ohikoa den DATEz) gauzatzeak zerikusia du *deitu* aditzak bere pieza lexikalean *dei* hitza txertatuta izatearekin. Horrek, badirudi, ABSri aukera ematen diola beste balio batez gauzatzeko” (Aldezabal, 2004: C eranskina, 251).

10. Entitate baten kokapena:

- gaia+kokapena

(72) **egon:** Ohean zegoen amona.

11. Entitate baten egoera:

- gaia+egoera

(73) **egon:** Abertzaleak eroso egotea izan da Ibarretzeren gobernuaren helburua.

12. Entitate baten deskribapena:

- gaia+ezaugarria

(74) **izan:** Orrien edukia, ordea, politegia zen benetakoa izateko.

13. Ustezko adierazpena:

- gaia

(75) **eman:** Delikadezak, sentiberatasunak, fineziak... izaturik hemendik ihes egin dutela ematen du.

EADBN landutako 100 aditzen azterketa sakona ikusi nahi izanez gero, jo bedi [Aldezabalen \(2004\)](#) tesi-laneko C eranskinera.

4.1.2 EPEC-DEP (BDT) corpora

Esana dugu, EPEC corpora predikatu-mailan etiketatzeko, dependentziekin etiketatutako corpusean oinarritu garelara (EPEC-DEP⁷⁴ ([Aldezabal et al., 2009a](#))), aditzaren mende agertzen den dependentzia orori *arg_info* etiketa esleitzeko.

1.1 atalean esan dugun bezala, EPEC corpusaren bi heren *Euskaldunon Egunkariatik* hartuak dira, eta, beste herena, *XX. mendeko euskararen corpus estatistikotik*. Corpusaren eskuzko etiketatze-lana hasi zenean, corpora fitxategietan banatu zen. Hala, *Euskaldunon Egunkariako* berri bakoitza

⁷⁴<http://ixa.si.ehu.es/Ixa/Produktuak/1306407157>

fitxategi batean sartu zen. *XX. mendeko euskararen corpus estatistikoari* dagokionez, ez ziren obra osoak hautatu, zatiak baizik; beraz, obra bereko esaldi-segida bat (1.000 hitz inguru) sartu zen fitxategi bakoitzean.

EPEC-DEP zuhaitz-bankuan, 200.000 hitz etiketatu dira eskuz Dependentsia Gramatikaren Teoriari (Tesnière, 1959) jarraituz. Teoria horretan, esaldiko hitzak binaka lotuz, esaldiaren zuhaitz sintaktikoa (dependentsia-zuhaitza ere deitua) lortzen da (ikusi 1.1 ataleko 1.1 irudia). Zuhaitz horietan, batetik, adabegietan dauden hitzen arteko gobernatzaile/mendeko erlazioak irudikatzen dira, eta bestetik, bi hitzen arteko loturan mendekoak betetzen duen funtzio sintaktikoa adierazten da dependentsia-etiketen bidez (Aranzabe, 2008).

Dependentsiekin etiketatutako corpus horretan, 1.211 aditz eta 30.740 agerpen daude guztira. Tesi-lan honetan, 271 aditz landu ditugu, eta horien artean corpusean 30 agerpenetik gora duten 151 aditzak daude. Horrekin corpusaren % 85 etiketatuta geratzen da.

4.1.3 Levinen (1993) aditzen euskarazko itzulpena

Aldezabalek (1998), Levinen sailkapenean oinarrituta, euskarazko eta ingelesezko aditzen arteko konparaketa ere egin zuen. Horretarako, Levinen (1993) aditz guztiak euskarara itzuli zituen, lehenik esanahiari (klase semantikoa-ri) eta gero jokaera sintaktikoari erreparatuta. Hots, ez zen itzulpen hutsa egin, baizik eta klase horretarako egitura sintaktiko antzekoena zuen/zuten aditzen ordainak hautatu ziren.

Levinen aditzen itzulpena, beraz, bi irizpide hauek (esanahia eta egitura sintaktikoa) jarraituta egin zuen, eta Aldezabalek (2010) itzulpen-lan hori egitean agertutako kasuistika azaltzen du:

1. Esanahia eta sintaxia bat datoz bi hizkuntzetan.

Hau da itzultzeko kasurik errazena. Adibidez, *to break* aditza 45.1 klasekoa da (*break verbs*) eta euskaraz *hautsi*, *apurtu* eta *puskatu* sinonimoekin itzul daiteke, hirurek dutelako esanahi berdina (klaseari begira: egoera-aldaketa) eta hirurek onartzen dutelako alternantzia kausatibo/inkoatiboa (76).

- (76) a. → *John broke the window.* (kausatiboa)
→ *The window broke.* (inkoatiboa)

- b. → *Jonek leihoa hautsi/apurtu/puskatu zuen.* (kausatiboa)
 → *Leihoa hautsi/apurtu/puskatu zen.* (inkoatiboa)

2. Zenbait ezaugarri sintaktiko partekatzen dituzte bi hizkuntzek.

Ingelesez *tell* eta *say* aditzen portaera sintaktikoa ez da berdina. *Tell* aditzak alternantzia datiboa onartzen du (*I tell sb. sth. / I tell sth. to sb.*) eta *say* aditzak, aldiz, ez (**I say sb. sth.*), baina euskaraz biak itzultzen ditugu *esan* aditzarekin. Euskaraz ez dago aldaera sintaktiko hori erakusten duen aditz desberdinik, eta *esan* aditzaren argumentuak ingelesezko bi aditzen berberak dira.

3. Batzuetan ez da sintaktikoki oso argi geratzen zergatik bereizten diren klase batzuk, eta, beraz, klase horietan sartzen diren aditzentzat ordainak topatzea ere ez da erraza.

Levinek (1993) 13.1 (*give verbs*) eta 13.3 (*verbs of future having*) klaseak bereizten ditu, eta bigarren klase horretan esanahi zehatzagoa duten aditzak (*assign, cede, concede, offer...*) sartzen ditu, baina ez orokorra den *give* aditza. Euskaraz ere antzeko zerbait egin daiteke, lehenengo klasean sartuz *eman* orokorra, eta bigarren klaserako utziz zehatzagoak diren beste aditz batzuk. *Esleitu*, esaterako, *assign* aditzaren ordain gisa jar daiteke; baina, berez, ez da arrazoirik ikusten *eman* ere klase honetan ez sartzeko, egitura sintaktikoak eta alternantziak bederen berdinak baitira (eta hori, berez, ez dator bat Levinen teoriarekin).

4. Ingelesezko aditzaren esanahia itzultzeko euskaraz bi hitz behar dira: aditza eta aparteko osagaia.

Askotan, ingelesezko aditzak barnean darama ekintza egiteko modua edo erabilitako tresna; adibidez, *funnel verbs* (9.3), *wipe verbs* (10.4), *spray/load verbs* (9.7), *drive verbs* (11.5), *poison verbs* (42.2) eta *verbs of instrument of communication* (37.4) klaseetan sailkatutako aditz askotan. Gehienetan aditz horiek euskarara itzultzeko bi hitz behar izaten ditugu: *ladle = burduntzaliáz zerbitzatu*. Hemen gertatzen dena da maiz ekintza egiteko modua lexikalizatuta dagoela ingelesez, baina euskaraz ez (ikusi 6.3.1 atala). Horrelako kasuetan, ez dugu aparteko aditz

88 Etiketatzeko aurretikoak: baliabideak, aurreprozesuak eta tresna

gisa aztertzen, hau da, aditza *zerbitzatu* da eta modua (*burduntzaliaz*) adjuntua izango litzateke.

5. Klase desberdinetako aditzak itzultzean euskaraz aditz bera erabiltzen da, baina objektua zehaztu egin behar da.

Euskaraz 13.4.2 (*equip verbs*) klaseko ingelesezko aditzak, 13.1 (*give verbs*) eta 13.2 (*contribute verbs*) klaseetako aditz berekin itzultzen dira. Adibidez, *charge* aditza *ardura eman* ordainarekin itzuli behar dugu euskaraz. Ingelesez *to charge somebody with a task* esan behar da (eta ez *give*). Euskaraz, aldiz, ez dago horrelako desberdintasunik, baina *charge* aditzaren itzulpena ematean, objektua (*ardura*) zehaztu beharra dago (*ardura eman*); eta *emanen* ohiko argumentuak hartzen ditu (alegia, NOR-NORI-NORK eta ez *withen* pareko lizatekeen beste postposizio-atzizkirik).

Kasuistika hori kontuan hartuta, argi geratzen da ingelesezko eta euskarazko aditzen argumentu-egiturak ez direla beti bat etorriko eta, beraz, ez eta egitura horietan agertuko diren alternantzia sintaktikoak ere.

Hala ere, lan honi esker Levenen klaseetan oinarritutako euskara-ingelesa aditzen zerrenda lortu dugu, eta oso baliagarria izan zaigu gure aditzak *Unified VerbIndex*arekin lotzeko eta horko informazio guztia jasotzeko⁷⁵. 4.1 taulan zerrenda horren lagin txiki bat ikus daiteke:

tell	37.1	esan, erran
tell	37.2	esan, erran
tense	45.4	teinkatu, tinkatu, gogortu
term	29.3	deitu, izendatu, -tzat hartu/eduki
terminate	55.1	bukatu, amaitu
terrify	31.1	izutu, izuarazi
terrorize	31.1	izua sartu, ikaratu
tether	22.4	sokaz lotu
thank	33	eskertu, eskerrak eman

4.1 taula: Levin (1993) eta euskal aditzen arteko loturaren hainbat adibide.

⁷⁵ Aditz-zerrenda eta ordainak ikusi nahi izanez gero:
<http://ixa2.si.ehu.es/aestarrona/Levin-ingelesa-euskara.txt>

4.2 Aurreprozesu automatikoak

Atal honetan, etiketatze-lana errazte aldera, etiketatzen hasi aurretik burututako prozedura automatikoak azalduko ditugu.

4.2.1 [Levinen \(1993\)](#) aditzen itzulpena PBrekin gurutzatuta eta corpusarekin lotuta

Euskarazko aditzak Levinen klaseetara loturik izanik eta PropBank-en eta VerbNet-en ere aditz bakoitzean Levinen klaseak zehazten direnez, PBn eta VN-n agertzen den ingelesezko aditzaren argumentu eta rol semantikoei buruzko informazio guztia eskura dugu.

Lotura horiek baliatuta, euskarazko aditza, ordain gisa duen ingelesezko aditzari dagokion PB-VN informazioaz hornitu dugu (itzulpenetarako kontuan hartu diren Levinen klaseetan oinarrituta, hain zuzen), informazio hori EPECen aditz horrek dituen agerpen orori esleitu diogu eta *AbarHitz* ([Díaz de Ilarraza et al., 2004](#)) tresnan ikusgarri jarri dugu, etiketatzailen lana errazteko asmoz (ikusi [4.3](#) atalean tresnaren aurkezpen zehatza).

Prozesu hori automatikoki egin da, baina ez da zuzenekoa izan kasu guztietan. Eragiketa hori zuzenekoa litzateke PB-VN-n Levinen klaseak berdindin mantendu izan balira, baina, esan dugun bezala, klase horiek berraztertzen eta egokitzen joan dira, eta, ondorioz, itzulpenetan erabili den klasea ([Levin, 1993](#)) ez dator beti bat PB-VN-n agertzen den klasearekin. Horiek horrela, lau egoera agertu zaizkigu:

EQUAL. ⁷⁶ Delako aditzean [Levinen \(1993\)](#) klasea eta gaur egun PB-VN-n ageri dena berdina da; % 74,92 dira horrelakoak. Adibidez: *glue* 22.4 (*erantsi, kolatu*)⁷⁷ (EQUAL⁷⁸).

SUBCLASS. Delako aditzak [Levinen \(1993\)](#) klase orokorrago bat zuen eta orain PB-VN-n klase berorren azpiklase bat du. Adibidez, [Levinen \(1993\)](#) *go* 51.1 (*joan*) aditzak 51.1.2 du PB-VN-n. Horiek % 9,46 dira.

⁷⁶EQUAL: berdina; SUBCLASS: azpiklasea; CHANGED: aldatuta eta MISSING: ez da agertzen.

⁷⁷Parentesi artean ematen ditugun euskarazko aditzak [Aldezabalek \(1998\)](#) emandako itzulpenak dira.

⁷⁸Kasuistika honen (hots, EQUAL, SUBCLASS, CHANGE eta MISSING hauen) kopuruak aldakorrak dira bertsioen arabera. Hemen ematen ditugun adibideak 2012ko PBko bertsioarekin konparatuta ateratakoak dira.

90 Etiketatzeko aurretikoak: baliabideak, aurreprozesuak eta tresna

CHANGED. Delako aditzak duen Levinen klasea ez dator bat PB-VN-n duenarekin. Hala ere, aditz horrek Levinen lanean klase bakarra badu, PB-VN-n duen klasekotzat jo da zuzenean, eta horren informazioa hartu da. Adibidez, *agonize* (*nahigabetu*, *atsekabetu*, *intziri egin*, *lantua jo*) aditzak Levinen (1993) 31.1 zuen, eta PB-VN-n 31.3. Hala, nahiz eta klasea beste bat izan, bakarra denez, PB-VNko informazioa zuzenean esleitu diogu euskarazko aditzari. Horrelakoak % 2,7 dira.

MISSING. Ingeleseko aditza ez dago PB-VN-n edo, baldin badago, ez du inolako Levin klaserik esleituta. % 12,8 dira horrelakoak. Azken bertsioan aditz berri asko landu dituzte, eta, beraz, lehen MISSING ziren aditz asko orain kategoria horretatik atera dira. Adibidez: PB-VN bertsio zaharrarekin konparatuta, *glom*⁷⁹ 22.3 MISSING zen, baina bertsio berrian dagoeneko badago, eta orain EQUAL etiketa du. Bestalde, *adjudicate* (*ebatzi*, *erabaki*, *epaia eman*) aditza ez zegoen PB-VNko bertsio zaharrean, eta, beraz, MISSING zen. Gaur egungo bertsioan gehitu egin dute, baina ez diote Levin klaserik esleitu. Beraz, ezin dugu loturarik egin, eta MISSING etiketa izaten jarraituko du.

4.2 taulan Levinen klaseen eta PB-VN-n agertzen diren klaseen arteko konparaketaren zenbait adibide ikus daitezke.

⁷⁹Aditz hau ez dago itzulia Aldezabalen (1998) lanean.

Aditza	Euskarazko itzulpena	Levin klasea	PB-VN klasea	Emaitzak
<i>adjudicate</i>	ebatzi, erabaki, epaia eman	29.4	-	MISSING
<i>tattoo</i>	tatuajea egin	29.1	25.1	CHANGED
<i>tell</i>	esan, erran	37.1	37.1-1	SUBCLASS
<i>tell</i>	esan, erran	37.2	37.2-1	SUBCLASS
<i>tense</i>	teinkatu, tinkatu, gogortu	45.4	45.4	EQUAL
<i>term</i>	deitu, izendatu, -tzat hartu / eduki	29.3	29.3	EQUAL
<i>terminate</i>	bukatu, amaitu	55.4	55.4	EQUAL
<i>terrify</i>	izutu, izuarazi	31.1	31.1	EQUAL
<i>terrorize</i>	izua sartu, ikaratu	31.1	31.1	EQUAL
<i>tether</i>	sokaz lotu	22.4	22.4	EQUAL
<i>thank</i>	eskertu, eskerrak eman	33	33	EQUAL

4.2 taula: [Levinen \(1993\)](#) klaseen eta 2012ko PB-VN klaseen arteko konparaketa.

Kasuistika honen aurrean, guk printzipioz EQUAL eta SUBCLASS gisa sailkatutako guztiak, eta CHANGEDeko kasu zehatz batzuk (klase desberdina, baina bakarra) hartu ditugu kontuan, EPECen informazioa automatikoki jartzeko.

Lotura egin ostean, agertu diren PB-VNko rol guztien zerrenda⁸⁰ atera da. Horrekin, aurrerantzean etiketatzeko baliatuko diren rolen zerrenda nahiko osoa lortu da⁸¹:

1. ADV (Adverbials)
2. Agent

⁸⁰Letra larriz agertzen diren rolak PBk adjuntuei esleitutakoak dira, eta letra xehez daudenak VNk argumentuei esleitutakoak.

⁸¹Zerrenda horri, [6.1.3](#) atalean azaldu bezala, beste rol bat gehigu diogu, *path* delakoa, hain zuzen ere.

92 Etiketatzeko aurretikoak: baliabideak, aurreprozesuak eta tresna

3. Asset
4. Attribute
5. BareArgM
6. Beneficiary
7. Cause
8. Destination
9. DIR (Directionals)
10. DIS (Discourse markers)
11. Experiencer
12. EXT (Extent markers)
13. Extent
14. Instrument
15. LOC (Locatives)
16. Location
17. Material
18. MNR (Manner markers)
19. MOD (Modals)
20. NEG (Negation)
21. Patient
22. PRD (Markers of secondary predication)
23. Predicate

24. PRP (Purpose clauses)⁸²
25. REC (Reciprocals)
26. Recipient
27. Source
28. Stimulus
29. Theme
30. TMP (Temporal markers)
31. Topic

4.2.2 EADBko 100 aditzen informazioaren esleipen automatikoa, postposizio-atzizkietan oinarrituta

EADBko 100 aditzak hartu dira, eta EPECen azaltzen direnei⁸³ EADBn definitutako rolak esleitu zaizkie automatikoki. Horretarako, osagai bakoitzak sintaktikoki gauzatzeko baliatzen duen postposizio-atzizkia hartu da oinarritzat. Esaterako, *esanek* bere adiera batean ABSan *gaia* du eta ERGan *experimentatzailea*; beraz, rol horiek automatikoki esleituko lirateke (77).

(77) *Zer_{ABS}gaia esan nian nik_{ERG}esper. hiri ...?*

Hori eginda, aditz jakin batean postposizio-atzizkiak edo bere mendeko osagaiak anbiguoak ez direnean, osagai horiek automatikoki etiketatuta eta desanbiguatuta geratzen dira. Anbiguotasuna dagoenean, aldiz, etiketatzaileak eskuz ebatzi behar izaten du.

Jarraian, anbigua ez den *adierazi* aditzaren adibide bat azalduko dugu (78). EADBk honako informazioa jasotzen du *adierazi* aditzarentzat:

⁸²PBko gidalerroetan helburuzko perpausak etiketatzeko PNC (*purpose, not cause*) etiketa erabiliko dutela esaten dute, baina etiketatuta topatu duguna PRP izan da. Esaterako, *take.01* aditz-adieran *it has taken measures to prevent cheating* adibidea ematen dute, eta *to prevent cheating* helburuzko perpausari *Argm-prp* etiketa jartzen diote. Guk ere PRP erabili dugu helburuzko perpausak etiketatzeko.

⁸³EADBko 100 aditz horietatik 99 agertzen dira EPECen, eta horiek dira guk landu ditugunak. Agertzen ez den aditz bakarra *loratu* da.

94 Etiketatzeko aurretikoak: baliabideak, aurreprozesuak eta tresna

1. `esperimentatzailea_ERG`; `gaia [-biz; -konkr]_ABS`
2. `esperimentatzailea_ERG`; `gaia [-biz; -konkr]_KONP`

(78) *Israelgo helikopteroek gune palestinarrek bonbardatu zituztela adierazi zuten lekukoek.*

(78)-ko *-ela* eta ergatibo postposizio-atzizkiak⁸⁴ kontuan hartuta, aurreprozesurako erabili dugun tresnak (79) adibidean ikus daitezkeen *arg_info* etiketak sortzen ditu, *Israelgo helikopteroek gune palestinarrek bonbardatu zituztela* mendeko perpausarentzat eta *lekukoek* subjektuarentzat:

(79) **ccomp_obj** (konpl, adierazi, bonbardatu, zituztela)
arg_info (-, adierazi, zituztela, -, -, gaia, -biz/-konkr)
ncsubj (erg, adierazi, lekukoek, lekukoek, subj)
arg_info (-, adierazi, lekukoek, -, -, esperimentatzailea, -)

Geratzen diren hutsuneak eskuz bete behar dira. Anbiguoak diren kasuetan, berriz, sistema automatikoak bi etiketa (edo gehiago) sortzen ditu; esaterako, *gertatu* aditzari bi adiera definitzen zaizkio EADBn:

1. `gaia_ABS`; `kokapena_INE` → entitate bat nonbait izatea (gertatzea)
2. `gaia_ABS`; `egoera_ABS` → entitate baten egoera

Ikusten denez, lehenengo adierako *gaia*, eta bigarren adierako *gaia* eta *egoera* argumentuak, denak gauzatzen dira postposizio-atzizki berarekin (ABS) eta, beraz, sistemak aukera guztiak emango ditu etiketatzaileak eskuz desanbiguatu dezan. Ikus (80) adibidea eta honi dagozkion etiketak.

(80) *Espezieen babespen egokia gerta dadin, habitat bera babestu egin behar da.*

ncsubj (**abs**, gerta, babespen, egokia, subj)
arg_info (-, gerta, babespen, -, -, **gaia**, -)
arg_info (-, gerta, babespen, -, -, **egoera**, -)

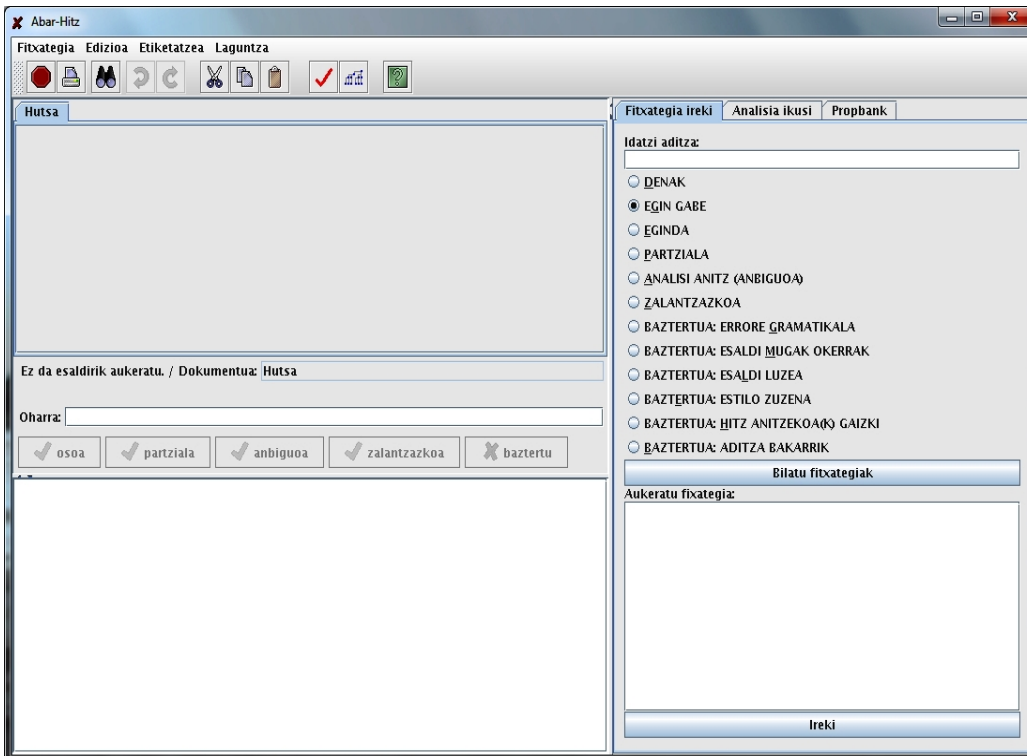
Bukatzeko, hurrengo atalean, etiketatzeko erabili dugun tresna aurkeztuko dugu irudiak baliatuz.

⁸⁴Postposizio-atzizki terminoa darabilgu orokorrean “sintaxi-atzizki”ak izendatzeko, deklinabide-atzizki nahiz erlazio-atzizki izan.

4.3 AbarHitz: etiketatzeko tresna

Predikatu-mailan etiketatzeko erabili dugun tresna etiketatze sintaktikorako erabilitako bera da: **AbarHitz**. Tresna honi buruzko informazio tekniko xehea nahi izanez gero, jo bedi [Díaz de Ilarraza *et al.*en \(2004\)](#) lanera. Hemen, labur, oinarritzko azalpenak emango ditugu interfazeari dagozkion irudi batzuk erakutsiz.

Tresna hau IXA Taldeko zerbitzarietan dago eta, beraz, tresna erabili ahal izateko, lehenik eta behin zerbitzari horietan sartu behar da (siuc04.si.ehu.es edo siuc01.si.ehu.es). Tresna martxan jartzeko, ezinbestekoa da X-Win programa aktibatuta izatea. Ondoren, zerbitzarira konektatu behar da, agindulerroan *abarhitz.xml* idatzi eta interfazea azaltzen da (4.1 irudia):

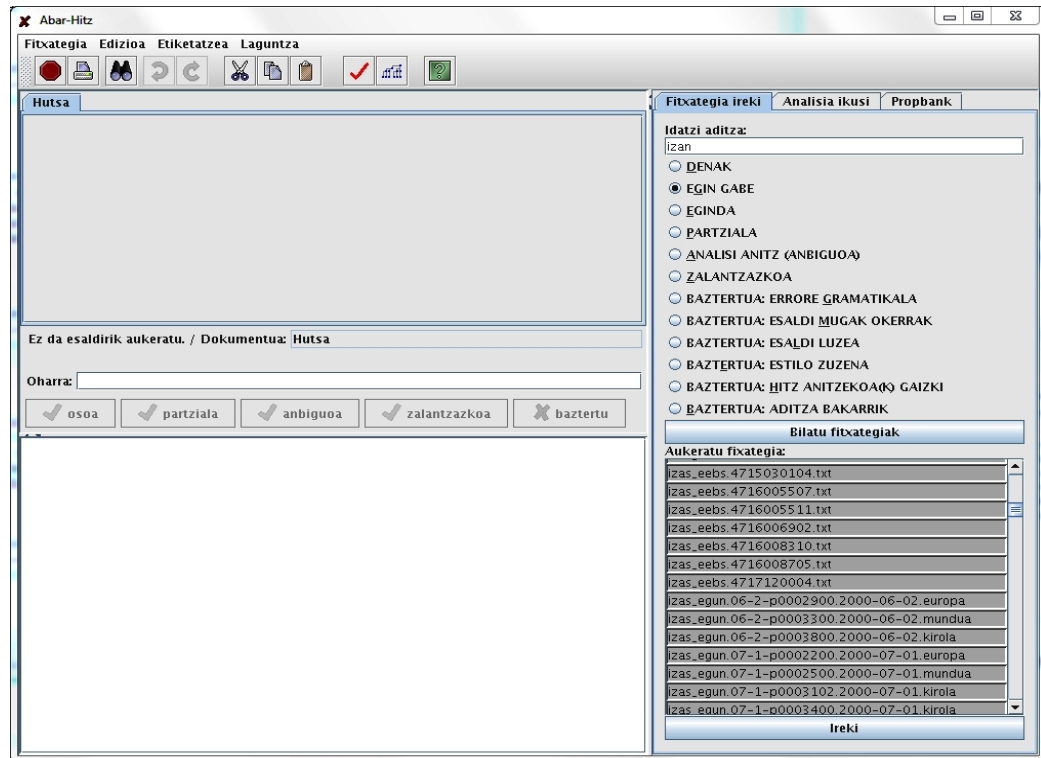


4.1 irudia: AbarHitz-en lanean hasteko interfazea.

Lanean hasteko, leihoaren eskuineko aldean kokatu, eta **Idatzi aditza** botoiaren azpian dagoen laukitxoan idatzi behar dugu landu nahi dugun aditza. Adibidez: *izan* eta **Bilatu fitxategiak** botoian klik eginda, botoiaren

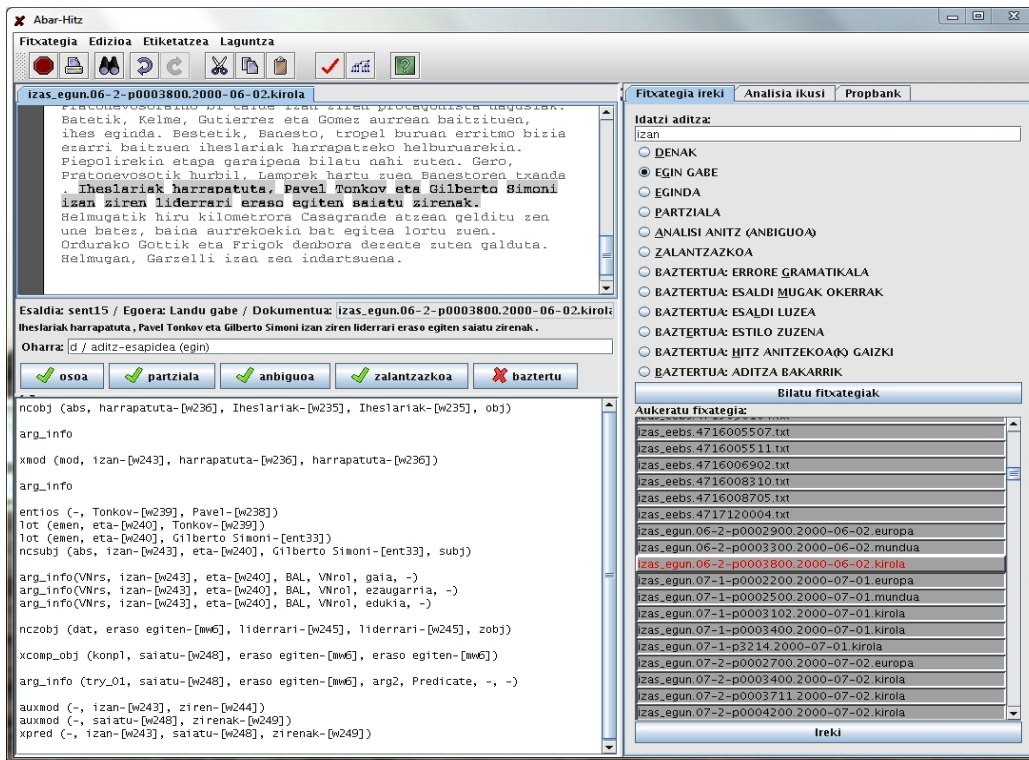
96 Etiketatzeko aurretikoak: baliabideak, aurreprozesuak eta tresna

beheko aldean, *izan* aditza duten adibideei dagozkien fitxategiak azalduko zaizkigu (4.2 irudia). Tresna lanean dagoela ziurtatzeko, saguaren kurtsorea “lanpetuta” agertuko da.



4.2 irudia: *izan* aditza duten fitxategien zerrenda, AbarHitz-en.

Horietako fitxategiren bat lantzen hasteko, saguarekin fitxategia hautatu dugu eta **Ireki** botoiaren gainean egingo dugu klik. Esan bezala, fitxategia ireki ondoren, fitxategi horretako esaldiak ezker aldeko goiko leihoan azalduko zaizkigu. Landu beharreko esaldiak letra-molde lodiaz nabarmenduta azalduko dira eta, saguarekin esaldi bat hautatzean, esaldia azpimarratua agertuko da eta beheko aldeko leihoan bere analisia agertuko da, dependentziak etiketatuta eta predikatuak etiketatzeko *arg_info* etiketekin (4.3 irudia).



4.3 irudia: Etiketatzen hasteko leihoa.

Etiketatzailaren lana errazteko, *AbarHitz* tresnak lantzen ari garen PBko balizko ordainen informazioa eskaintzen du (ikus 4.2.1 atala). Horretarako, lantzen ari garen esaldian aztergai den aditzaren gainean kokatuko dugu sagua eta eskuineko botoia sakatuko dugu. Ondoren, eskuineko goiko aldean dagoen menuan **PropBank** aukeratu eta beheko leihoan ikusiko dugu dagoen informazioa (4.4 irudia).

98 Etiketatzeko aurretikoak: baliabideak, aurreprozesuak eta tresna

The screenshot shows the Abar-Hitz software interface. The main window displays a text document titled "izas_egun.06-2-p0003800.2000-06-02.kirola". The text is in Basque and discusses a political event. Below the text, there are several buttons: "osoa", "parziala", "anbigua", "zalantzakoa", and "baztertu". The right-hand pane shows the semantic analysis for the word "exist". It includes the following information:

- PREDICATE:** exist
- ID:** exist:01 NAME: be VERBNET: 29.6 47.1-1
- ROLE1:** thing existing VNTHETA (39.6): Agent VNTHETA (47.1-1): Theme
- ROLE2:** secondary attribute
- EXAMPLE:** ...a case of political corruption that "trace" simply doesn't exist.
- ASPECT:** FORM: PERSON: TENSE: VOICE: ARG(1): "trace"=that=political corruption ARG(m)(DIS): simply ARG(m)(NEG): n't REL: exist
- EXAMPLE:** John exists as a figment of my imagination.
- ASPECT:** ns FORM: full PERSON: third TENSE: present VOICE: active ARG(1): John REL: exists ARG(2)(as): a figment of my imagination
- EXAMPLE:** No mechanism "trace" exists for finding a middle ground.
- ASPECT:** FORM: PERSON: TENSE: VOICE: ARG(1): [No mechanism] "trace" for finding a middle ground REL: exists

4.4 irudia: Aztergai den aditzari dagozkion PBko ordainen informazio-leihoak.

Etiketatzeko hasteko, *arg_info* etiketen eremuak betetzen hasiko gara, eta tresnak berak eskainiko digu eremu bakoitzean jar daitezkeen balioen menua (4.5 irudia).

The screenshot shows the Abar-Hitz software interface. The main window displays a document with the following text:

izas. egun.06-2-p0003800.2000-06-02.kirola

mes eguna, bestetik, basteo, tropei buruz eritimo biza ezarri baitzuen iheslariak harrapatzeko helburuarekin. Piepolirekin etapa garaipena bilatu nahi zuten. Gero, Pratoevosotik hurbil, Lamprek hartu zuen Banestoren txanda. **Iheslariak harrapatuta, Pavel Tonkov eta Gilberto Simoni izan ziren liderrari eraso egiten saiatu zirenak.** Helmugetik hiru kilometroa Casagrande atzean gelditu zen une batez, baina aurrekoekin bat egitea lotu zuen. Ordurako Gottik eta Frigok denbora desente zuten galduta. Helmugetan, Garzelli izan zen indartsuena.

Esaldia: senti15 / Egoera: Landu gabe / Dokumentua: izas. egun.06-2-p0003800.2000-06-02.kirola
Iheslariak harrapatuta, Pavel Tonkov eta Gilberto Simoni izan ziren liderrari eraso egiten saiatu zirenak.

Oharra: / aditz-esapidea (egin)

oso partziala anbigua zalantzakoa baztertu

ncobj (abs, harrapatuta-[w236], Iheslariak-[w235], Iheslariak-[w235], obj)

arg_info

xmod (mod, izan-[w243], harrapatuta-[w236], harrapatuta-[w236])

arg_info

entios (-, Tonkov-[w239], Pavel-[w238])
lot (emen, eta-[w240], Tonkov-[w239])
lot (emen, eta-[w240], Gilberto Simoni-[ent33])
ncsubj (abs, izan-[w243], eta-[w240], Gilberto Simoni-[ent33], subj)

arg_info(VNrs, izan-[w243], eta-[w240], VNrol, galia, -)
arg_info(VNrs, izan-[w243], eta-[w240], B...
arg_info(VNrs, izan-[w243], eta-[w240], B...

nczobj (dat, eraso egiten-[w246], liderrari-[w245], eraso egiten-[w246])

xcomp_obj (konpl, saiatu-[w248], eraso egiten-[w246], saiatu-[w248])

arg_info (try_01, saiatu-[w248], eraso egiten-[w246], saiatu-[w248])

auxmod (-, izan-[w243], zirenak-[w244])
auxmod (-, saiatu-[w248], zirenak-[w249])
xpred (-, izan-[w243], saiatu-[w248], zirenak-[w249])

The right panel shows the analysis for the highlighted text:

PREDICATE: exist

ID: exist:01 NAME: be VERBNET: 39.6 47.1-1

ROLE1: thing existing VNTHETA (39.6): Agent
VNTHETA (47.1-1): Theme
ROLE2: secondary attribute

EXAMPLE: ...a case of political corruption that "trace" simply doesn't exist.

ASPECT: FORM: PERSON: TENSE: VOICE:
ARG(1): "trace"=that=political corruption
ARG(m)(DIS): simply
ARG(m)(NEG): n't
REL: exist

EXAMPLE: John exists as a figment of my imagination.

ASPECT: ns FORM: full PERSON: third TENSE: present VOICE: active
ARG(1): John
REL: exist
ARG(2)(as): a figment of my imagination

EXAMPLE: No mechanism "trace" exists for finding a middle ground.

ASPECT: FORM: PERSON: TENSE: VOICE:
ARG(1): [No mechanism] ["trace" for finding a middle ground]
REL: exist

4.5 irudia: Balentziaren eremuari dagokion menua.

Balio bakoitza betetzean, “koma” sakatu eta hurrengo eremuaren menua agertuko zaigu (4.6 irudia).

100Etiketatzeko aurretikoak: baliabideak, aurreprozesuak eta tresna

The screenshot shows the Abar-Hitz software interface. The main window displays a document with the following text: "Ihes eginda, Estetik, Bañesto, Erapel burua eritimo bizia ezarri baitzuen iheslariak harrepatzeko helburuarekin. Piepolirekin etapa garrantzitsua bilatu nahi zuten. Gero, Pratohevostotik hurbil, Lamprek hartu zuten Bañestorea txanda. Iheslariak harrepatuta, Pavel Tonkov eta Gilberto Simoni izan ziren liderrari eraso egiten saiatu zirenak. Helmugetatik hiru kilometrora Casagrande atzean gelditu zen une batez, baina aurrekoekin bat egitea lortu zuen. Ondurako Gottik eta Frigok dambora deusente zuten galduta. Helmugeta, Garrelli izan zen indartsuena." The text "Iheslariak harrepatuta, Pavel Tonkov eta Gilberto Simoni izan ziren liderrari eraso egiten saiatu zirenak" is highlighted in blue. Below the text, there are several analysis panels. The "Fitxategia ireki" panel shows the document title "izas_egun.06-2-p0003800.2000-06-02.kirola". The "Analisis ikusi" panel shows the predicate "exist" and its arguments: "ROLE(1): thing existing VNTHETA (39.6): Agent VNTHETA (47.1-1): Theme ROLE(2): secondary attribute". The "Propbank" panel shows the predicate "exist" and its arguments: "ASPECT: FORM PERSON: TENSE: VOICE: ARG(1): 'trace'=that=political corruption ARG(2)(DIS): simply ARG(2)(NEG): n't REL: exist". The "Oharra" panel shows the note "d / aditiz-esapidea (egin)". The "Ezaldia" panel shows the sentence "Iheslariak harrepatuta, Pavel Tonkov eta Gilberto Simoni izan ziren liderrari eraso egiten saiatu zirenak." The "Osoa" panel shows the sentence "Iheslariak harrepatuta, Pavel Tonkov eta Gilberto Simoni izan ziren liderrari eraso egiten saiatu zirenak." The "Zalantzazkoa" panel shows the sentence "Iheslariak harrepatuta, Pavel Tonkov eta Gilberto Simoni izan ziren liderrari eraso egiten saiatu zirenak." The "Baztertu" panel shows the sentence "Iheslariak harrepatuta, Pavel Tonkov eta Gilberto Simoni izan ziren liderrari eraso egiten saiatu zirenak." The "ncobj" panel shows the object "Iheslariak harrepatuta, Pavel Tonkov eta Gilberto Simoni izan ziren liderrari eraso egiten saiatu zirenak". The "xmod" panel shows the modifier "Iheslariak harrepatuta, Pavel Tonkov eta Gilberto Simoni izan ziren liderrari eraso egiten saiatu zirenak". The "arg_info" panel shows the arguments "Iheslariak harrepatuta, Pavel Tonkov eta Gilberto Simoni izan ziren liderrari eraso egiten saiatu zirenak". The "nczobj" panel shows the object "Iheslariak harrepatuta, Pavel Tonkov eta Gilberto Simoni izan ziren liderrari eraso egiten saiatu zirenak". The "xcomp_obj" panel shows the object "Iheslariak harrepatuta, Pavel Tonkov eta Gilberto Simoni izan ziren liderrari eraso egiten saiatu zirenak". The "arg_info" panel shows the arguments "Iheslariak harrepatuta, Pavel Tonkov eta Gilberto Simoni izan ziren liderrari eraso egiten saiatu zirenak". The "auxmod" panel shows the auxiliary "Iheslariak harrepatuta, Pavel Tonkov eta Gilberto Simoni izan ziren liderrari eraso egiten saiatu zirenak". The "xpred" panel shows the predicate "Iheslariak harrepatuta, Pavel Tonkov eta Gilberto Simoni izan ziren liderrari eraso egiten saiatu zirenak". The "Predicate" panel shows the predicate "exist". The "Product" panel shows the product "exist". The "Proposition" panel shows the proposition "exist". The "PRP" panel shows the preposition "exist". The "REC" panel shows the recipient "exist". The "Recipient" panel shows the recipient "exist". The "Source" panel shows the source "exist". The "Stimulus" panel shows the stimulus "exist". The "Theme" panel shows the theme "exist". The "TMP" panel shows the time "exist".

4.6 irudia: VNko rolen eremuari dagokion menua.

Lantzen ari garen esaldia bukatu ondoren, gordetzeko klik egingo dugu saguarekin *osoa*, *zalantzazkoa*... botoietan. Esaldiari oharrak (ikusi 5.3 atala, erabilitako oharren azalpenerako) jartzeko aukera ere ematen du tresnak, barra-menuko *Oharrak* ikonoan klik eginda.

Bukatzeke, esan bezala, *AbarHitz*ek eremu batean ager daitezkeen balioak zerrendatu eta aukeratzeko moduan jartzen ditu. Balio horietan hizkuntzariok batzuetan aldaketak egin behar izaten ditugu, eta, ondorioz, informatikariek aldaketa horiek aplikatu behar dituzte. Hala ere, gerta daiteke etiketatzaileok nahi dugun etiketa zehaztuta ez egotea (aldaketa oraindik ez delako egin, akats bat dagoelako...); hori gertatzen denean, etiketatzaileak balio berri hori eskuz sartzeko aukera ere badu. Horrela, datua ondo sartuta geratzen da, eta gero egiten dira behar diren eguneraketak.

Bestalde, esan beharra daukagu lanean jardun ahala sortzen zaizkigun beharren arabera tresna hobetzen goazela etengabe.

4.4 Laburbilduz

Kapitulu honetan, tesi-lan honetan erabili ditugun aldez aurretiko baliabi-deak, etiketatze-lana errazteko burutu ditugun prozedura automatikoak eta etiketatze-erabili dugun tresna aurkeztu ditugu.

Garrantzi berezia eman diogu **EADBri**, hau da, **Aldezabalek (2004)** bere tesian egindako lanari, hori izan baita gure lanaren abiapuntu nagusia etiketatze-erabili dugun tresna aurkeztu ditugu. Alde batetik, **Aldezabalen (2004)** oinarri teorikoak azaldu ditugu, “alternantzia” eta “balio sintaktiko-semanticoko” kontzeptuetan arreta berezia jarritz, eta, bestetik, rol semantikoaren zerrenda eta horiek konbinatuta proposatzen diren predikatu-mota orokorrak aurkeztu ditugu.

Ondoren, gure predikatu-mailako etiketatzearen oinarrizko geruza izan den **EPEC-DEP (BDT) corpora** aurkeztu dugu labur.

Jarraian, **Aldezabalek (1998)** egindako **Levinen (1993) aditzen euskarazko itzulpenaz** aritu gara. Lan horri esker, Levinen klaseetan oinarritutako euskara-ingelesa aditzen zerrenda izan dugu eskura. Hori oso baliagarria izan zaigu, PBko ordaina aukeratzeko eta baita landutako euskarazko aditzak *Unified Verb Indexen* jasotzen diren baliabide semantikoekin (VN, FN, WN eta OntoNotes) lotzeko ere. Zerrenda hori erabilita, **EPEC corpuseko aditzak PBrekin lotu** ditugu eta informazio hori etiketatze-erabili dugun *AbarHitz* tresnan ikusgai jarri dugu, etiketatzailearen lana errazte aldera. Etiketatze-lana azkartzeko eta errazteko helburu horrekin, eta EADBn eta dependentzia-etiketetan agertzen den postposizio-atzizkien informazioa oinarri hartuta, *arg_info* etiketetan **EADBk ematen duen informazioa gehitu dugu automatikoki** (rola eta HMa).

Bukatze-erabili dugun *AbarHitz* (Díaz de Ilarraza *et al.*, 2004) tresna aurkeztu dugu, irudiak baliatuz.

102Etiketatzeko aurretikoak: baliabideak, aurreprozesuak eta tresna

Metodologia

Aurretiko tresna, baliabide eta prozedura automatikoak aplikatuta ere ez da erraza izan etiketatzeko metodologia egokiena finkatzea. Lanean aritu ahala, metodologia fintzen joan gara, eta azkenean definitu duguna azalduko dugu hemen. Atal honetan, metodologia hori definitzeko egindako urratsak azalduko ditugu ([Aldezabal *et al.*, 2013](#)).

Hiru fase nagusi bereizi ditugu:

- Lehen hurbilpena
- Metodologiaren oinarrien finkatzea
- Behin betiko metodologia

Hiru fase horiek azaltzeko, lehenik, taula batean fasearen azalpen eskematikoa emango dugu, eta, ondoren, sakonki aztertuko dugu.

5.1 Lehen hurbilpena

Lehenik, [5.1](#) taulan, atal honetan azalduko duguna adierazi dugu laburki:

- 3 hizkuntzalarik *esan*, *adierazi* eta *eskatu* aditzen 50 agerpen etiketatu
- Ereduaren egokitasuna egiaztatu
- Eskuliburuaren lehen bertsioa osatu

5.1 taula: Metodologiaren lehen hurbilpeneko pausoen laburpena.

Lehen fase honetan, bi helburu nagusi izan ditugu: i) predikatu-mailako etiketatzea egiteko hautatutako eredia euskararako egokia dela egiaztatzea, eta ii) etiketatze-prozesuan lagungarri izango zaigun eskuliburuaren lehen bertsioa osatzea. Eskuliburu hori da etorkizunean EPEC corpus osoa etiketatzeko erabiliko dugun oinarria.

Hasteko, 3 etiketatzailerik, taldean garatutako *Basque Dependency Treebank* (BDT ([Arantzabe, 2008](#); [Aldezabal et al., 2009a](#))) oinarri hartuta eta EADBn jasotako informazioa lagungarri izanda, *esan*, *adierazi* eta *eskatu* aditzen 50 agerpen etiketatu dituzte.

Bi arrazoi nagusi izan ditugu hiru aditz horiek aukeratzeko. Alde batetik, nahiko maiztasun handiko aditzak dira hirurak (*esan*: 642 (% 2,09); *adierazi*: 199 (% 0,65) eta *eskatu*: 190 (% 0,62))⁸⁵, eta, bestalde, ez dira aditz oso konplexuak anbiguotasunari begira (alde batera utzi ditugu lehen hurbilpen honetan *egin* edo *izan* moduko aditzak, oso anbiguoak direlako eta askotan agertzen direlako hitz anitzeko unitate lexikaletan).

Jarraian, eredia egokia dela egiaztatzeko egindako azterketa azalduko dugu ([Agirre et al., 2006a](#)). Horretarako, lehenik eta behin, corpora PB-VN eredura etiketatzeko oinarritzko metodologia finkatu behar da:

1. Aditza aukeratu
2. Aditzak EADBn duen informazioa jaso
3. Aditz horren ingelesezko (PB-VNko) ordaina aukeratu
4. Aditz horren corpuseko zenbait agerpen etiketatu

⁸⁵Corpuseko aditz guztien maiztasunak eta portzentajeak ikusteko, jo bedi [B](#) eranski-nera.

Erabilitako metodologiaren azalpena emateko, *esan* aditza baliatuko dugu. Behin aditza aukeratuta, ikus dezagun zehazki nola eman ditugun metodologiako pausoak:

- Etiketatzailleak EADBk aditz horrentzat ematen duen informazioa begiratzen du. *Esan* aditzarentzat EADBk bi predikatu nagusi edo adiera definitzen ditu: i) adierazpenezko jarduera, bi argumentu eta bi aldaera sintaktiko dituena; eta ii) ezaugarri-esleitzea, hiru argumentu eta aldaera sintaktiko bakarra dituena:

1. **Adierazpenezko jarduera:**

- esperimentatzailea [+giz]_ERG, gaia [-konkr]_ABS
- esperimentatzailea [+giz]_ERG, gaia _KONP

2. **Ezaugarri-esleitzea:**

- iturria [+giz]_ERG, helburua _DAT, ezaugarria _ABS

- EADBk ematen duena ikusita, etiketatzaileak PB-VNra joko du ordain egokien bila, eta horretarako, semantikaz gain, aditzen antzekotasun sintaktikoa ere kontuan izango du. Hala, ingelesezko ordain egokia bilatzen du, 4.2.1 atalean azaldu dugun Levinen sailkapenean oinarritutako euskarazko eta ingelesezko aditzen lotura (*mappinga*) erabiliz. *Esan* aditzaren kasuan, lehen adieraren ordain egokiak *to say* eta *to tell* izan daitezke, eta bigarren adierarena *to call*. 5.2 taulan ikus ditzakegu hiru ordain horien PB-VNko *rolsetak*:

Adierazpenezko jarduera		Ezaugarri-esleitzea
say.01	tell.01	call.01
vncls: say-37.7	vncls: tell-37.1	vncls: dub-29.3
Arg0: agent	Arg0: agent	Arg0: agent
Arg1: topic	Arg1: topic	Arg1: theme
Arg2: recipient	Arg2: recipient	Arg2: predicate
Arg3: attributive		

5.2 taula: *say.01*, *tell.01* eta *call.01* adieren sarrerak PB-VN-n

5.2 taulan ikus daitekeen bezala, lehen adierarentzat aukeratu ditugun *say.01* eta *tell.01* aditz-adieretan, EADBn zehazten ez den hirugarren

argumentu bat dago, euskaraz datiboan (DAT) gauzatuko litzatekeena. Askotan gertatzen zaigu PB-VNk EADBk baino argumentu gehiago markatzea⁸⁶, hau da, EADBN adjuntu gisa aztertzen diren osagai asko PB-VN-n argumentu gisa lantzea. Bestalde, *say.01* en kasuan, *tell.01* en agertzen ez den Arg3_ *Attributive* argumentu bat dugula ikusten dugu⁸⁷. Euskaraz, *Attributive* hori instrumentalean (INS) ala *-i buruz* postposizioaren bidez gauzatzen da. Gure helburua PB-VN eredia euskarazko corpusa etiketatzeko baliagarria den ala ez frogatzea denez, horrelako kasuetan, eta adierari kalterik egiten ez dionez, euskarazko aditzari PB-VNk dakarrena gehitu diogu (horrelako kasuetan hartutako erabakiak eta irizpideak xehe azaltzen ditugu 6.1 atalean).

Bigarren adierari dagokionez, ordain egokia, esan bezala, *call.01* dela iruditu zaigu. Kasu honetan EADBko adiera eta ingelesezko aditz-adiera bat datoz argumentu-kopuruan; beraz, ez dago inolako arazorik *esan* aditzaren bigarren adiera PB-VNko *call.01* ordainaren argumentu eta rolekina etiketatzeko.

- Aditzaren agerpenak PB-VNk ematen duen informazioari jarraiki etiketatzen dira. Ikusi dugunez, bai lehen adieran, bai eta bigarrenean ere, ingelesezko eta euskarazko aditzen egitura sintaktiko-semanticoko oso antzekoak dira. Honenbestez, informazio hori guztia baliatuta *esan* aditzaren corpuseko agerpenak etiketatzeko prest gaude.

Lehen azterketa honi esker, eta hiru aditzen (*esan*, *adierazi*, *eskatu*) 50 agerpenak arazorik gabe etiketatu ondoren, PB-VN eredia euskarazko aditzak predikatu-mailan etiketatzeko egokia dela ondorioztatu ahal izan dugu (Agirre *et al.*, 2006a).

EADB abiapuntu egokia da euskal aditzak lantzen hasteko eta, ingelesezko aditzekiko hainbat desberdintasun aurkitu baditugu ere (batez ere argumentu/adjuntu bereizketa egiteko erabilitako irizpideak direla eta), orokorrean egokitzapenak modu errazean egin daitezkeela ikusi dugu (ikusi 6.1 atala).

Era berean, etiketatze-prozesua burutzeko gidalerroen lehen bertsioa osatu dugu. Eskuliburu hau etiketazaileen eta lexikoiaaren editoreen lanerako

⁸⁶PBn, askotan, argumentu edo adjuntua den erabakitzeo corpuseko maiztasunean oinarritzen dira.

⁸⁷VNk ez dakar argumentu hori, beraz, badirudi ingelesez ere ez dagoela argi argumentu ala adjuntu den.

oinarrizko tresna da, eta, aditzak landu ahala, osatuz goaz etengabe.

5.2 Metodologiaren oinarrien finkatzea

Hasteko, [5.3](#) taulan, atal honetan azalduko ditugunak adierazi ditugu laburki:

- **EADBko 100 aditzen zenbait agerpenen eskuzko etiketatzea**
 - 3 etiketatzailerik aditz bakoitzeko 40 agerpen desberdin (22 aditz (2.640 agerpen))
 - 3 etiketatzailerik aditz bakoitzeko 20 agerpen desberdin (77 aditz (4.620 agerpen))
- **Emaitzak**
 - * Eskuliburuaren bertsio osatua
 - * EADBn landuta dauden aditzen lexikoa (BVI)
- **Metodologiaren ebaluazioa: 3 aditz: *adierazi, izan eta etorri***
 - 2 etiketatzailerik aditz bakoitzeko 20 fitxategi bi fasetan:
 - * 1. fasea: 143 agerpen guztira
 - * 2. fasea: 207 agerpen guztira
 - Ondorioak metodologiari begira
- **EADBko 100 aditzen gainontzeko agerpenen etiketatze erdi-automatikoa**
 - Abiapuntua:
 - * Eskuliburuaren bigarren bertsioa irizpide finduekin
 - * EADBko aditzen PB-VNren arabera eskuzko lexikoa (BVI) eta konbinazio sintaktiko-semantikoak
 - Corpusetik konbinazio sintaktiko-semantikoak atera eta datu horiek automatikoki esleitu. Ondoren, eskuz errebisatu
 - Lexikoa erabili etiketak automatikoki sortzeko. Ondoren, eskuz errebisatu (2 etiketatzailerik)
 - Konbinazio sintaktiko-semantikoetatik ateratako datuekin lexikoa aberastu
- **Levin klaseen baliagarritasunaren azterketa (97 aditz)**
 - Etiketatzailerik bakarrik

5.2.1 EADBko 100 aditzen zenbait agerpenen eskuzko etiketatzea eta BViren sorrera

Bigarren fase honetan, eta ereduaren egokitasuna frogatu ondoren, corpusaren eskuzko etiketatzea eraman da aurrera. Gure helburua oraingoan, corpus osoa etiketatzea da, baina, hasteko, EADBn landuta dauden 100 aditzak⁸⁸ hautatu ditugu, horien euskarazko azterketa egina dugulako eta aurretik egingandako lana aprobetxatzeko bidea ematen digulako.

Fase honetan, 3 etiketatzailer egon dira lanean eta bakoitzak aditz bakoitzeko 40 agerpen etiketatu ditu; hots, aditz bakoitzeko 120 agerpen inguru etiketatu dira. Helburua aditz bakoitzeko eskuz etiketatutako lagin bat izatea da, ondoren, aditzaren gainontzeko agerpenak prozesu automatikoen bidez etiketatzeko. Aztertutako lehen 22 aditzen 120 agerpenak etiketatuta izan ondoren (guztira 2.640 agerpen), ikusi genuen ez zela beharrezkoa horrenbeste agerpen etiketatzea, adierak eta egiturak behin eta berriz errepikatzen zirelako. Hala, gainontzeko aditzetan 20 agerpen etiketatzea erabaki genuen; hau da, 60 agerpen hiru etiketatzailer artean, guztira 4.620 (77 aditz x 60 agerpen).

Urrats honetan egindako lanari esker, emaitza bikoitza lortu dugu. Alde batetik, eskuliburuaren bertsio osatua (Aldezabal *et al.*, 2010b), eta, bestetik, EADBn landuta dauden eta EPECen agertzen diren 99 aditz horien lexikoa PB-VN ereduari jarraiki (*Basque Verb Index* (BVI)), hori guztia eskuz etiketatutako 7.260 (2.640 + 4.620) agerpenetako laginean oinarrituta.

5.2.2 Metodologiaren ebaluazioa

Aurrera egin aurretik, gure metodologia egokia den ziurtatzeko eta eskuliburuaren baliagarritasuna bermatzeko, oinarritzko ebaluazio bat egin dugu (Aldezabal *et al.*, 2011). Ebaluazio hori egiteko, hiru aditz aukeratu ditugu: *adierazi* (199 agerpen, corpuseko agerpen guztien % 0,65), *izan* (4.832 agerpen (% 15,72)) eta *etorri* (229 agerpen (% 0,75)). Ebaluazioa bi faseetan egin dugu. Lehen fasea burutu eta emaitzak aztertu ondoren, hainbat ondorio atera ditugu, eta horien arabera, eskuliburua osatu eta moldatu dugu. Jarraian, bigarren faseari ekin diogu, lehenengotik ateratako ondorioek emaitzak hobetzea lortzen duten ala ez egiaztatzeko.

⁸⁸4.2.2 atalean esan bezala, hauetatik 99 agertzen dira EPEC-DEP corpusean; *loratu* da agertzen ez den aditza.

Fase bakoitzean aditz bakoitzeko 20 fitxategi⁸⁹ etiketatu ditu etiketatzaile bakoitzak (fase bakoitzean lagin desberdina), eta 2 etiketatzaileen arteko adostasuna neurtu dugu. 20 fitxategi horietan aditz bakoitzeko hainbat agerpen aurkitu ditugu. Lehen fasean, *adierazi* aditza 27 aldiz agertu da, 74 aldiz *izan* aditza, eta 42 aldiz *etorri* aditza (143 agerpen guztira); eta bigarren fasean, 27 aldiz *adierazi*, 138 aldiz *izan*, eta 42 aldiz *etorri* (207 agerpen guztira).

Hiru aditz horiek aukeratu ditugu, kasuistika zabala aztertzeko aukera emango digutelakoan. *Adierazi* aditzak adiera bakarra du EADBN, baina sarri erabiltzen da kazetaritza-testuetan, eta gure corpusaren zati handi bat kazetaritza-testua denez, corpusaren zati handi bat erraz etiketatzeko aukera ikusi dugu aditz hori aztertuta. *Izan* aditzak 3 adiera ditu eta, gainera, corpusean maiztasun gehien (corpuseko agerpen guztien % 15,72) duen aditza da. PropBank-en ez dute *to be* aditza etiketatu; horrelako aditzak euskarri-aditzak direla eta, hortaz, eduki semantikoa galtzen dutela esaten dute. Baina guk saiakera egin nahi izan dugu, horrelako aditz batek sor ditzakeen arazoak aztertzeko aukera izateko. Azkenik, *etorri* aditzak 4 adiera ditu, eta, gainera, horietako batzuk ez dira oso erraz bereizten. Horrenbestez, alde aurretik esan daiteke honek sortuko dituela zailtasun handienak etiketatzean. Gainera, oso maiz agertzen da hitz anitzeko unitate lexikaletan (*bat etorri, burura etorri*, etab.), eta hori ere desadostasunerako iturri izan daiteke.

Gure etiketatzeko metodologiari jarraiki, lehenik eta behin EADBN dagoen informazioa jaso, eta aditz horien adiera bakoitzarentzat ingelesezko ordain egokiena bilatzen saiatu gara; *adierazi* (5.4); *izan* (5.5) eta *etorri* (5.6):

⁸⁹4.1.2 atalean esan bezala, fitxategi bakoitzean *Euskaldunon Egunkariako* berri bat edo *XX. mendeko euskararen corpus estatistikotik* bildutako 1.000 hitz inguruko obra zatiak daude.

EADB	PB-VN	
<i>adierazi1- jarduera</i>	state.01	express.01
esperimentatzailea (ERG) gaia [-biz, -konkr] (ABS)	Arg0: agent Arg1: topic Arg2: recipient Arg3: - (attributive)	Arg0: agent Arg1: theme Arg2: recipient

5.4 taula: *adierazi* aditzaren EADBko informazioa, dagozkion PB-VN ordainen informazioarekin alderatuta.

EADB	PB-VN
<i>izan1- entitate baten kokapena</i>	be.02
gaia (ABS) kokapena (INE)	Arg1: - (thing that is)
<i>izan2- entitate baten deskribapena</i>	be.01
gaia (ABS/KONP) ezaugarria (ABS)	Arg1: theme Arg2: attribute
<i>izan3- entitate baten edukitzea</i>	have.03
edukitzailea (ABS)[+biz] edukia (ABS)[-biz]	Arg0: theme1 ⁹⁰ Arg1: theme2

5.5 taula: *izan* aditzaren EADBko informazioa, dagozkion PB-VN ordainen informazioarekin alderatuta.

⁹⁰VN-n etengabe ari dira lexikoia eguneratzen eta, azkenaldian, 2.1.3 atalean esan dugun bezala, euren rolak LIRICS proiektuko rol estandarrekin berdintzeko bidea hartu dute. Horren ondorioz, hainbat rol aldatu egin dituzte; esaterako, *have.03* adierako Arg0 argumentuari *Pivot* rola ematen diote 3.2 bertsioan. Tesi-lan hau hasi genuenean, VN-n zeuden datuak mantendu ditugu BVI lexikoian.

EADB	PB-VN
etorri1- desplazamendua	come.01
gai ukitua (ABS)	Arg1: theme
abiapuntuko kokapena/bidea (ABL)	Arg2: - (extent)
helburuko kokapena (ALA)	Arg3: source ⁹¹
	Arg4: destination
etorri2- sortze-prozesua	come.03
gai sortua (ABS) [-konkr]	Arg1: theme
iturria (ABL) [-biz] / (DAT) [+biz]	Arg2: source
etorri3- edukitzea	be.02
edukia (ABS)[-biz]	Arg1: - (thing that) is
edukitzailea (INE)[-biz]	
etorri4- deskribapena	be.01
gaia (ABS)	Arg1: theme
ezaugarria (ABS)	Arg2: attribute

5.6 taula: *etorri* aditzaren EADBko informazioa, dagozkion PB-VN ordainen informazioarekin alderatuta.

Lehen fasean, etiketatzaile bakoitzak bere aldetik eta bestearekin hitz egin gabe etiketatu ditu agerpenak, eta bi etiketatzaileek lagin bera etiketatu dute. Fase hori bukatzean aztertu dugun lehenengo gauza adiera aukeratzean etiketatzaileek (E1 eta E2) izan duten adostasun-maila izan da. Emaitzak 5.7 taulan ikus daitezke.

Gauza bera neurtu dugu Cohen's Kappa neurria⁹² erabilita (Siegel eta Castellan, 1988; Carletta, 1996), eta emaitzak 5.8 taulan daude ikusgai.

5.7 taulan ikusten denez, adiera, eta ondorioz, ingelesezko ordaina aukeratzean, adostasun maila nahiko altua izan da *adierazi* eta *izan* aditzetan; ez, ordea, *etorri* aditzean.

Adierazi aditzean adostasuna erabatekoa izan da, baina hau alde aurretik espero genuen zerbait zen, izan ere, aditz horrek adiera bakarra baitu. Hala ere, adiera bakarra izan arren, bi ordain erabil daitezke *adieraziren* or-

⁹¹Aurreko oin-oharrean azaldutakoari jarraiki, gure lana hasi genuen garaian, Arg3 argumentuari *Source* rola esleitzen zion VNk, baina gaur egungo datuak (3.2 bertsioa) begiratzen baditugu, *Initial Location* jartzen duela ikusiko dugu (hain zuzen EADBtik hurbilago).

⁹²Etiketatzaileren arteko desadostasuna/adostasuna neurtzeko gehien erabiltzen den neurria da hori.

Aditza	PB-VN	E1	E2	Ados.	Desados.	%
Adierazi	state.01	49	49	49	0	100
Adierazi	express.01	5	5	5	0	100
Izan	be.01	143	139	139	4	97.20
Izan	be.02	12	14	12	2	85.71
Izan	have.03	27	29	27	2	93.10
Etorri	come.01	29	26	22	11	89.65
Etorri	come.03	2	7	0	9	0
Etorri	be.01	2	0	0	2	0

5.7 taula: Etiketatzailen arteko adostasun-maila adiera aukeratzean.

adierazi	1.000
izan	0.939
etorri	-0.120

5.8 taula: Cohen's Kappa neurriaren emaitzak adiera kontuan hartuta.

dain gisa ingelesez, eta bata ala bestea aukeratzean ere erabateko adostasuna egon da etiketatzailen artean.

Izan aditzaren kasuan 3 adiera genituen aukeran, eta hemen ere adostasun maila oso altua izan da. Hau ezusteko atsegina izan da, aditz kopulatiboa izaki, adierak desberdintzea zaila izango zela pentsa baikenezakeen; baina corpusa etiketatzean bestelako errealitatea topatu dugu, eta ez da batere zaila izan adierak desberdintzea eta adostasuna izatea etiketatzailen artean. 178 kasutan ordain berbera aukeratu dute etiketatzaille biek eta 8 aldiz ordain desberdina. (81)-ean ordain desberdina aukeratu duen adibide bat erakusten dugu:

(81) *Kasparovi kendu dio Kramnik gazteak koroa, hamabost urtean harena izan ostean.*

Etiketatzaille batek *have.03* jarri du eta besteak *be.01*. Etiketatzaille batek objektua edukitzea deskribapen gisa ulertu du (*be.01*), eta besteak edukitzearen ekintza gisa (*have.03*).

Etorri aditzaren kasuan, bi etiketatzailleek 22 aldiz aukeratu dute ordain berbera eta beste 22 aldiz ez dira bat etorri; beraz, Kappa neurria ere oso baxua izan da. Adostasuna batez ere lehen adiera aukeratu denean egon da. Gainontzeko adierak aukeratzean nahasmen handiagoa izan da, eta horrek

pentsaraz diezaguke adiera horien arteko mugak ez direla oso argiak. (82)-an desadostasun-adibide bat ikus dezakegu:

(82) *Oso gaztetatik datorrio xakeko zaletasuna.*

Etiketatzaille batek lehen adiera aukeratu du “entitate baten desplazamendua” (*come.01*), eta bigarrenak bigarren adiera “sortze-prozesua” (*come.03*).

Bestelako neurketak ere egin ditugu Kappa neurria erabilita: adiera (ingelesezko ordaina) eta balentzia kontuan hartuta lehenengoa (5.9 taula), eta adiera, balentzia eta rola kontuan hartuta bigarrena (5.10 taula):

Ingelesezko ordaina + balentzia	
adierazi	1.000
izan	0.950
etorri	0.232

5.9 taula: Cohen’s Kappa neurriaren emaitzak PBko ordaina (adiera) eta balentzia kontuan hartuta.

Ingelesezko ordaina + balentzia + rola	
adierazi	0.783
izan	0.846
etorri	0.231

5.10 taula: Cohen’s Kappa neurriaren emaitzak PBko ordaina (adiera), balentzia eta rola kontuan hartuta.

5.10 taulan rola gehitzerakoan *adierazi* eta *izan* aditzen emaitzak jaitsi egiten direla ikusten dugu (batez ere *adierazi* aditzean). Pentsatzekoa da zenbat eta parametro gehiago kontuan hartu, orduan eta desadostasun handiagoa egongo dela. Hala ere, eskuz errebisatu ditugu emaitzak eta desadostasunak adjuntuen roletatik datozela ikusi dugu; beraz, jaitziera hori ez da adierazgarria gure azterketari begira. Fenomeno hori PBn ere aipatzen dute (Palmer *et al.*, 2005). Etiketatzailleen adostasuna neurtzen dutenean, adjuntuak edo ArgM-ak kontuan hartuta, emaitzak pixka bat jaisten direla ikusten da (ikusi aurrerago 5.14 taula), eta rola aukeratzean desadostasunik handienak ADV, MNR eta DIS rola esleitzerakoan gertatu direla esaten dute:

“Discrepancies between annotators tended to be less on numbered arguments than on the selection of function tags [...]. Certain types of function tags, particularly ADV, MNR, and DIS, can be difficult to distinguish” (Palmer *et al.*, 2005: 16).

Ebaluazioaren lehen fasea bukatuta, hainbat ondorio atera ditugu bigarren fasean lagungarri izango zaizkigulakoan.

Hasteko, eskuliburuan zehaztu beharreko hainbat gai aipatuko ditugu.

5.2.2.1 Adjuntuak

Adjuntuen edo modifikatzaileen atala zehaztu beharra ikusi dugu, lehen esan bezala, hainbat desadostasun hemendik etorri baitira. Hala ere, ezin dugu ahaztu badela hizkuntzaren baitan dagoen anbiguotasuna, eta horretan etiketatzailearen senak agintzen duela. Esaterako, *hitzaldian adierazi* kasuan, inesiboak zer adierazten du? Denbora ala lekua? Eta *amaitzear zegoela* moduko adibide batean, zer ulertu behar da? Modua ala denbora? Badirudi biak izan daitezkeela zuzenak etiketatzailearen interpretazioaren arabera. Horregatik, honelako anbiguotasunek sortutako desadostasunak ez dira adierazgarriak gure azterketari dagokionean.

5.2.2.2 Hitz Anitzeko Unitate Lexikalak

Hitz Anitzeko Unitate Lexikalek (HAUL) ere desadostasunak sortu dituzte, etiketatzaile batek egitura batzuk HAUL gisa hartu dituelako eta besteak ez. Esaterako, (83) adibidean etiketatzaile batek *zertara etorri* HAUL gisa tratatu du (eta, beraz, ez du esaldia etiketatu), eta besteak, aldiz, ez (eta ondorioz, etiketatu egin du).

(83) *Sharonen jarrera probokatzailea zertara datorren galdetu zuen Mubarak.*

Jakina da HAULen gaia irristakorra dela eta HAUL asko ez daudela aurrez definituta. Eskuliburuan ez zegoen ondo zehaztuta horrelakoetan nola jokatu, eta, hala, hau ere desadostasun-iturri izan da HAULetan parte hartzeko joera duten aditzetan.

5.2.2.3 Bokatiboak

Bokatiboen arazoa ere aipatu beharra dago. Eskuliburuan ez dugu adierazi bokatiboak nola etiketatu behar diren. Hala, agertu den kasu bakarrean etiketatzaile bakoitzak mementoko erabakia hartu du. Horrelako kasuak asko ez badira ere, komeni da eskuliburuan azaltzea nola etiketatuko ditugun. PropBank-eko gidalerroei jarraiki, bokatiboak ArgM_ *DIS* etiketatzea erabaki dugu, eta horrela zehaztu dugu eskuliburuan.

Horiek dira eskuliburua osatze aldera ateratako ondorioak; baina badira bestelako desadostasunak, zerikusi gehiago dutenak etiketatzen hasi aurretiko aditz-sarreraren zehaztapen-lanarekin, eskuliburuarekin baino.

5.2.2.4 EADB edo PB-VN jarraitu? Sarreraren zehaztapen-lana

Ikusi bezala, adostasun nabarmena egon da aditzaren adiera aukeratzera-koan, baina desadostasunak ondoren etiketatu beharreko eremuetan agertu dira, besteak beste, balentzia eta roletan. Etiketatzaile batek EADBN dagoen informazioa jarraitu du etiketatzerakoan, eta besteak PropBank-ek eskaintzen zion informazioa. Horrelako desadostasunak batez ere *etorri* aditzean aurkitu ditugu.

Adibidez, *come.01* adierak Arg2_ *Extent* argumentu bat du, euskaraz posible ez dena (PBN bertan ere *rare* marka du, eta ez dago adibiderik). Beraz, euskaraz rol hori ez da existitzen (ikusi gai honi buruz 6.2.1 atala). Etiketatzaile batek Arg2 hori erabili du existitzen den “abiapuntuko kokapena” rolarentzat, eta besteak, aldiz, Arg2 hori hutsik utzi eta Arg3 jarri dio “abiapuntuko kokapenari”. Horren ondorioz, hainbat desadostasun egon dira balentzia eta rolak etiketatzerakoan, nahiz eta adiera hautatzean bi etiketatzaileak bat etorri.

Arg1 argumentua etiketatzean ere desadostasun bat aurkitu dugu. PBk, jakina denez, *Theme* rola esleitzen dio beti Arg1 argumentuari, baina, esana dugunez, guk ez dugu irizpide berbera jarraitu gure etiketatze-lanean (ikusi 3.1 atala), PBko gidalerroei jarraiki, hain zuzen ere (Babko-Malaya, 2005). Gidalerro horietan jartzen duenaren arabera, balentziak eta rolak elkarren arteko dependentziarik ez duten bi maila dira. Beraz, *come.01* gisako aditz inakusatibo batean (aldaera iragangaitza baino onartzen ez duena), gure us-tetan ekintza egiten duena eta jasaten duena entitate berbera dira, eta, ondorioz, Arg0 etiketatzea proposatzen dugu. Horrelako kasuetan, etiketatzaile batek gure eskuliburua jarraitu du, eta besteak PBko etiketatzeko modua

jarraitu du oharkabeen.

Desadostasun-iturri horietatik ondorio hauxe ateratzen dugu: aditza etiketatzen hasi aurretik, aditz horren **adiera guztien zehaztapena egin behar da**; hau da, etiketatzen hasi aurretik, etiketatzailen artean aditzaren sarrera kanonikoa proposatu behar da, argumentuak eta rolak zehatz azalduz. Hala, zehaztapen-lan hori egin dugu aukeratutako aditzekin, ebaluazioaren bigarren fasea hasi baino lehen:

adierazi

1- **Jarduera**

VNrs⁹³: state.01 / express.01

BAL: Arg0, VNrol: agent, EADBrol: esperimentatzailea_ERG

BAL: Arg1, VNrol: topic, EADBrol: gaia_ABS/ELA_KONP, HM: -biz, -konkr

izan

1- **Entitate baten kokapena**

VNrs: be.02

BAL: Arg0, VNrol: theme, EADBrol: gaia_ABS

BAL: Arg1, VNrol: location, EADBrol: kokapena_INE

2- **Entitate baten deskribapena**

VNrs: be.01

BAL: Arg0, VNrol: topic, EADBrol: gaia_ABS/KONP

BAL: Arg1, VNrol: attribute, EADBrol: ezaugarria_ABS

3- **Entitate bat edukitzea**

VNrs: have.03

BAL: Arg0, VNrol: theme, EADBrol: edukitzailea_ERG

BAL: Arg1, VNrol: theme, EADBrol: edukia_ABS

etorri

1- **Kokapen-aldaketa**

⁹³Hemen erabiltzen ditugun laburdurak 2.3 atalean daude azalduta.

VNrs: come.01

BAL: Arg0, VNrol: theme, EADBrol: gai ukitua_ABS

BAL: Arg1, VNrol: source/path, EADBrol: abiapuntuko kokapena/bidea_ABL

BAL: Arg2, VNrol: destination, EADBrol: helburuko kokapena_ALA

2- Sortze-prozesua

VNrs: come.03

BAL: Arg0, VNrol: theme, EADBrol: gai sortua_ABS, HM: -konkr

BAL: Arg1, VNrol: source, EADBrol: iturria_ABL, HM: -animate/_DAT,

HM: +animate

3- Entitate bat edukitzea

VNrs: be.02

BAL: Arg0, VNrol: theme, EADBrol: edukia_ABS, HM: -biz

BAL: Arg1, VNrol: location, EADBrol: edukitzailea_INE, HM: -biz

4- Entitate baten deskribapena

VNrs: be.01

BAL: Arg0, VNrol: topic, EADBrol: gaia_ABS

BAL: Arg1, VNrol: attribute, EADBrol: ezaugarria_ABS

Sarrerren zehaztapen-lan hau egin ondoren, etiketatzaile biek lagin berdina etiketatu dute elkarrekin hitz egin gabe, eta emaitzak aztertuta, *etorri* aditzaren emaitzak nabarmen hobetu dira, [5.11](#) taulan ikusten den bezala:

adierazi	0.854
izan	0.910
etorri	0.781

5.11 taula: Cohen's Kappa neurriaren emaitzak, sarrerak zehaztu ondoren.

Adierazi aditzaren kasuan emaitzak jaitsi egin dira pixka bat, baina ez dirudi desadostasun hori garrantzitsua denik, kasu gehienetan etiketatzaile batek *state* eta besteak *express* aukeratu duelako gertatu baita.

[5.12](#) taulan, adieraz gain, balentzia ere kontuan izanda lortutako emaitzak ikus daitezke, eta [5.13](#) taulan, berriz, adiera, balentzia eta rola kontuan izanda lortutakoak:

Ingelesezko ordaina + balentzia	
adierazi	0.922
izan	0.930
etorri	0.818

5.12 taula: Cohen's Kappa neurriaren emaitzak, balentzia ere kontuan hartuta.

Ingelesezko ordaina + balentzia + rola	
adierazi	0.808
izan	0.869
etorri	0.740

5.13 taula: Cohen's Kappa neurriaren emaitzak, balentzia eta rola ere kontuan hartuta.

5.9 eta 5.12 taulak konparatuta, *etorri* aditzaren emaitzak asko hobetzen direla ikusten dugu (0.232 vs 0.818), eta hori zen gure helburu nagusia; izan ere, aditz horretan izan ditugu arazorik larrienak. Beste bi aditzetan (*izan* eta *adierazi*), aldiz, emaitzak pixka bat jaitsi dira, baina ez dugu us-te adierazgarria denik, 5.10 eta 5.13 taulak konparatuta ikusten den bezala, neurketan rola ere sartzen dugunean, emaitza guztiek gorantz egiten baitute.

2. fase honetan, etiketatu beharreko hiru eremuak neurtuta (adiera, balentzia eta rola), aditz guztien emaitzak hobetu egiten dira. Dirudenez, beraz, gure zehaztapen-lanak bere ondorio onak izan ditu.

Argi dago, honenbestez, aditz bakoitzaren sarrera alde aurretik landuta eta zehaztuta badago, etiketatze-lanaren kalitatea hobea dela, eta, ondorioz, kalitatezko eskuzko etiketatzea bermatzeko ezinbestekoa dela aditzen sarre-rak banan-banan lantzea etiketatzen hasi aurretik.

Gure ebaluazio honen helburua ez da izan corpus etiketatuaren kalitatea neurtzea, baizik eta metodologiaren egokitasuna balioztatzea eta desadostasunen arrazoiak bilatzea. Horregatik, ezin ditugu etiketatzailen arteko adostasunari buruzko gure emaitza hauek 5.14 taulan ikus daitezkeen PBn lortutako emaitzekin konparatu (Palmer *et al.*, 2005).

Hala ere, gure irudipena da PBn ageri diren emaitza altu horiek zerikusizuzena dutela Arg0/Arg1 argumentuak etiketatzeko hartutako erabakiekin (ikus 3.1 atala). Izan ere, Arg0 eta Arg1 argumentuak zuzenean daude lotuta subjektu eta objektu funtzio sintaktikoekin, hots, corpusean gehien agertzen diren elementuekin. Gainera, Arg0 guztiak *Agent* eta Arg1 guztiak *Theme*

ArgM-ak kontuan hartuta	Rola identifikatzean	0.93
	Rola sailkatzean	0.93
	Bien arteko konbinazioa	0.91
ArgM-ak kontuan izan gabe	Rola identifikatzean	0.94
	Rola sailkatzean	0.96
	Bien arteko konbinazioa	0.93

5.14 taula: Cohen's Kappa neurriaren emaitzak PropBank-en.

etiketatzea erabaki dutenez, adostasun maila altua izatea espero daitekeena da. Hala ere, azterketa sakonagoa egin beharko genuke gai honen gainean, ondorio garbiak atera ahal izateko.

Beste hizkuntzetan garatutako PropBank-etan ez da etiketatzailen arteko adostasuna neurtu. Gehienetan etiketatze-prozesua erdi-automatikoa izan da, eta ez dute ebaluazio orokorrik ematen. Aurkitu dugun kasu bakarra frantsesarena izan da (Van Der Plas *et al.*, 2010). Horiek frantseseko corpus bat eskuz etiketatzeko PBk izan dezakeen baliagarritasuna neurtu dute. Ez dute Kappa neurria erabili, *F-measure*⁹⁴ baizik (5.15 taulan ikus daitezke emaitzak), eta, beraz, ezin ditugu euren emaitzak zuzenean gurekin konparatu.

	Aditza aukeratzean	Argumentua aukeratzean
Entrenamendu-fasea	% 46	% 62
Etiketatzeko-fasea	% 59	% 69

5.15 taula: Frantseseko corpusean aditzak eta argumentuak aukeratzean etiketatzailen artean egon den adostasunaren *F-measure* neurria erabilia.

Hala ere, PB erdua ingelesa ez den beste hizkuntza batean etiketatzeko egokia den ala ez aztertzean, gure lanarekiko antzekotasunak aurkitu ditugu, bai metodologian, bai eta emaitzen azterketatik ateratako ondorioetan ere.

Metodologikoki 3 antzekotasun topatu ditugu:

- Entrenamendu-fasearen ondoren etiketatzailen arteko adostasuna biltzea.

⁹⁴Ohiko F puntuazio orekatua (*F1 score* edo *F-measure*) doitasunaren eta estalduraren batezbesteko harmonikoa da.

- PBko ordain egokia aukeratzea eta adostasun maila neurtzea.
- Ebaluazio kualitatiboa egitea.

Lehen antzekotasuna etiketatzeke proposatutako metodologian ikus dezakegu. [Padóri \(2007\)](#) jarraiki, bi fase nagusi bereizi dituzte metodologian: i) entrenamendu-fasea eta ii) etiketatze-fasea. Baina entrenamendu-fasearen barruan, etiketatzailen arteko adostasuna bilatzeko fase bat gehitu dute, eta fase horren ondoren entrenamendu-fasean lortutako emaitzak asko hobetzen direla egiaztatu dute. Hori gure metodologian gehitutako aditz-sarrerren zehaztapen-lanarekin bat etor daiteke. Etiketatzailak aditzaren sarrera definitzeko ados jartzen dira, eta, horren ondoren etiketatzean, emaitzak asko hobetzen dira.

Bigarren antzekotasunari dagokionez, guk egin dugun bezala, horiek ere frantsesezko aditzaren PBko ordain egokia aukeratzen dute, eta, ondoren, horren argumentuekin etiketatzen dute frantsesezko aditza. Beraz, guk bezala, PBko ordaina aukeratzean egon den adostasuna neurtu dute, eta % 59ko adostasuna dagoela ikusi dute. Hala ere, aditza aukeratzean PBko ordaina kontuan izan beharrian, VNko klasea kontuan hartuz gero (hau da, bi etiketatzailak PBko ordain desberdina aukeratu dute, baina biak dira VN-klase berekoak), emaitzak asko hobetzen dira eta % 81eko adostasuna lortzen dute.

Hirugarren antzekotasun garrantzitsua, emaitzen azterketa kuantitatiboa egiteaz gain, azterketa kualitatiboa ere egin dutela izan da. Euskaraz ikusi dugunaren antzera, HAULak desadostasun-iturri nabarmena direla ikusi dute. Askotan, etiketatzailak HAUL bat topatzean, ez du jakiten nola jokatu eta ez dio ingelesezko ordainik esleitzen. Beraz, eskuliburuan horrelakoan aurrean nola jokatu zehaztu behar dela ondorioztatzen dute (guk egin dugun bezala). Esapide idiomatikoetan ere arazoak izan dituzte, horrelakoetan zaila baita zuzenean ordain egokia topatzea. Hori arazo orokorra da hizkuntzen arteko azterketak egiten direnean ([Burchardt et al., 2009](#)), eta, guk egin dugunaren antzera, azterketa sakonagoa behar duen gaia dela ondorioztatzen dute.

Laburbilduz, corpusaren ebaluazio orokorra egin ez badugu ere, ebaluazio honetatik ateratako ondorioekin, eskuliburuaren bigarren bertsioa osatu dugu ([Aldezabal et al., 2010b](#)), eta, ikusitako hutsuneak betetzeaz gain, etiketatzeke erabiliko ditugun irizpideak ere findu eta zehaztu ditugu, eskuzko lana errazte eta hobetze aldera. Bestalde, aditzen sarrerak lantzearen garrantzia ere azpimarratu dugu. Aditz bakoitza etiketatzen hasi aurretik, adierak,

argumentuak, rolak, PB-VNra egin beharreko egokitzapenak etab. aztertu behar direla eta aditzaren sarrera lexikala proposatu behar dela argi geratu da. Hala, EADBn landu ziren 99 aditz horien lexikoa (BVI) osatu dugu, PB-VN ereduari jarraituz (ikusi [A](#) eranskina).

5.2.3 EADBko 100 aditzen gainontzeko agerpenen etiketatze erdi-automatikoa

Lexikoi hau oso tresna baliagarria da etiketatu gabe geratu diren agerpenen etiketatze erdi-automatikoa aurrera eramateko. Corpusean agertu diren 99 aditz horien agerpen guztiak etiketatze bide erdi-automatikoak aztertu ditugu, lan hau guztia eskuz egiteak denbora gehiegi eskatuko lukeelako.

5.2.3.1 Corpuseko konbinazio sintaktiko-semanticok baliatuz

Hasteko, eskuz etiketatutako corpusetik aditz bakoitzaren “lexikoa” edo konbinazio sintaktiko-semanticok automatikoki erauzi ditugu; hau da, aditz-adiera bakoitzeko, *arg_info* etiketen konbinazio desberdin guztien zerrenda lortu dugu. Adibidez, *aldata* aditzaren kasuan, [5.16](#) taulan ikusten diren konbinazio sintaktiko-semanticok erauzi ditugu automatikoki:

Aditza	PropBank ordaina	VerbNet rola	Post.-atz. ⁹⁵
aldata	alter.01/change.01	Agent-Patient-NEG	erg-par-neg
aldata	alter.01/change.01	Patient-NEG	abs-neg
aldata	alter.01/change.01	Patient-TMP	abs-ine
aldata	alter.01/change.01	Patient-ADV	abs-abs
aldata	alter.01/change.01	Patient-MNR	abs-gen
aldata	alter.01/change.01	Patient-LOC	abs-
aldata	alter.01/change.01	Patient-PRP	abs-helb
aldata	alter.01/change.01	Agent-Patient	erg-abs

5.16 taula: *aldata_alter.01/change.01* adieraren konbinazio sintaktiko-semanticok.

Eskuzko etiketatze ateratako konbinazio sintaktiko-semanticok horiek izanda, eta berriz ere prozedura automatikok erabilita, postposizio-atziz-

⁹⁵Post.-atz.: Postposizio-atzizkia.

ki bakoitza zein rolekin eta zein portzentajetan agertu den atera dugu⁹⁶. Adibidez, 5.17 taulan, *aldata_change.01/alter.01* adieran postposizio-atzizki bakoitza zein rolekin eta zein portzentajetan agertu den ikus dezakegu⁹⁷:

Postposizio-atzizkia	VN rola	Portzentajea
ABL	Product	% 50
ABL	Material	% 50
ABS	Patient	% 85
ABS	ADV	% 7
ABS	MNR	% 4
ABS	TMP	% 2
ALA	Product	% 100
BALD	DIS	% 100
DENB	TMP	% 100
ERG	Agent	% 88
(...)		

5.17 taula: *aldata_change.01/alter.01* adieran postposizio-atzizki bakoitza zein rolekin eta zein portzentajetan agertu den.

Datu horiek izanda, prozedura automatikoari ekinez, % 50eko portzentajea gainditzen dutenak zuzenean gehitu dira agerpen berrien etiketetan. Prozedura hau 99 aditzetan egin da eta asko erraztu du eskuzko etiketatzea. Eskainitako laguntza hori zenbakitan neurtzeko ebaluazioa egin dugu.

Horretarako, alde batetik, portzentaje hauek aplikatuta ateratako etiketa guztiak hartu ditugu, eta horien gaineko eskuzko errebisioa egin ondorengo etiketekin konparatu ditugu. Etiketak berdinak diren kasuetan, hau da, aldatetarik egin ez den kasuetan, automatikoki zetorrena zuzena izan da: kasuen % 69,17. Gainontzekoetan (% 30,82), etiketatzailleak zuzenketaren bat egin behar izan du. Beraz, esan dezakegu proposatutako prozedura automatiko hau oso baliagarria izan zaigula eskuzko errebisioa egiterakoan denbora eta lana aurrezteko.

⁹⁶Portzentaje horiek delako postposizio-atzizkiaren agerpen-kopuru osoaren arabera koak dira.

⁹⁷Taula hau adibide gisa baino ez dugu jarri, eta bertan ez ditugu postposizio-atzizki, rol eta konbinazio bakoitzaren portzentaje guztiak jaso.

5.2.3.2 BVI baliatuz

Hala ere, kontuan izanda automatikoki sortutako konbinazio sintaktiko-semantiko horietan ez dela balentzia zehazten, datuak osoagoak izan daitezen eta eskuzko errebisioa are gehiago errazteko, eskuz egindako lexikoia ere erabiltzea erabaki dugu. BVIn dugun informazioa etiketetan txertatzeko, 99 aditzen azterketa egin dugu eta aditzak multzoka bildu ditugu. Horrela, BViko informazioa multzoka aplikatzeko aukera izan dugu, aditz errazenetatik hasi eta zailenetara. 7 multzo bereizi ditugu:

1. **Ingelesezko ordain bakarra / adiera bakarra:** 41 aditz (99 aditz-multzotik % 41,41) ditugu multzo honetan: *joan, ahaztu, konturatu, entzun, ikusi, bisitatu, erabili, gosaldtu, mintzatu, irakurri, abestu, baiezta-tu, bazkaldtu, edan, ehizatu, grabatu, jan, landatu, barkatu, erreparatu, besarkatu, existitu, ezkondu, flotatu, haserretu, isildu, hornitu, jolastu, kezkatu, konparatu, lastandu, mugitu, oroitu, otu, topatu, ukitu, uler-tu, zeharkatu, zintzilikatu, dedikatu, irakin*. Eskuz egindako lexikoian oinarrituta, etiketa automatikoki sortu dugu eta eskuz errebisatu da. Adibidez, 5.18 taulan, *joan* aditzak lexikoian duen sarrera ikusten da.

joan_go.01

Arg0	theme	gai ukitua (ABS)
Arg1	source	abiapuntuko kokapena (ABL)
Arg2	destination	helburuko kokapena (ALA/ABU)

5.18 taula: *joan_go.01* adiera BVIn.

Lexikoian argumentu eta rol bakoitza zein postposizio-atzizkirekin gauzatzen den agertzen da; hala, delako postposizio-atzizki hori dependenzietan bilatzen dugu, eta dagozkion balentzia eta rola esleitzen dizkiogu automatikoki.

2. **Adiera bakarra / ingelesezko ordain bat baino gehiago:** Multzo honetan 13 aditz (% 13,13) ditugu: *lortu, eskaini, erantzun, ikasi, bilakatu, eutsi, gainditu, jasan, kokatu, landu, moztu, adierazi, aritu*. Espero bezala, nahiz eta ordain bat baino gehiago iruditu egokia, argumentuak eta rolak ez dira aldatzen (rolak aldatzen direnean, ez ditugu automatikoki landu). Beraz, lehen eremuan posible diren ordain guztiak jarri eta gainontzean eskuz egindako lexikoian oinarrituta etiketak

sortu ditugu automatikoki; gero, eskuz errebisatu da. Adibidez, 5.19 taulan *ikasi* aditzaren BVIko sarrera ikusten da:

Arg0	agent	esperimentatzailea [+giz] (ERG)
Arg1	topic	jarduera (ABS/KONP/INE)
Arg2	source	- (ABL)

5.19 taula: *ikasi_learn.01/study.01* adiera BVIn.

3. **Adiera bakoitzak ordain bat:** 16 aditz (% 16,16) sailkatu ditugu multzo honetan: *jokatu, eman, hil, aldatu, erre, gertatu, eraman, iritsi, ekarri, eragin, esan, etorri, izan, egon, jaso, amaitu*. Gauzak zailtzen hasten dira hemen eta ezin da dena automatikoki egin. Aditzez aditz joan behar dugu, ezin baitira erregela orokorrak egin. Aditzak aztertzerakoan, adierak desberdintzeko balio diguten ezaugarriak bilatu ditugu eskuz egindako lexikoian, eta gero ezaugarri horren arabera adierari dagozkion argumentuak eta rolak esleitu dizkiogu etiketari automatikoki. Multzo honetako adibide bat *izan* aditza da. 5.20 taulan BVIko *izan* aditzaren sarrera ikus daiteke:

izan1_be.02	Arg0	theme	gaia (ABS)
	Arg1	location	kokapena (INE)
izan2_be.01	Arg0	topic	gaia (ABS/KONP)
	Arg1	attribute	ezaugarria (ABS)
izan3_have.03	Arg0	theme	edukitzailea (ERG)
	Arg1	theme	edukia (ABS)

5.20 taula: *izan* aditzaren sarrera BVIn.

Izan aditzaren kasuan, horrelako erregelak egin behar izan ditugu:

- INEa topatzen badugu, jarri *be.02* adieraren balentziak eta rolak.
- KONP topatzen badugu, jarri *be.01* adieraren balentziak eta rolak.
- ERGa topatzen badugu, jarri *have.03* adieraren balentziak eta rolak.

Gauza bera egin dugu multzo honetako 16 aditzekin.

Ondorengo 4 multzoetan ezin izan dugu BVIko informazioa automatikoki aplikatu; hauetan, aurreko prozesu automatikoan lortutako konbinazio sintaktiko-semantikoak baino ezin izan ditugu aplikatu, eta gai nontzekoa eskuz egin behar izan dugu.

4. **Adiera bakoitzak ordain bat baino gehiago eta ordainaren arabera rol desberdinak PB-VN-n:** Multzo honetan 10 aditz (% 10,10) ditugu: *eskatu*, *asmatu*, *sartu*, *aurkitu*, *jarri*, *agertu*, *bete*, *argitu*, *elkartu*, *pasatu*. 5.21 taulan, adibidez, *eskatu* aditzaren lehen adierentzat erabili diren ordainak ikus ditzakegu:

eskatu _ 1			
eskatu1_ask.02	Arg0 Arg1 Arg2	agent proposition patient	esperimentatzailea (ERG) gaia (ABS/KONP) - (DAT)
eskatu1_order.02	Arg0 Arg1 Arg2 Arg3	agent theme beneficiary source	esperimentatzailea (ERG) gaia (ABS/KONP) - (DES) - (DAT)
eskatu1_demand.01	Arg0 Arg1 Arg2	agent proposition patient	esperimentatzailea (ERG) gaia (ABS/KONP) - (DAT)
eskatu1_claim.01	Arg0 Arg1 Arg2	agent topic recipient	esperimentatzailea (ERG) gaia (ABS/KONP) - (DAT)

5.21 taula: *eskatu* aditzaren lehenengo adiera BVI-n.

5.21 taulan ikusten dugu lehen adieran ingelesezko 4 ordain (desberdin) ditugula, eta ordain bakoitzak rol desberdinak dituela PB-VN-n. Horrelako kasuetan ezin izan dugu automatikoki ezer egin. Etiketatzaileak erabaki behar du agerpen bakoitzarentzat ordain egokiena zein den, eta dagozkion balentzia eta rolak eskuz esleitu.

5. **Adiera bakoitzean ordain bat baino gehiago eta, gainera, adiera berriak:** 3 aditz (% 3,03) daude multzo honetan: *egin*, *jo*, *atera*. *Egin* aditzak, adibidez, adiera bakarra du EADBn, baina ingelesez hainbat aditz erabiltzen dira gure *eginen* adiera orokor hori adierazteko

(ikusi 5.22 taula), eta, gainera, corpusean EADBn landuta ez dauden bestelako adierak ere aurkitu ditugu.

egin_1

egin1_make.01/build.01/construct.01	Arg0 Arg1 Arg2 Arg3	agent product material beneficiary	iturria (ERG) gai sortua [-biz] (ABS) - (INS) - (DAT/DES)
egin1_do.02	Arg0 Arg1 Arg2	agent product instrument	iturria (ERG)) gai sortua [-biz] (ABS) - (INS/SOZ)
egin1_ask.02	Arg0 Arg1 Arg2	agent topic recipient	iturria (ERG) gai sortua [-biz] (ABS) - (DAT)
egin1_compose.02	Arg0 Arg1 Arg2	agent product beneficiary	iturria (ERG) gai sortua [-biz] (ABS) - (DAT/DES)
egin1_practice.01	Arg0 Arg1 Arg2	agent theme instrument	iturria (ERG) gai sortua [-biz] (ABS) - (INS/SOZ)
egin1_manufacture.01	Arg0 Arg1 Arg2	agent product beneficiary	iturria (ERG) gai sortua [-biz] (ABS) - (DAT/DES)
egin1_hold.04	Arg0 Arg1 Arg2	agent theme instrument	iturria (ERG) gai sortua [-biz] (ABS) - (INS/SOZ)
egin1_score.01	Arg0 Arg1 Arg2	agent product beneficiary	iturria (ERG) gai sortua [-biz] (ABS) - (DAT)

5.22 taula: *egin* aditzaren lehen adiera BVIn.

Adiera berrien artean (84) adibidekoa dugu, eta 5.23 taulan ikus daiteke adibide hori etiketatzeko sortu dugun adiera berria.

(84) *Ordu batzuk egiten dituzte hantxe*

Bigarren adiera berria (85) adibidean ikus dezakegu, eta adibide hau etiketatzeko lexikoian sortu dugun adiera berria 5.24 taulan ikus daiteke.

(85) *Pentsatzekoa da kide gehiago izateak are zailago egingo duela hori.*

egin_spend.02

Arg0	agent	esperimentatzailea (ERG)
Arg1	asset	denbora (ABS)
Arg2	goal	jarduera (INE)

5.23 taula: Corpora etiketatzean *egin* aditzarentzat aurkitutako 1. adiera berria.

egin_make.06

Arg0	agent	esperimentatzailea (ERG)
Arg1	theme	gaia (ABS)
Arg2	predicate	ezaugarria (ABS)

5.24 taula: Corpora etiketatzean *egin* aditzarentzat aurkitutako 2. adiera berria.

Kasuistika zabal hori ikusita, ezer gutxi egin daiteke automatikoki aditz honekin (*egin*), eta etiketatzaileak erabaki behar du agerpen bakoitzean zer den egokiena.

6. **Euskarazko adiera desberdinetan ingelesezko ordain bera:** 2 aditz (% 2,02) ditugu multzo honetan: *deitu* eta *isuri*. Aditz horiek nahiko korapilatsuak dira, eta zailtasunak izan ditugu adierak bereizteko eta ingelesezko ordain egokiak aukeratzeko; beraz, etiketatze-lana erabat eskuz egin behar izan dugu.
7. **Euskaraz bi adiera, ingelesez bakarra:** Hemen aditz bakarra (% 1,01) dugu: *hasi* (*begin.01/start.01*). Ingelesez adiera bakarra denez, bi adieretan VNko rol berberak jarri ditugu. Euskaraz, aldiz, EADBn bi adiera bereizten direnez, rol desberdinak jarri dizkiogu (5.25), eta eskuz erabaki behar izan dugu agerpen bakoitzari zein rol dagozkion.

hasi			
1- hasi_begin.01/start.01	Arg0	agent	kausa [+giz] (ERG)
	Arg1	theme	gai ukitua [-biz] (ABS)
	Arg2	instrument	- (INS/SOZ)
2- hasi_begin.01/start.01	Arg0	agent	esperimentatzailea (ABS)
	Arg1	theme	jarduera (INE/SOZ)
	Arg2	instrument	- (INS/SOZ)

5.25 taula: *hasi* aditzaren sarrera BVIn.

Ikusten den bezala, metodo erdi-automatikoak (konbinazio sintaktiko-
-semantikoak + lexikoia) lehen 3 multzoetan aplikatu ditugu, eta horre-
kin, EADBN landutako aditzen % 71 landu dugu erdi-automatikoki eta mo-
du zehatzean. Gainontzeko aditzetan, automatikoki ateratako konbinazio
sintaktiko-semantikoen portzentajeetan oinarrituta, % 50 gaintzen duten
“postposizio-atzizki+rol” pareak esleitu ditugu automatikoki, eta, gero, eskuz
errebisatu eta osatu da.

5.2.3.3 BVI aberasten

Automatikoki erauzitako konbinazio sintaktiko-semantikoetatik ateratako da-
tuekin **eskuz egindako lexikoia osatu** dugu, lexikoia eskuz egitean ezinez-
koa delako corpusean agertuko den kasuistika guztia aurreikustea. Etiketatz
erdi-automatikoari esker, 99 aditzen beste 12.015 agerpen etiketatu ditugu.
Guztira 99 aditz eta 143 adiera aztertu dira lehen fase honetan, eskuzko lana
eta bide erdi-automatikoak konbinatuz. Gainera, lexikoia postposizio-atzizki
berriekin aberastea lortu dugu konbinazio sintaktiko-semantiko horiek erabi-
lita. Abiapuntuko lexikoia egiteko, EADBN agertzen zen informazioa erabili
genuen oinarri gisa, eta han 9 postposizio-atzizki baino ez ziren aztertzen;
hori dela eta, puntu honetan, corpusetik erauzitako konbinazio sintaktiko-se-
mantikoek lexikoia aberasteko bidea eman digute, eta alde aurretik aurrei-
kusita ez genituen postposizio-atzizkiak gehitu ditugu lexikoian. Adibidez,
EADBN *jo* aditzaren “ezaugarri-esleitze” adieran, *ezaugarria* adierazten duen
argumentua absolutiboan (ABS) gauzatzen dela zehazten da, baina kombina-
zio sintaktiko-semantiko horiei esker, corpusean prolatiboan (PRO) ere ager-
tu dela ikusi dugu, eta, beraz, postposizio-atzizki hori gehitu dugu BViko
adieran (86):

- (86) *Gaddafik persona non grata-tzat jotzen zuen lider palestinarra.*
 jo_consider.01: arg2:Predicate:pro

jo4_consider.01

Arg0: agent, abiapuntuan [+giz] (ERG)

Arg1: theme, helburua (ABS)

Arg2: predicate, ezaugarria (ABS/**PRO**)

Osatze-lan hau oso aberasgarria izan da, aditzen sarrera lexikalak osatzea ekarri baitu, eta, horrekin batera, corpora automatikoki etiketatzeko aukera gehiago (estaldura handiagoa, alegia).

6. kapituluaren sakonki aztertuko dira etiketatze-lanaren urrats honetan atera ditugun hizkuntzen arteko aldeak, PB-VN eta EADBren arteko adostasunak eta desadostasunak, eta euskarazko aditzen inguruko fenomenoak.

Ondorengo urratsean, EADBn landuta ez dauden corpuseko gainontzeko aditzen etiketatze erdi-automatikoan nola jardun dugun azalduko dugu.

5.2.4 Levinen klaseen baliagarritasunaren azterketa

EADBn landuta ez dauden aditzen kasuan ez dugu inolako informaziorik aditzen adiera, argumentu eta rolei buruz. Gure helburua, kasu honetan, ingelesezko ordainik egokiena topatu, eta aditz horrek PB-VN-n duen informazioa euskarazko aditzari zuzenean esleitzea izango litzateke.

Hori egiteko, Levinen klaseetan oinarritu gara, hipotesia baita klase bereko aditzek jokaera sintaktiko-semanticoa bera izango dutela. Hala, [Levinen \(1993\)](#) aditz guztien euskarazko itzulpenetik (ikusi [4.1.3](#) atala) abiatu gara; EADBn ez dauden aditzak hartu ditugu, zein klasetakoak diren begiratu dugu, eta aurreko fasean landu ditugun 99 aditzetatik klase hori bera zeinek duen bilatu dugu. [5.26](#) taulan ikus daitezke atera zaigun zerrendak duen itxura. Lehen zutabean, landu gabeko aditz berria eta dagokion ingelesezko ordaina edo ordainak (parentesi artean) ikus daitezke. Bigarrenean, ingelesezko ordainaren Levin klasea agertzen da, eta, hirugarrenean, klase hori bera duten aurreko fasean landutako aditzak.

Lehen pausoan, klase bakarra duten aditzak begiratu ditugu; ikusi [5.26](#) taulan, esaterako, *jakin* eta *ezagutu* aditzak. Ondoren, aditz horiei dagokien landutako aditzen eremuan, aditz bakarra dutenak aukeratu ditugu; [5.26](#) taulan, *ezagutu* aditza (nabarmenduta agertzen dena), hain zuzen (landutakoen

eremuan *ikusi* baino ez duelako). Baldintza horiek betetzen dituzten 97 aditz ditugu, eta horiek landu dira fase honetan.

Aditz berria (ingeleseko ordaina)	Levin klasea	Landutako aditzak
irabazi (carry)	11.4	jaso, eraman
irabazi (earn, win)	13.5.1	eskatu, lortu, iritsi, topatu, eraman, jaso, ulertu, hartu, hautatu, ekarri, aurkitu
jakin (know)	29.5	adierazi, asmatu, onartu
utzi (admit, allow)	29.5	adierazi, asmatu, onartu
utzi (cease)	55.1	amaitu, hasi
utzi (leave)	13.4.1	eman, hornitu
utzi (leave)	13.5.1	eskatu, lortu, iritsi, topatu, eraman, jaso, ulertu, hartu, hautatu, ekarri, aurkitu
utzi (leave)	13.3	egokitu, atera, eman, eskaini, hautatu, onartu
utzi (relinquish)	13.2	aldatu, eman
ezagutu (recognize, spot)	30.2	ikusi
(...)		

5.26 taula: [Levinen \(1993\)](#) klase bera partekatzen duten etiketatu gabeko eta etiketatutako aditzen zerrenda.

Hasiera batean, landutako aditza (EADBn dagoena) adiera bakarrekoa den kasuak bakarrik hartu ditugu (27 aditz); hau da, aditz berria, landutako aditz baten klase berekoa denean, eta landutako aditz horrek adiera bakarra duenean. Horrelakoak dira *ezagutu/ikusi*, *idatzi/grabatu* edo *begiratu/entzun* bikoteak, esaterako. *Ezagutu*, *idatzi* eta *begiratu* EADBn ez dauden aditz berriak dira, eta *ikusi*, *grabatu* eta *entzun* aditzek Levin klase bera dute, etiketatuta daude eta adiera bakarra dute BVIn; beraz, automatikoki esleituko dizkiegu aditz berriei landutako aditzen balentzia eta rolak. Hala, 97 aditzetatik, modu honetan 27 aditz etiketatu ditugu automatikoki (% 28)⁹⁸.

⁹⁸Adjuntuak eskuz etiketatu ditugu, BVIn argumentuen informazioa baino ez baitigu ematen.

27 aditz horietatik 15 eskuz desanbiguatu behar dira, argumentu bat baino gehiago dagoelako kasu berean gauzatzen dena, eta, beraz, kasu hori agertzen denean, automatikoki bi etiketa sortzen zaizkigu, gero eskuz erabakitzeko aukera bietatik zein den zuzena.

Demagun *kritikatu* aditza landu behar dugula. Kontuan izanda *kritikatu* aditzaren klase bereko EADBko aditza *barkatu*⁹⁹ dela, 5.27 taulan ikusten den bezala definitu dugu bere sarrera:

kritikatu_criticize.01

Arg0	agent	abiapuntua [+giz] (ERG)
Arg1	theme	helburua (ABS/DAT/KONP)
Arg2	predicate	gaia (ABS/MOT)

5.27 taula: *kritikatu* adieraren sarrera BVIn.

5.27 taulan ikusten dugun bezala, Arg1 eta Arg2, biak, gauzatu daitezke absolutiboan. Orduan, dependentzietan ABSa agertzen denean, bi etiketa sortuko zaizkigu automatikoki (87), eta gero eskuz aukeratu beharko dugu zein den kasu bakoitzean zuzena.

- (87) *Ian Paisley apaizaren DUP alderdi unionista erradikala izan da itxiera kritikatu duen bakarra.*

ncobj (abs, kritikatu, itxiera, itxiera, obj)

arg_info (critize_01, kritikatu, itxiera, Arg1, **theme**, -, -)

arg_info (critize_01, kritikatu, itxiera, Arg2, **predicate**, -, -)

Gainontzeko 70 aditzak ezin izan ditugu modu berean landu, klase berekoa den landutako aditzak adiera bat baino gehiago duelako; beraz, ezin izan dugu automatikoki jakin zein adieraren argumentuak eta rolak esleitu behar zaizkion. Hori ebazteko, eskuz errebisatu ditugu 70 aditz horiek, eta landutako aditzaren zein adieraren pareko diren erabaki dugu, adiera horren argumentuak eta rolak esleitu ahal izateko. Adibidez, *egiaztatu/agertu* bikotean, biek dute 29.4 Levin klasea, baina *agertu* aditzak 3 adiera ditu eta

⁹⁹*Barkatu* aditzaren sarrera BVIn:

forgive.01

Arg0: agent, abiapuntua [+giz] (ERG)

Arg1: theme, helburua [+giz] (DAT/ABS)

Arg2: predicate, gaia [-biz] (ABS)

ezin dugu egiaztatu, bide automatikoak erabiliz, zein adieraren balentzia eta rolak esleitu behar dizkiogun aditzari.

Eskuzko lan honi esker, urrats honetan lantzekoak ziren 97 aditzen lexikoa osatu dugu eta BVIn gehitu ditugu sarrerak. Behin lexikoa izanda, automatikoki landu ezin izan diren 70 aditzen 30 agerpen etiketatu dira eskuz. Eskuz etiketatutakoa oinarri hartuta, eta 5.2.3 atalean azaldutako urratsak jarraituz, etiketa automatikoak sortu ditugu, eta, bukatzeko, dena eskuz errebisatu dugu. Horrela, fase honetan landu behar genituen 97 aditzen agerpenak etiketatuta geratu dira corpusean.

EADBN ez dauden 97 aditz berri horiek lantzeko jarraitu ditugun urratsak azaldu ondoren, bidean aurkitu ditugun hainbat zailtasun aipatuko ditugu, gure corpora etiketatzeko lanean Levin klaseek eskain dezaketen laguntza-
ren inguruan ondorioak ateratzeko asmoz. Jarraian, zailtasun horien artean nagusi diren laurak azalduko ditugu¹⁰⁰:

1. Batzuetan, nahiz eta aditzak klase berekoak izan, egitura sintaktiko-semantikoak ez datoz inondik ere bat; adibidez, *ireki/jaso* aditz-parean, biak dira 40.3.2 klasekoak, eta klase horri VNk 5 rol esleitzen dizkio: *Agent, Patient, Theme, Cause* eta *Recipient*. Ez dirudi *ireki* aditzak rol horiek guztiak dituenik; beraz, ezin dizkiogu automatikoki esleitu *jaso* aditzak dituen argumentuak eta rolak. Honelako hainbat adibide aurkitu ditugu, batez ere *ikusi* aditzaren klase bera duten aditzekin (*estali/ikusi; ohartu/ikusi...*).
2. 5.2.3 ataleko 99 aditzak lantzean (EADBN landu zirenak, hain zuzen), gure abiapuntua Levenen zerrenda itzulia izan zen. Hortik ateratzen genuen ingelesezko lehen hautazko ordaina, baina zenbait kasutan ordain hori ez zela egokia erabaki eta beste bat hautatzen genuen. Hala, zerrenda hori orain aztertzean, gerta liteke egindako EADBko aditzari ordaina aldatu izana, eta, beraz, hasierako zerrenda itzulian zuen ordainaren orde, beste klase bateko bat jarri izana. Ondorioz, aditz berriaren ingelesezko ordaina eta egindako aditzari jarritako ordaina ez dira orain klase berekoak, eta, hortaz, ezin ditugu automatikoki esleitu argumentuak eta rolak. Adibidez, *aldatu* aditzaren ingelesezko balio-kide gisa *alter* eta *change*, biak 26.6 eta 45.4 klaseetakoak, eta *shift*, 11.1 klasekoa, ditugu. *Aldatu* aditza etiketatu genuenean, ingelesezko

¹⁰⁰Automatikoki etiketatzea ezinezkoa izan denean, eskuz etiketatu da.

ordain gisa *alter* eta *change* jartzea erabaki genuen (eta, beraz, *shift-11.1* baztertu egin genuen). Orain, aditz berriak Levinen klaseekin aztertzen hasi garenean, (88) adibidean ikusten dena topatu dugu:

(88) garraiatu (*transport*) 11.1 aldatu

(88)-an ikusten den bezala, *garraiatu* aditz berriaren ingelesezko ordaina *transport* litzateke, eta *transport* hori *shift* aditzaren klase berekoa da, baina guk ez genuenez *shift* jarri *aldatu* aditzaren ordain gisa, gure *aldatu* aditzaren ordainak (*alter* eta *change*) eta *garraiatu* aditzaren ordain den *transport* hori ez dira klase berekoak, eta, beraz, ez dituzte argumentu eta rol berak. *Alter* eta *change* aditzek *Agent*, *Patient*, *Product* eta *Material* rolak dituzte VN-n, eta, *transportek*, aldiz, *Agent*, *Theme*, *Source* eta *Destination*. Hala, ezin dizkiogu *aldatu* aditzari esleitutako argumentuak eta rolak zuzenean *garraiatu* aditzari esleitu.

3. Beste kasu batzuetan, Levinen aditzen euskarazko itzulpena ez dirudi oso zuzena denik, eta, beraz, arazoak daude euskarazko aditza eta us-tezko bere ingelesezko ordaina parekatzeko; esaterako, *oldartu* aditza ingelesezko *victimize* aditzaren itzulpen gisa emanda zegoen zerrendan (89):

(89) oldartu (*victimize*) 33 barkatu

Baina ezin dizkiogu *oldartu* aditzari *barkatu* aditzaren argumentu eta rolak esleitu, ez dirudielako *victimize* hori baliokide egokia denik.

4. Zenbaitetan, landutako aditzak adiera bakarra dauka, baina horrek ez du esan nahi aditz berriak ere adiera bakarra izango duenik; beraz, nola etiketatu automatikoki adiera berri horiek eskuz landu gabe?

Kasuistika hori guztia aztertu ondoren, esan dezakegu euskararako dugun Levinen sailkapena oso mugatua dela prozedura automatikoak erabilita aditz eta agerpen berriak etiketatzeko. Ondorioz, proposatzen dugun behin betiko metodologian ez dugu Levinen klaseena baliatzea aurreikusi. Proposatutako metodologia horrek eskuzko etiketatzea eta prozedura automatikoak batzen ditu, eta hurrengo atalean azalduko dugu zabalago.

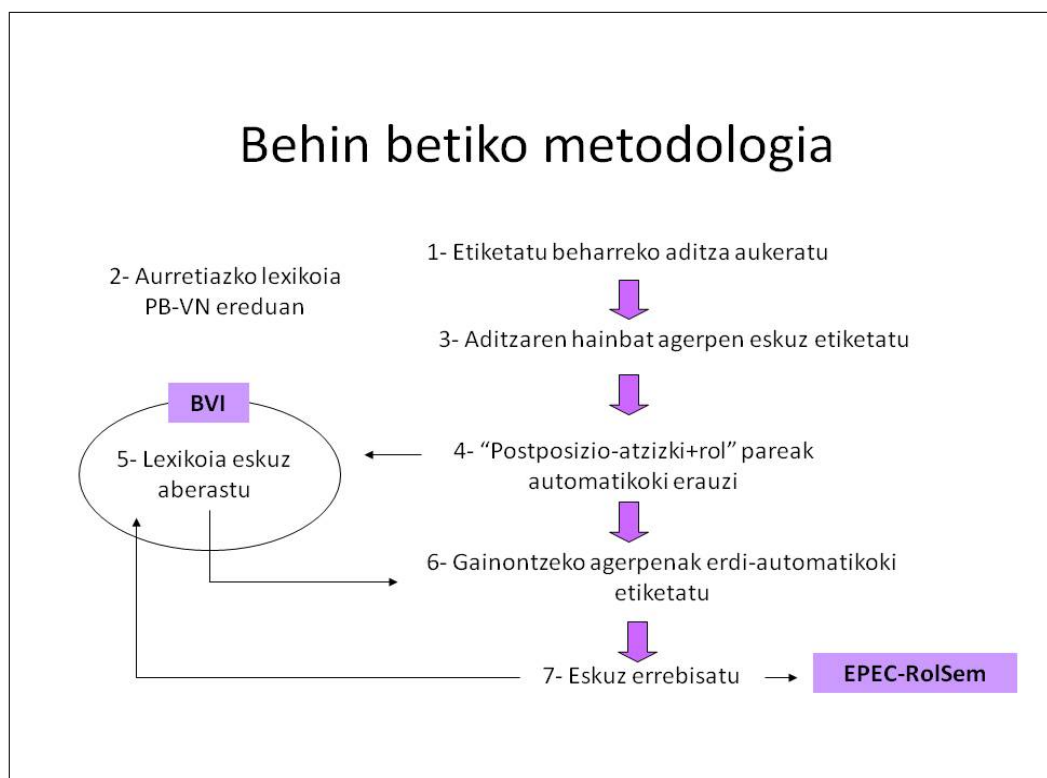
5.3 Behin betiko metodologia

Orain arte aztertutako bideak kontuan hartuta, aditz berriak etiketatze metodologia proposatu nahi dugu. Etiketatze-lanean zehar aztertutako aukerak ikusita, hainbat ondorio atera ditugu:

- BVI ondo definitzeak etiketatze-lanaren kalitatea bermatzen du.
- Prozedura automatikoen bidez lortutako konbinazio sintaktiko-semantikoek BVI aberasteko bidea eskaintzen digute.
- Prozedura automatikoak behar ditugu eskuzko etiketatze-lana errazteko.

Hainbat urrats definitu ditugu proposatutako azken metodologia honetan (ikusi [5.1](#) irudia):

1. Etiketatu beharreko aditza(k) aukeratu
2. PB-VN ereduari jarraituta aditzak lexikoian izango duen sarrera definitu
3. Aditzaren hainbat agerpen eskuz etiketatu (20-30 bitarte)
4. Eskuz etiketatutako laginetik konbinazio sintaktiko-semantikoak automatikoki atera
5. Konbinazio sintaktiko-semantiko horiekin hasierako lexikoia eskuz aberastu
6. Aditzaren gainontzeko agerpenak erdi-automatikoki etiketatu, lexikoia eta konbinazio sintaktiko-semantikoak baliatuta
7. Eskuz errebisatu



5.1 irudia: Behin betiko metodologian proposatutako urratsak.

Eskuzko etiketatze-lanean zehar esaldiei oharrak jartzeko beharra izan dugu, eta *AbarHitz* tresnak horretarako aukera ematen digu (ikusi 4.3 atala). Jarraian, erabili ditugun oharren zerrenda eta euren azalpena emango ditugu:

- **Aditz-esapidea:** Aztergai dugun aditza aditz-esapide baten parte dela uste dugunean, ohar hau jartzen diogu esaldiari. Horrelakoetan, Hitz Anitzeko Unitate Lexikala (HAUL) ziurra bada (Urizar, 2012), ez dugu etiketatzen eta oharra jartzen dugu (*bat etorri, lo egin...*); ez bada ziurra, aldiz, oharra jartzeaz gain, etiketatu egiten dugu (*zertara etorri, traizio egin, urratsak eman...*).
- **Aditz-perifraisi aspektuala:** Aztergai dugun aditzak, dagozkion adieren balioa adierazi beharrean, ondoan duen aditzaren aspektuari buruzko informazioa ematen duenean jarri dugu ohar hau. Adibidez: *ikusten joango gara / hasiz joan zen*. Horrelako esaldiak ez ditugu etiketatzen.

- **Aditz-perifraasi modala:** Aztergai dugun aditza modua adierazten duen perifraasi baten parte denean, ohar hau erabiltzen dugu. Adibidez: *Ez dakit hori botatzerik dagoen / arrazoi duelakoan nago.*
- **Argumenturik gabea:** Zenbaitetan, aditzak ez du argumentu-egiturarik aztertzen ari garen esaldian, eta, beraz, ezin diogu argumenturik edo rolik esleitu. Adibidez: *Joan den astean [...] / Ikusten dugunez [...].*
- **Balio berria:** Corpusean etengabe aurkitzen ditugu aurreikusi gabeko adierak, hau da, EADBN landuta ez dauden adierak. Horrelakoei jartzen diegu ohar hau (jo bedi 6.3.2 atalera adibideak ikusteko).
- **Ordain berria:** Levinen aditzen zerrenda euskaratua kontuan hartuta, tresnak euskarazko aditzentzat egoki izan daitezkeen PB-VNko ordainak eskaintzen dizkigu. Ordain horien artean egokirik topatu ez eta beste bat aukeratzen badugu, ohar hau jartzen diogu agerpenari. Esaterako, euskarazko *ahaztu* aditzarentzat tresnak *omit* aditza proposatzen digu¹⁰¹, baina, corpuseko adibideak ikusita, *forget* aditza egokiagoa dela iruditu zaigu. Beraz, *ahaztu* aditzaren agerpenak PBko *forget.01* aditzarekin etiketatuta ditugu, eta oharren eremuan “ordain berria” oharra jarri diegu.
- **d:** Dependentziak gaizki etiketatuta daudela uste dugunean, konpondu eta ohar hau jartzen dugu. Hala ere, EPEC-DEPen emandako irizpideak jarraitzen ez dituztenak bakarrik aldatuko ditugu (Aldezabal *et al.*, 2007a). Adibidez, *bi esibizio egitea onartu zuen Goenatzok* adibidean, *bi esibizio* izen sintagma *egin* aditzaren adizlagun gisa (ncmod) etiketatuta da, baina, argi dago akats bat izan dela, eta objektu zuzenaren etiketa behar duela (ncobj), beraz, dependentzia-etiketa aldatu eta “d” oharra jarri dugu.
- **Estilo zuzena:** Zenbait esaldi, estilo zuzena dela eta, ez dira dependentzietan etiketatuta, eta, beraz, esaldi horiek ez daude predikatu-mailan etiketatuta dugun EPEC-DEP corpusean. Hala ere, badira batzuk, eta, horrelakoetan, predikatu-mailako etiketatzea egitea erabaki dugu,

¹⁰¹Levinen (1993) lanean ez da *forget* aditza agertzen, eta, ondorioz, Levinen aditzen zerrenda euskaratuan ere ez. Horregatik, tresnak ez digu *forget* aditza proposatzen *ahaztu* aditzaren ordain gisa.

arg_info etiketaren hirugarren eremuan puntuazio-ikurra jarritz (gunearekin lotura egiten duen elementu gisa), eta oharren eremuan “estilo zuzena” oharra idatziz (90).

- (90) “*Bi buruzagiek, bai Barakek, bai Arafatek, badakite ahalegin handi bat egiteko prestatu behar dutela*”,-[w84]¹⁰² *esan zuen israeldarren zein palestinarren buruzagiak ohartaraziz.*

auxmod (-, esan, zuen)

ncmod (gen, buruzagiak, zein, palestinarren)

ncobj (abs, ohartaraziz, buruzagiak, buruzagiak, obj)

lot¹⁰³ (emen, zein, israeldarren)

lot (emen, zein, palestinarren)

xmod (mod, esan, ohartaraziz, ohartaraziz)

arg_info (say_01/tell_01, esan, ohartaraziz, argM, MNR, -, -)

ccomp_obj (konpl, esan, punt_koma-[w84], punt_koma-[w84])

arg_info (say_01/tell_01, esan, punt_koma-[w84], Topic, gaia, -konkr)

- **Erabilera desegokia:** Adibideren batean Euskaltzaindiaren arauetatik urruntzen den erabileraren bat topatzen badugu, etiketatzen saiatzen gara, baina oharretan “erabilera desegokia” jartzen dugu. Adibidez, *errulelara jokatu* agertu zaigu eta badakigu erabilera desegokia edo okerra dela, *erruletan jokatu* beharko lukeelako; beraz, *errulelara* horri dagokion rola jartzen diogu, baina oharra ere gehitzen diogu.

5.4 Laburbilduz

Kapitulu honetan, etiketatzeko metodologian emandako pausoak azaldu ditugu xehe. Etiketatze-lana hiru fase nagusitan banatu dugu:

1. Metodologiaren lehen hurbilpena

¹⁰²Puntuazio-ikurraren identifikadorea.

¹⁰³**lot:** loturazko elementuak.

2. Metodologiaren oinarrien finkatzea

3. Behin betiko metodologia

- Lehen fasean, hautatutako **eredua euskararako egokia dela egiaz-tatu** dugu. Horretarako, 3 hizkuntzalarik *esan*, *adierazi* eta *eskatu* aditzen 50 agerpen etiketatu dituzte, eta lan horretatik ateratako informazioarekin etiketatze-prozesuan ezinbestekoa izan den **eskuliburuaren lehen bertsioa osatu** dugu.
- Bigarren fasean, eta eredua egokia dela frogatu ondoren, bi helburu nagusi izan ditugu: i) corpus osoa etiketatzea, eta ii) behin betiko metodologiaren oinarriak finkatzea. Hasteko, EADBn landutako 100 aditzak eskuz etiketatzeari ekin diogu. Urrats honetan, 3 etiketatzaile aritu dira lanean, eta lortutako emaitza nagusiak izan dira eskuliburuaren bertsio osatua eta EADBn landuta dauden aditzen lexikoa (BVI) sortzea. Eskuzko lan honi esker, EADBn landuta dauden 100 aditzetatik EPECen agertzen diren 99 aditzen lexikoa sortu dugu eta corpuseko 7.244 agerpen etiketatu ditugu.

Eskuzko lan hori guztia egin bitartean, gure **metodologiaren egokitasuna ebaluatu** nahi izan dugu, hau da, orain arte jarraitutako metodoarekin etiketatzearen kalitatea bermatzen ote dugun frogatu nahi izan dugu. Horretarako, hiru aditz aukeratu ditugu (*adierazi*, *izan* eta *etorri*) eta bi etiketatzaile jarri ditugu aditz horien agerpenak etiketatzen (20 fitxategi aditz bakoitzeko: *adierazi*, 27 agerpen, *izan*, 74, eta *etorri*, 42; guztira, 143 agerpen). Ondoren, bi etiketatzaileen arteko adostasuna/desadostasuna neurtu dugu. Emaitza onak izan dira *adierazi* eta *izan* aditzetan, baina txarrak *etorri* aditzean. Desadostasunaren ebaluazio kualitatiboa egin ondoren, ondorio batzuk atera ditugu:

- Hainbat gai eskuliburuan zehaztu behar dira: adjuntuen etiketatzea, HAULEkin nola jokatu edota bokatioak nola etiketatu.
- Etiketatzen hasi aurretik aditzaren sarrera zehaztu behar da, hau da, aditzak lexikoian izango duen sarrera definitu behar da.

Bi ondorio nagusi horiek kontuan izanda, etiketatzaileek aditz bakoitzeko beste 20 fitxategi (*adierazi*, 27 agerpen; *izan*, 138, eta *etorri*, 42; guztira, 207) etiketatu dituzte, eta emaitzak nabarmen hobeak izan dira. Beraz, ebaluazio horretatik ateratako ondorioekin, eskuliburuaren

bigarren bertsioa osatu dugu (Aldezabal *et al.*, 2010b), eta, etiketatzen hasi aurretik egindako zehaztapen-lana baliatuta, EADBn landutako 99 aditzen lexikoia (BVI) sortu dugu.

Jarraian, 99 aditz horietan etiketatu gabe geratu diren agerpenen etiketatzeari ekin diogu. Lan hori guztia eskuz egiteak denbora eta lan gehiegi eskatuko lukeela ikusita, etiketatzeko **bide erdi-automatikoak** aztertu ditugu. Etiketatze erdi-automatikoak aurrera eramateko ezinbestekoa izan da aurretik egindako eskuzko lana, bai eta sortutako lexikoia ere.

Hasteko, eskuz etiketatutako laginetik aditz bakoitzaren konbinazio sintaktiko-semantiko guztiak erauzi ditugu automatikoki, hau da, etiketatutako agerpen bakoitzean agertutako “postposizio-atzizki+rol” pareak. Horietatik maizen agertzen direnak (corpusean % 50eko portzentajea gainditzen dutenak) automatikoki esleitu dizkiegu etiketatu gabeko agerpenei. Prozedura hori EADBn landutako 99 aditzetan egin da eta asko erraztu du eskuzko etiketatzea.

Automatikoki erauzitako konbinazio sintaktiko-semantiko horietan ez da balentzia zehazten (Arg0, Arg1...). Beraz, eskuzko errebisioa are gehiago errazteko, posible izan den kasuetan, BVIn bildutako informazioa automatikoki txertatu dugu etiketetan.

Prozedura automatiko horiei esker, BVI lexikoia corpusean agertutako informazioarekin osatzeko aukera izan dugu. Hasieran egindako eskuzko lana eta bide erdi-automatikoak konbinatuta, 99 aditz eta 143 adiera landu dira guztira.

Bigarren fase hau bukatzeko, **EADBn landuta ez dauden aditzak etiketatzeko bideak** aztertu ditugu. Horretarako, Levinen klaseetan oinarritu gara, lan horren hipotesi nagusiaren arabera, klase bereko aditzek jokaera sintaktiko-semantiko bera baitute. Hainbat azterketa egin ondoren, gure ondorioa izan da euskararako dugun Levinen sailkapena oso mugatua dela prozedura automatikoak erabilia alde aurretik eskuz landu gabeko aditz berriak etiketatzeko. Urrats honetan, beste 97 aditz etiketatu dira (EADBn landuta ez daudenak), klaseak baliatuta, automatikoki egin ahal izan dena eta eskuz egin dena bilduta. Metodologiaren urrats bakoitzeko etiketatze-kopuru zehatzetarako, ikusi 7.1 ataleko 7.1 taula.

- Hirugarren fasean, bigarren fasean egindako azterketa guztiak kontuan

hartuta, **aditz berriak etiketatzeko behin betiko metodologia proposatu** dugu. Metodologia horrek eskuzko lana eta bide erdi-automatikoak biltzen ditu eta honako urrats hauek ditu:

- Etiketatu beharreko aditza(k) aukeratu
- PB-VN ereduari jarraituta aditzak lexikoian izango duen sarrera definitu
- Aditzaren hainbat agerpen eskuz etiketatu
- Eskuz etiketatuko laginetik konbinazio sintaktiko-semanticoko automatikoki erauzi
- Konbinazio sintaktiko-semanticoko horiekin hasierako lexikoa eskuz aberastu.
- Aditzaren gainontzeko agerpenak erdi-automatikoki etiketatu, lexikoa eta konbinazio sintaktiko-semanticokoak baliatuta
- Eskuz errebisatu

EADBko 100 aditzak etiketatzean azaldutako kasuistika

Etiketatzeari azaldutako kasuistika guztia aurkezteko, hiru azpiataletan banatuko dugu atal hau¹⁰⁴:

1. EADB eta PB-VN ereduaren arteko adostasunak eta desadostasunak
2. Hizkuntzen arteko aldeak, hau da, ingelesarekiko aurkitu ditugun aldeak
3. Euskarazko aditzen inguruan azaldutako fenomenoak

6.1 EADB eta PB-VN ereduaren arteko adostasunak eta desadostasunak

Gure etiketatze-lana EADBn landutako aditzekin hasi dugu eta, beraz, etiketatzeko hautatutako PB-VN ereduko informazioa eta EADBn jasotzen dena bateratu egin behar izan ditugu. Bateratze-lan horretan, zenbaitetan ez da arazorik egon bi ereduak bat etorri direlako, baina beste zenbaitetan, ereduak ez dira bat etorri, eta desadostasun horiek gainditzeko hainbat erabaki hartu behar izan ditugu.

¹⁰⁴Kapitulua honetan aztertzen ditugun gaiak [Estarrona et al.en \(2014a\)](#) lanean jasota daude.

Hasteko, bi ereduak bat etorri diren kasurik arruntena aurkeztuko dugu eta, jarraian, ereduak bat etorri ez diren kasuetan nola jokatu dugun azalduko dugu.

6.1.1 EADBko eta PB-VNko informazioa bat etorri da

Lehenik eta behin kasurik arruntena azalduko dugu: bi ereduak bat egiten dute landutako aditzaren egitura sintaktiko-semanticoa. Horrelako kasuetan, zuzenean aplikatu ahal izan dugu PB-VN ereduak.

Horrelako kasua da *aldatu* aditza. Aditz honek bi adiera edo predikatu orokor ditu euskaraz, EADBn aztertutakoaren arabera betiere:

1. Kokapen- nahiz egoera-aldaketa
2. Ordezkatze-jarduera

Lehen adieraren ordain egokiak *alter.01* eta *change.01* dira. Biak dira klase berekoak eta, beraz, argumentu eta rol berdinak dituzte PB-VN-n. 6.1 taulan ikusten den bezala, lehen adiera horrentzat EADBn ematen den informazioa PB-VN-n ematen denarekin bat dator, bai argumentu-kopuruan, bai eta rol semanticotik ere:

PB-VN: <i>alter.01/change.01</i>	EADB: <i>aldatu_1</i>
Arg0: causer of transformation (agent)	kausa (ERG)
Arg1: thing changing (patient)	gai ukitua (ABS)
Arg2: end state (product)	helburua/kokapena/egoera (ALA/INS)
Arg3: start state (material)	abiapuntua (ABL)

6.1 taula: PB-VNk eta EADBk *alter.01/change.01* eta *aldatu_1* adierentzat ematen duten informazioa.

PB-VNko *Agent* EADBko *kausa* rolaekin parekatzen da, eta berdin *Patient* eta *gai ukitua* rola. Hirugarren eta laugarren argumentuak *end state* eta *start state* direla esaten da PBn, eta *Product*¹⁰⁵ eta *Material* rola esleitzen zaizkie VN-n. Informazio hori bat dator EADBko *helburua* eta *abiapuntua* rola adierazten dutenarekin, *helburua* aldaketaren emaitza baita, eta *abiapuntua* aldaketaren aurretik zegoena.

¹⁰⁵VNren 3.2 bertsioan, Arg2 argumentuari *Result* rola esleitzen diote.

Bigarren adieran ere bat datoz bi ereduak. Adiera hori itzultzeko, PB-VNko *replace.01* aditza aukeratu dugu; 6.2 taulan ikus dezakegu PB-VN-n eta EADBn bakoitzak duen informazioa:

PB-VN: <i>replace.01</i>	EADB: <i>aldatu_2</i>
Arg0: replacer (agent)	esperimentatzailea (ERG)
Arg1: old thing (theme1) ¹⁰⁶	gaia (ABS)
Arg2: new thing (theme2)	gaia (INS)

6.2 taula: PB-VNk eta EADBk *replace.01* eta *aldatu_2* adierentzat ematen duten informazioa.

Bietan ikusten ditugu 3 argumentu, eta rola ere bat datoz. Lehen argumentuan, PB-VN-n *Agent* ageri da, eta EADBn horren pareko den *esperimentatzailea*. Ondorengo bi argumentuek *Theme* rola jasotzen dute PB-VN-n, eta EADBn ere *gaia*.

Hori guztia ikusita, PB-VNko argumentuak eta rola zuzenean esleituko dizkiegu *aldatu* aditzaren lehen eta bigarren adierei.

6.1.2 EADBko eta PB-VNko informazioa ez da bat etorri

Aurreko kasuan ikusi dugun bezala, ereduak bat datozenean ez dago inolako arazorik, eta zuzenean esleitzen diogu PB-VNko informazioa euskarazko aditzari; baina gauzak ez dira beti hain errazak, zenbait kasutan ereduak ez baitute bat egiten.

PB-VNk aditzaren mende agertzen den elementu bat argumentu den ala ez erabakitzeke, maiztasunari begiratzen dio; hau da, delako elementua oso maiz agertzen bada aditz horrekin, argumentu gisa etiketatzen dute. Normalean, PB-VNk EADBk baino argumentu gehiago esleitzen dizkie aditz-adierari, eta guk, kasu horietan, gure irizpide orokorrari jarraiki, PB-VNk esleitzen dituen argumentuak esleitu dizkiogu euskarazko aditz-adierari ere (Agirre *et al.*, 2006a), betiere argumentu horiek euskarazko adierari kalterik egiten ez badiote, hau da, EADBko postposizio-atzizki onargarrien¹⁰⁷ artean

¹⁰⁶VNren 3.2 bertsioan *Theme1* eta *Theme2* aldatu egin dituzte, eta orain *Theme* eta *Co-theme* agertzen dira.

¹⁰⁷EADBn aditz-adiera bakoitzean, ezinbestekoak diren postposizio-atzizkiez gain (argumentuak), postposizio-atzizki onartuak eta ez-onartuak zeintzuk diren zehazten da.

badaude. 6.3 taulan ikusten den *lortu* aditzaren bidez azalduko dugu hobeto egoera hau:

PB-VN: <i>obtain.01</i>	EADB: <i>lortu_1</i>
Arg0: receiver (agent)	esperimentatzailea (ERG)
Arg1: thing gotten (theme)	gaia (ABS)
Arg2: received from (source)	-

6.3 taula: PB-VNk eta EADBk *obtain.01* eta *lortu_1* adierentzat ematen duten informazioa.

6.3 taulan ikusten dugunez, PB-VNk hiru argumentu esleitzen dizkio eta EADBk bi; baina *Source* (ABL) hori kasu onargarrien artean dago euskarazko adieran:

“Kasu azpikategorizatuak: ERG, ABS
 Kasu onartugabeak: ELA_KONP, ALA
 Kasu onartuak: INE, INS, SOZ, ABL, DAT”

(Aldezabal, 2004: C eranskina, 30)

Gainera, aditz horrek adiera edo predikatu orokor bakarra du; hau da, ez dago beste adiera batetik bereizi beharrik, eta, beraz, ez dago arazorik argumentu gisa etiketatzeko.

Zenbait kasutan ikusi dugu PB-VN-n agertzen den argumentu bat, nahiz eta maiztasun gutxikoa izan, euskarazko adierak desberdintzeko lagungarri izan daitekeela, eta horrelakoetan ere PB-VNri jarraitzea eta euskarazko adieran argumentu hori gehitzea erabaki dugu. Ikus dezagun, esaterako, *esan* aditzaren kasua. EADBN bi adiera ditu:

1. Adierazpenezko jarduera bat
 - esperimentatzailea [+giz] (ERG); gaia [-konkr](ABS/ELA_KONP)
2. Ezaugarri-esleitze bat
 - abiapuntua [+giz] (ERG); helburua (DAT); ezaugarria (ABS)

6.4 taulan ikusten den bezala, PB-VNk *esanen* lehen adierari dagokion *say.01 rolsetean* beste bi argumentu ematen ditu, *hearer* eta *attributive*, hain zuzen ere:

say.01

Arg0	sayer (agent)
Arg1	utterance (topic)
Arg2	hearer (recipient)
Arg3	attributive

6.4 taula: *say.01* adiera PB-VN-n.

Azken argumentua (*Arg3_attributive*) oso maiztasun gutxiko argumentua da (VN-n ez diote rolik esleitzen, hau da, ez dute argumentutzat jotzen), baina *esan* aditzaren kasuan adiera bereizlea da, hau da, *esan* aditzaren bigarren adieran, ezinezkoa da *zertaz* edo *-ri buruz* moduko elementu bat agertzea aditzaren mende; beraz, argumentu hori adieren artean bereizteko baliagarria da, eta gehitu egin dugu euskarazko *esan* aditzaren lehen adieran.

Asko izan dira horren antzeko kasuak. Gure irizpide orokorra PB-VN ereduari jarraitzea denez, EADBN argumentu ez diren horiek euskarazko adierari kalterik egiten ez diotenean edo adiera bereizle direnean, argumentu gisa etiketatuta ditugu¹⁰⁸.

6.1.3 Path rola gehitu

Ereduaren arteko aldeekin jarraituz, EADBN bai, baina PB-VN-n agertzen ez den *Path* (bidea) rol semantikoa da beste alde bat¹⁰⁹.

Corpusa etiketatzen joan ahala, beharrezkoa iruditu zaigu rol semantiko hori gehitzea. Hualde eta Ortiz de Urbinak (2003) ere *pasatu* aditzerako aipatzen dute rol hori *ibaia pasatu dut* adibidea ematen dutenean¹¹⁰. (91)-an *pasatu* aditzaren bi adibide ikus daitezke (biak ere EPEC corpusekoak), bata ablatiboan eta bestea inesiboan:

¹⁰⁸Lan honetan hasi ginenetik, aldaketak ikusi ditugu VN-n. Garai batean, VN norma-lean bat etortzen zen argumentu-kopuruan EADBrekin, hau da, PBk baino argumentu gutxiago esleitzen zituen, baina azkenaldian, VNk PBrekin berdintzeko bidea egin du. Garai batean, PBren eta VNren artean desadostasuna zegoenean, guk PBren alde egin genuen, eta badirudi VNk ere erabaki berbera hartu duela bere azken bertsioetan.

¹⁰⁹Esan beharra daukagu VNren 3.2 bertsioan *Path* rolaren pareko izan daitekeen *Trajectory* rola gehitu dutela; esaterako, 51.1 klasean (Levinen (1993) arabera *verbs of inherently directed motion*) sartzen dituzten *come.01* eta *go.01* aditz-adieretan. Gure lana hasi genuenean ez zegoen horrelakorik, eta horregatik sortu genuen *Path* rola.

¹¹⁰“Only the path-structure *pasatu* takes a transitive auxiliary” (Hualde eta Ortiz de Urbina, 2003: 373).

- (91)
- [...] *haietako norbaiten ama garrasika hasi zitzaion, errieta botatzen, ondotik nor pasatzen zen jakin gabe.*
 - *Labezomorro bat pasa da bere aurrean, patxadaz, inguruarekiko axolagabe.*

Argi dago horrelakoetan ablatiboa eta inesiboa ez direla “abiapuntuko kopapena”, baizik eta aditzak adierazten duen mugimendua egiterakoan egiten den “bidea”.

6.2 Ingelesarekiko aldeak

Nabaria denez, hizkuntza baterako sortutako eredia beste hizkuntza batera aplikatzea ez da beti zuzenean eta arazorik gabe egiten den bidea. Ingelesa eta euskara oso hizkuntza desberdinak dira eta, beraz, pentsatzekoa da ingeleserako sortutako eredia euskarara aplikatzean, hainbat arazo edo zailtasun topatuko ditugula. Atal honetan, bi hizkuntzen artean aurkitutako aldeak azalduko ditugu.

6.2.1 Euskaraz posible ez den argumentu bat agertzen da PB-VN-n

Euskaraz posible ez diren argumentuak ez ditugu kontuan hartu, eta, beraz, ez ditugu etiketatu (eta, jakina, ez eta BVIn definitu ere). Horien artean, ingelesezko mugimenduzko zenbait aditzetan agertzen den *Extent* argumentua daukagu. Ikus dezagun *etorri* aditzaren adibidea. 6.5 taulan ikus dezakegu *come.01* aditz-adierak PB-VN eredian duen informazioa:

come.01

Arg1	entity in motion / “comer” (theme)
Arg2	extent – rare
Arg3	start point (source)
Arg4	end point (destination)

6.5 taula: *come.01* adiera PB-VN-n.

PBn bertan esaten dute, 6.5 taulan ikusten den bezala, Arg2_ *Extent* hori ez dela ohikoa (*rare* marka du ingelesez) eta, gainera, ez dute horren adibide bat bera ere ematen. VNk ez dio rolik esleitzen Arg2 argumentuari

VNko *escape-51.1* klasean (Levinen (1993) 51.1 *Verbs of Inherently Directed Motion*), ez dutelako *Extent* rola aurreikusten.

EADBN ere ez dugu argumentu hori *etorri* aditzaren sarreran (92):

- (92) *etorri_1*: gai ukitua_ABS; abiapuntuko kokapena/bidea_ABL; helburuko kokapena_ALA

Izan ere, euskaraz ez da posible zenbateko distantzia *etorri* den esatea (93).

- (93) **Mikel 200 metro etorri da handik hona.*

Beraz, Arg2_*Extent* hori ez dugu kontuan hartu BVIko *etorri* aditzaren lehen adiera definitzerakoan, eta ondoren datozen argumentuak aurreratu egin ditugu, 6.6 taulan ikus daitekeen bezala:

etorri_1		
Arg0 ¹¹¹	theme	gai ukitua (ABS)
Arg1	source	abiapuntuko kokapena (ABL)
Arg2	destination	helburuko kokapena (ALA)

6.6 taula: *etorri_1* adiera BVI lexikoian.

To go aditzean ere ikus dezakegu *Extent* rola, baina kasu honetan badira adibideak, nahiz eta ez diren *to go* aditzaren benetako adibideak. Ikus dezagun, 6.7 taulan, *go.01* aditz-adiera PB-VN-n:

go.01	
Arg1	entity in motion / “goer” (theme)
Arg2	extent
Arg3	start point (source)
Arg4	end point (destination)

6.7 taula: *go.01* adiera PB-VN-n.

6.7 taulan ikusten dugunez, PB-VNk 4 argumentu esleitzen dizkio *go.01* aditz-adierari, baina Arg2_*Extent* hori *go up* eta *go down* aditzek dituzten adibideetan baino ez da agertzen, (94) adibidean ikusten dugun bezala:

¹¹¹Ikusi 3.1 atala, Arg0/Arg1 aukeratzeko irizpidearen azalpenerako.

(94) *Imports have gone down 33%*

Arg1: Imports

Argm-dir: down

Arg2: 33%

Woolworth went up 1 3/4 to 59 1/2

Arg1: Woolworth

Rel: went

Argm-dir: up

Arg2: 3/4

Arg4: to 59 1/2

Ikusi dugu, beraz, ingelesez ere *Extent* argumentu hori ez dela *to go* aditz arruntarentzat erabiltzen, baina PBn *go up* eta *go down* aditzek ez dute aparteko sarrerarik *phrasal verb* gisa, eta *go.01* aditz-adieraren azpian sartzen dituzte (ikusi 6.2.2 atala); horregatik agertzen da *go.01* aditz-adieran Arg2_Extent argumentua.

Euskaraz *go up* eta *go down* horiek *igo* eta *jaitsi* aditzekin parekatuko genituzke, eta aditz horietan posible da zenbateko distantzia igo edo jaitsi den esatea, baina ez *joan* aditzean. Beraz, *etorri* aditzaren kasuan egin dugun bezala, ez dugu Arg2_Extent hori kontuan hartuko BVIko sarrera definitzean (6.8 taula):

joan_1		
Arg0	theme	gai ukitua (ABS)
Arg1	source	abiapuntuko kokapena (ABL)
Arg2	destination	helburuko kokapena (ALA/ABU)

6.8 taula: *joan* aditzaren sarrera BVIin.

6.2.2 Mugimenduzko aditzak: aditz-hizkuntzak vs satellite-hizkuntzak

Motion verbs delako aditzak bereziki aztertu izan dira hainbat hizkuntzatan. Mugimenduaren kontzeptua hizkuntza guztietan agertzen da, baina hizkuntza bakoitzak mugimendua adierazteko modu desberdinak erabiltzen

ditu, hots, lexikalizazio-eredu desberdinak¹¹². Talmy-ren tipologiari jarraituta (Talmy, 1985), euskara *verb framed* hizkuntza (aditz-hizkuntza) izango litzateke, gaztelania edo turkiera bezala, eta ingelesa, aldiz, *satellite-framed* hizkuntza (satelite-hizkuntza), alemana bezala. Satelite-hizkuntzek aditzetik kanpo uzten dute mugimenduaren norabidea adierazten duen elementua (*out, in, up, etab.*), eta aditz-hizkuntzek, ostera, aditzean bertan adierazten dute norabide hori. Horren adibide dira: *igo (go up)/jaitsi (go down)* edota *atera (take out/come out/go out)/sartu (come in/go in)*.

Beraz, euskarak eta ingelesak lexikalizazio-eredu desberdinak erabiltzen dituzte, eta horrek arazoak sortu dizkigu mugimenduzko aditzetan euskarazko aditzari ingelesezko PB-VN ordaina topatzerakoan. Ingelesezko *phrasal verb* horiek ez daude sistematikoki landuta PB-VN-n; horren ondorioz, askotan zaila izan da ordain egokia topatzea. Adibidez, *atera* aditzaren “kokapen-aldaketa” adierarentzat ingelesezko ordain egokia topatzea zaila izan da: ingelesez *take out, come out* edo *go out* erabiltzen dute predikatu orokor hori adierazteko. PB-VN-n badago *take out (take.11)*, baina ez *atera* aditzaren adiera horretan, baizik eta *lortu* adieran (*obtain*). Gauza bera gertatzen da *come out (come.09)* aditzarekin, *appear* esanahia adierazteko erabiltzen dute. PBn *go out (go.17)* ere agertzen da, baina 2 argumentu baino ez ditu, eta gure “kokapen-aldaketak” 3 edo 4 argumentu eskatzen ditu EADBren arabera:

- *atera*-DA: gai ukitua (ABS); abiapuntuko kokapena/bidea (ABL); helburuko kokapena (ALA)
- *atera*-DU: kausa (ERG); gai ukitua (ABS); abiapuntuko kokapena/bidea (ABL); helburuko kokapena (ALA)

Azkenean, *come out.01, take out.01* ordainak sortu ditugu gure *atera* aditzaren “kokapen-aldaketa” adierarentzat. Alegia, ingelesez esanahi hori nola adierazten den kontuan hartu dugu, eta gure *atera* aditzaren “kokapen-aldaketa” adieraren argumentuak eta rolak jarri dizkiogu. Ikusi 6.9 taulan *atera_1* aditz-adierak BVIn duen informazioa.

¹¹²Gai honen inguruan ikusi Jaka-ren (2009) lana, mugimenduzko ekintzek ingelesetik euskararako itzulpenetan sortzen dituzten arazoei eta horien aurrean itzultzaileek bilatutako irtenbideei buruz.

atera_1 (come_out.01/extract.01/take_out.01/take.01)

Arg0	agent	kausa (ERG)
Arg1	theme	gaia (ABS)
Arg2	source	abiapuntuko kokapena (ABL/DAT)
Arg3	destination	helburuko kokapena (ALA)

6.9 taula: *atera_1* adieraren sarrera BVI lexikoian.

Bestalde, *go up*, *go down*, *come out* eta antzekoak, esan dugun bezala, *go* eta *come* aditzen sarreretan daude sartuta (ikusi 6.2.1 atala). Adibidez, *go.01* aditz-adieraren azpian, lehen aipatutako (95) adibidea aurkitzen dugu PBn.

(95) *Imports have gone down 33%*

Hori dela eta, *igo* aditzaren ordaina aukeratzean, ezin izan dugu *go up* aukeratu, PB-VN-n ez baitago *phrasal verb* hori aparte landuta. Horren orde, *rise*, *raise*, *ascend* eta antzeko ordainak erabili behar izan ditugu. Gauza bera gertatu zaigu antzeko beste aditz batzuekin.

6.2.3 Euskaraz aditz bakar batekin esaten dena, ingelesez ordain desberdinekin itzultzen da

Bi egoera daude horrelako kasuetan¹¹³, eta egoera bakoitzak ondorio desberdinak ditu argumentuei eta rolei dagokienez:

1. Euskarazko aditz-adiera bat adierazteko ingelesezko ordain bat baino gehiago erabil daitezke, baina aditz-adiera horien argumentuak eta rolak ez dira aldatzen klase bereko aditzak direlako. Horren adibidea da *amaitu* aditzaren lehen balio semantikoarekin (egoera-aldaketa) gertatu zaiguna. Ingelesez *amaitu* berdinean esan daiteke *to finish* eta *to end*, eta gure ustez biak dira egokiak euskarazko *amaitu* aditzaren adiera hau adierazteko. PB-VN-n begiratzen badugu, *finish.01* eta *end.01* VN klase berekoak (55.4-1 klasea) direla ikusten dugu, eta, beraz, biek

¹¹³Jakin badakigu kontrako fenomeno ere gertatzen dela, hau da, ingelesez aditz bakar batekin esaten dena, euskaraz aditz desberdinekin adieraztea. Adibidez, ingelesezko *to break* aditza euskaraz *apurtu*, *puskatu* edota *hautsi* izan daiteke. Fenomeno horretan ez dugu sakontzen, izatez, horrek ez digulako arazorik sortzen.

dituzte argumentu eta rol berberak: *Agent, Theme, Instrument*. Gure *amaitu* aditzaren “egoera-aldaketa” adierak bi argumentu ditu: *kausa* eta *gai ukitua*, eta ez dago inolako arazorik *amaituren* balio semantiko hori PB-VNko *finish.01* eta *end.01* aditz-adieren argumentu eta rolekin etiketatzeko.

2. Euskaraz aditz bakar batekin esaten dena, ingelesez bat baino gehiagorekin esaten da, testuinguruaren arabera. Horren adibide garbia da *egin* aditza (bai eta *hautsi, eskatu* edo *pasatu* ere). Euskaraz hainbat gauza esateko erabiltzen da *egin* aditza, baina, esanahia gorabehera, aditz-adiera bat eta bakarra da beti; ingelesez, ordea, esanahi jakin bakoitzarentzat aditz desberdina erabiltzen da. Esate baterako, galderak euskaraz *egin* egiten dira, baina ingelesez ez dira egiten, *galdetu* (*to ask*) egiten dira. Horregatik, *galdera bat egin* adibidean, ingelesez *to ask* aditza erabili beharrean gaude, eta honek desadostasunak sor ditzake EADB eta PB-VNren artean (argumentu-kopuruari eta rolei dagokienez). Gainera, ingelesez *egin* aditza *to make* (*prestakuntza txiki bat egin du*) eta *to do* (*aurreko jardueran jasotako informazioa erabiliz, egizu honako hau*) aditzekin itzul daiteke testuinguruaren arabera. Bestelako aditzak ere erabili ditugu kasuan kasuko ingelesezko itzulpena emateko (96)¹¹⁴.

- (96)
- a. Kanta asko *egin* zituen: *compose.02*
 - b. Baina biek ere joko alaiegia *egiten* zuten ACB baterako: *practice.01*
 - c. Baserrietan sagardoa etxeko sagarrekin *egiten* zen: *manufacture.01*
 - d. Ateen bidez *eginiko* erabilpen anitzeko zirkuitu konbinazionalak: *build.01/construct.01*
 - e. Bilera bat *egin*: *hold.04*
 - f. Boticak 27 puntu *egin* zituen: *score.01*

¹¹⁴Badakigu adibide horietako batzuetan bestelako aditzak erabil zitezkeela; esaterako *kanta asko osatu zituen* edota *ateen bidez eraikitako erabilpen anitzeko zirkuitu konbinazionalak*, baina, izatez, corpusean *egin* agertu da, eta horri erantzun behar diogu guk. Beraz, hizkuntzaren berezko ezaugarria izan ala hizkuntza “pobretuari” dagokion zerbait izan, corpusean agertzen denari irtenbidea bilatu beharrean gaude.

Horrek arazo bat sortu digu *egin* aditzaren rolak etiketatzean. Izan ere, *eginen* sortzezko balio bakar horrek bi argumentu ditu: *iturria* eta *gai sortua*; baina, aldiz, ingelesezko ordainek hiru argumentu edo gehiago dituzte, eta, gainera, rolak aldatu egiten dira batetik bestera (97).

- (97) a. *compose.02*: agent, product, beneficiary
 b. *practice.01*: agent, theme, instrument
 c. *manufacture.01*: agent, product, beneficiary
 d. *ask.02*: agent, topic, recipient
 e. *build.01*: agent, product, material, beneficiary

PB-VN ereduari jarraituz etiketatzen ari garenez, gure erabakia izan da ordain bakoitzari dagozkion argumentuak eta rolak etiketatzea, baina jakin badakigu, euskara hutsezko lexikoa egiten dugunean, *egin* aditzak ezingo dituela rol desberdinak jaso agerpen bakoitzean, baldin eta ez baditugu adiera desberdinak definitzen. Hau da, dela *galdera bat egin*, dela *maiteari poesia bat egin*, *egin* aditzak beti izan beharko ditu argumentu eta rol berberak. Bietan erabakitzen badugu “sortze-prozesu” bat dagoela, rol semantikoak berdinak izango dira batean zein bestean: *iturria* eta *gaia* (eta *helburua* aukeran, alternantzia modura). Hala ere, Arg1-ean ageri diren *Product*, *Theme* edo *Topic* rolak oso antzeko objektuak dira, eta, hortaz, *Themerekin* (gaia) adieraz daitezke. Horretaz gain, hirugarren argumentua beti da *Instrument*, hots, *egin* beharrekoa “zerekin” edo “zein tresnarekin” egiten den adierazten duena. Orobat, *Beneficiary/Recipient* “norentzat” edo “nori” *egin* zaion *egin* beharrekoa (*helburua*). EADBn ez dira azken bi horiek argumentutzat jo, eta iruditzen zaigu gehiago direla adjuntuak, baina, bestalde, *egin* aditzak adiera bakarra izanda, ez dago arazorik argumentu gisa etiketatzeko.

Atal honetan, *hautsi* aditzean gertatzen den salbuespen berezia ere aipatu nahiko genuke. Aditz honek balio semantiko bakarra du EADBn, entitate baten egoera-aldaketa, hain zuzen ere. Alternantzia kausatibo/inkoatiboa du, eta bi argumentu azpikategorizatzen ditu: i) ergatiboan gauzatzen den *kausa* eta ii) absolutiboan gauzatzen den *gai ukitua*. PBko ordaina *break.01* da, eta horrek ere *Agent* eta *Patient* rolak betetzen dituzten argumentuak hartzen ditu (*Instrument* argumentua ere agertzen da). Baina bada ingelesez testuinguru jakin bat, hausten dena gorputz-atal bat denean, hain zuzen ere, non *to break* erabili beharrean, *to fracture* erabiltzen den, eta *fracture.01* aditz-adieraren rolak *Experiencer* eta *Patient* dira. EADBn ere salbuespen gisa aztertzen da testuinguru jakin hori:

“Hala ere, jabetzazko harreman hau ez da bakarrik agertzen DATen presentzian. Izan ere, DU-2 *bssan*, ABS entitatea gorputz-atala denean, ERG entitatearekikoa da jabetza hori (*besoa hautsi du*). [...]. Kasu hauetan, orduan, *gai ukitua* gorputz-atala denez, ERG ere -honen jabea izaki- nolabait *gai ukitu* bihurtzen da, eta *kausa* balioa galtzen du. [...]. Beraz, erabilera hauek DU-2 *bssaren* barruan salbuespen lirasteke *kausa* osagaiari buruz orokorrean esan dugunarekiko” (Aldezabal, 2004: 132).

EADBn ere salbuespena dela onartzen denez, BVI lexikoian *fracture.01* etiketatzen dugunean, euskarazko rolak ere aldatu egingo ditugu, eta, *kausaren* ordez, *gai ukitua* jarriko diogu ERGari ere.

6.2.4 Kausatibo/inkoatibo egiturak lexikalizatzeko bide desberdinak

Euskaraz, DU-DA erabilera duten aditzekin zalantza izan ohi da, egitura horiek adiera beraren alternantzia gisa edo adiera desberdineko egitura gisa landu behar ote diren.

Ikus dezagun, esaterako, *hil*, *sartu*, *atera*, *apurtu*, *jaitsi* eta *zimurtu* aditzak nola agertzen diren *Elhuyar Hiztegia*¹¹⁵, *Hauta Lanerako Euskal Hiztegia* (Sarasola, 1995) eta *Euskal Hiztegi Modernoan* (Elhuyar, 2000):

- *Hil*

Elhuyar Hiztegia: 1 da ad. *morir, fallecer* [...]. 3 du ad. *matar, asesinar, quitar la vida*.

Hauta Lanerako Euskal Hiztegia: 1. da ad. Bizitzetik gelditu. [...]. 2. du ad. Bizia kendu. Ik. garbitu 2; akabatu 2; erail.

Euskal Hiztegi Modernoa: da/du ad. 1. Izaki bizidunaren bizia amaitu, berez edo norbaitek eraginda.

- *Sartu*

Elhuyar Hiztegia: 1 da/du ad. *entrar, meter(se), introducir*.

Hauta Lanerako Euskal Hiztegia: 1. da ad. Barrura joan, barrura igaro. [...]. 2 du ad. (Norbait edo zerbait) barrura joanarazi, barrura igaroarazi.

¹¹⁵<http://hiztegiak.elhuyar.org/> (es/eu hiztegi elebiduna).

Euskal Hiztegi Modernoa: da/du ad. 1. Kanpoko aldetik barrura joan edo eraman.

- *Atera*

Elhuyar Hiztegia: 1 du ad. *sacar*. [...]. 7 da ad. *salir*.

Hauta Lanerako Euskal Hiztegia: du ad. (*1562,1596) 1. (Zerbait) dagoen tokitik at ezarri, barnetik kanpora eraman. Kanpora atera. [...]. 4. da ad. Irten. Ik. ilki.

Euskal Hiztegi Modernoa: da/du ad. 1. Zerbait edo norbait dagoen tokitik kanpora joan edo eraman.

- *Apurtu*

Elhuyar Hiztegia: 1 da/du ad. *romper(se) en pedazos, destrozar*.

Hauta Lanerako Euskal Hiztegia: da-du ad. [...]. 2. (*1745, 1800) Apurretan xehatu, desegin. Ik. txikitu; birrindu.

Euskal Hiztegi Modernoa: da/du ad. 1. Hautsi.

- *Jaitsi*

Elhuyar Hiztegia: 1 da/du ad. *bajar(se), descender; apearse*.

Hauta Lanerako Euskal Hiztegia: da ad. (1545) Toki batetik, beherago dagoen beste batera joan. Ik. jautsi; beheratu. da-du ad. [...]. || du ad. (XIX b.) Toki batetik, beherago dagoen beste batera eraman.

Euskal Hiztegi Modernoa: da/du ad. 1. Toki batetik beherago dagoen beste batera joan edo eraman.

- *Zimurtu*

Elhuyar Hiztegia: 1 da/du ad. *arrugar(se), llenar(se) de arrugas*.

Hauta Lanerako Euskal Hiztegia: da-du ad. (1643) 1. Zimurrak atera edo sortu, zimurrez bete.

Euskal Hiztegi Modernoa: da/du ad. 1. Zimurrak sortu; zimurrez bete.

Ikusten den bezala, *apurtu* eta *zimurtu* aditzetan 3 hiztegiek berdin jokatzen dute eta ez dute alternantzia bakoitzarentzat adiera desberdin bat proposatzen; hau da, gauzak berez *apurtzea* edo *zimurtzea*, eta beste zer-baiten edo norbaiten eraginez *apurtzea* edo *zimurtzea*, adiera berean sartzen dituzte. Baina ez da gauza bera gertatzen *hil*, *sartu* eta *atera* aditzekin. 3 aditz horietan desberdintasunak ikusten ditugu hiztegien artean. *Euskal Hiztegi Modernoak* ez ditu inoiz bi adiera desberdin proposatzen, eta, beraz, alternantzia gisa tratatzen ditu honelakoak; *Hauta Lanerako Euskal Hiztegiak*, aldiz, alternantzia bakoitza adiera desberdin batean sartzen du, eta *Elhuyar Hiztegiak* bi adiera proposatzen ditu *hil* eta *atera* aditzentzat, baina bakarra *sartu* aditzarentzat. Bukatzeko, *jaitsi* aditzaren kasuan, adiera bakarra proposatzen dute *Elhuyar Hiztegiak* eta *Euskal Hiztegi Modernoak*; eta *Hauta Lanerako Euskal Hiztegiak*, nahiz eta 2 adiera desberdin ez proposatu, alternantzia kausatiboa inkoatiboaren azpiadiera gisa ematen du. Zergatik desberdintasun horiek? Zein da arrazoia batzuetan alternantzia moduan lantzeko eta besteetan adiera desberdinak proposatzeko? Inguruko hizkuntzen eraginik ba ote dago bereizketa horretan?

Ikus dezagun, adibidez, ingelesez eta gaztelaniaz zer gertatzen den.

Bi hizkuntza horietan *hil*, *sartu* eta *atera* aditzen alternantzia kausatibo/inkoatiboa bi pieza lexikal desberdinekin gauzatzen da (98)¹¹⁶:

- (98)
- *Hil*
 - a. *hil* [da] >to die/morir
 - b. *hil* [du] >to kill/matar
 - *Sartu*
 - a. *sartu* [da] >to enter (to come in, to go in)/entrar
 - b. *sartu* [du] >to put/meter
 - *Atera*
 - a. *atera* [da] >to go out/salir
 - b. *atera* [du] >to take out/sacar

¹¹⁶Hizkuntza bakoitzak bere jokamoldea du: frantsesak, esaterako, ingelesak eta gaztelaniak bezala jokatzen du *hil* eta *sartu* aditzen kasuan, baina ez *atera* aditzaren kasuan: *hil* [da]>*mourir* eta *hil* [du]>*tuer*; *sartu* [da]>*entrer* eta *sartu* [du]>*mettre*; *atera* [da/du]>*sortir*.

Aztertzen ari garen beste bi aditzen kasuan (*apurtu*, *zimurtu*), aldiz, bi aldaerak adierazteko pieza lexikal bera erabiltzen dute gaztelaniak eta ingelesak (99)¹¹⁷:

- (99) • *Apurtu*
 a. *apurtu* [da/du] >to break/romper(se)
 • *Zimurtu*
 a. *zimurtu* [da/du] >to wrinkle/arrugar(se)

Azkenik, *jaitsi* aditzaren kasuan, desberdin jokatzen dute gaztelaniak eta ingelesak (100)¹¹⁸:

- (100) a. *jaitsi* [da] >to go down/bajar/descender
 b. *jaitsi* [du] >to take down/bajar/descender

Ingelesak bi pieza lexikal desberdin erabiltzen ditu, eta gaztelaniak, aldiz, bakarra bi kasuetan.

Kausatibo/inkoatibo egituretan bi pieza lexikal desberdin erabiltzen dituzten hizkuntzetan, ingelesa eta gaztelania kasu, ez dute alternantzia kausatibo/inkoatiboa dagoela irizten¹¹⁹. Euskararen kasuan, hiztegietan ikusi dugun bezala, zalantzak daude horrelakoak lantzerakoan.

Gure kasuan, BVI lexikoia egitean ingelesezko eredu hartu dugunez oinarri, ingelesez bi pieza lexikal desberdin daudenean, alternantzia bakoitzari ingelesezko aditz desberdin bat esleitu behar izan diogu, nahiz eta EADBn proposatutakoari jarraiki, eta (101) adibidean ikusten den bezala, euskaraz adiera bakarra izan ingelesezko bi aditzen kasuan: entitate baten egoera-aldaketa, *kausa* (aldaera inkoatiboan agertzen ez dena) eta *gai ukitua* rolak eskatzen dituena.

¹¹⁷Baita frantsesak ere: *zimurtu* [da/du] > (se) *froisser*; *apurtu* [da/du] > (se) *casser*.

¹¹⁸Kasu honetan, frantsesak gaztelaniaren bideari jarraitzen dio, eta bi aldaeretan pieza lexikal bera erabiltzen du: *jaitsi* [da] > *descendre/baisser* eta *jaitsi* [du] > *descendre/baisser*.

¹¹⁹“En este trabajo no se han considerado como alternancias aquellas estructuras en las que participan verbos léxicamente diferentes, como los pares *matar/morir*, *comprar/vender*, *dar/recibir*, *like/please*. Respecto al primero de estos pares, se establece entre los dos ítems una relación de causa, donde *matar* se puede parafrasear por “causar morir” (Lakoff 1965, Lyons 1968)” (Vázquez *et al.*, 2000: 94).

- (101) 1. *hil*: *kill.01*
- Arg0: agent, kausa (ERG)
 - Arg1: patient, gai ukitua (ABS)
 - Arg2: instrument, - (SOZ/INE/INS)
2. *hil*: *die.01*
- Arg1: theme, gai ukitua (ABS)
 - Arg2: instrument, - (SOZ/INE/INS)

Bestalde, nahiz eta alternantzia kausatibo/inkoatiboen multzoan sartu ez eta EADBko kontua gehiago izan, zenbait aditzetan adierak desberdintze-ko zailtasunak izan ditugu. Euskarazko bi adiera desberdintzea zaila bada, eta ikusten badugu ingelesez biak adierazteko aditz berbera erabiltzen dutela, beharbada sakonago aztertu beharko genuke hor benetan bi adiera ote dauden erabakitzeko.

Hasi aditzak, esaterako, EADBn bi adiera ditu. Lehenengoan (DA-DU), entitate baten egoera-aldaketa adierazten da, entitatea irautes hasten den neurrian, parte-hartzaile nagusi gisa *gai ukitua*, eta, batzuetan, *kausa* dituen. Bigarrenean (DA), entitate baten jardueraren hasiera adierazten da; baina ingelesez biak *begin.01/start.01* aditz-adierekin adierazten dira, eta zaila egin zaigu bien artean desberdintzea:

1. Egoera-aldaketa

- DA: *Dena kasualidadez hasi zen.*
- DU: *Hiru lagunek gose greba hasi dute.*

2. Entitate baten jardueraren hasiera

- DA: *Urrian hasi ziren prestaketarekin/Zinemagintzan hasi zine-nean [...]*

Ez dugu argi ikusten hemen benetan bi adiera dauden edo ez: zer desberdintasun dago *urrian hasi ziren prestaketarekin* eta *hiru lagunek gose greba hasi dute* esaldien artean? Ez al da bietan entitate baten jardueraren hasiera? Oraingoz, bi adiera desberdin ditugu lexikoian (6.10 taula), baina, ikusten denez, argumentu-kopuru bera eta VNko rol berberak dituzte biek.

1- <i>begin.01/start.01</i>	2- <i>begin.01/start.01</i>
Arg0: Agent, kausa [+giz] (ERG)	Arg0: Agent, esperimentatzailea (ABS)
Arg1: Theme, gaia ukitua [-biz] (ABS)	Arg1: Theme, jarduera (INE/SOZ)
Arg2: Instrument, - (INS/SOZ)	Arg2: Instrument, - (INS/SOZ)

6.10 taula: *hasi* aditzaren sarrera BVI lexikoian.

Aurrera begira, euskara hutsezko lexikoa egiten dugunean, *hasi* aditzari bi adiera edo adiera bakarra esleitzen diogun erabaki beharko dugu.

Etxeparek (2003) *hasi* aditz aspektualen multzoan sartzen du, eta, behar-bada, bere izaera berezi horrek ere badu zerikusia adierak bereizteko dugun zailtasun horrekin.

6.2.5 Ingelesez existitzen ez diren aditzak

Euskaraz askotan gertatzen da predikatuari esanahia ematen dion elementua aditzetik kanpo geratzea, eta ingelesez, aldiz, gauza bera esateko hitz bakar bat izatea¹²⁰. Baina alderantzizko norabidean ere gauza bera gertatzen da, hots, ingelesak izatea predikatuari esanahia ematen dion elementua aditzetik kanpo eta euskaraz hitz bakar bat egotea hori adierazteko; adibidez, *to have lunch* (bazkaldu) edo *to get lost* (galdu)¹²¹. Horrelako kasuetan, PB-VN-n ez dugu euskarazko aditzentzat ordain egokirik aurkitu, eta, beraz, *have_lunch.01* eta *get_lost.01* aditz-adierak asmatu ditugu gure euskarazko aditzen ordain egoki gisa, eta EADBk ematen dituen argumentuak eta rolak esleitu dizkiegu asmatutako “aditz-esapide” horiei. Ikusten dugunez, fenomeno horrek bi norabideetan sortzen dizkigu arazoak euskarazko eta ingelesezko ereduak elkarrengana hurbiltzeko.

6.3 Euskarazko aditzen inguruan

Atal honetan, euskararen beraren ezaugarriak direla eta, etiketatzean aurkitu ditugun hiru gai interesgarri aipatuko ditugu.

Hasteko, HAULEtarako joera duten aditzak etiketatzean aurkitutako hainbat kasu zalantzazko eta horien aurrean hartutako erabakiak aipatuko ditugu. Jarraian, corpora etiketatzean aurkitutako balio semantiko berrien hainbat

¹²⁰Ikus 6.3.1 ataleko (104) adibidea fenomeno horren erakusgarri.

¹²¹Baita 6.2.2 atalean ikusi ditugun *go up/go down* eta antzekoak ere.

adibide aurkeztuko ditugu, eta, bukatzeko, aditz aspektualen inguruan jardungo dugu.

6.3.1 Hitz Anitzeko Unitate Lexikaletan parte hartzeko joera handia duten aditzak

Badira Hitz Anitzeko Unitate Lexikaletan (HAUL) parte hartzeko joera berezia duten aditzak (*eman, jarri, hartu, pasatu, egin, etorri...*)¹²², eta horrelakoak etiketatzeak zalantza bat baino gehiago sortu digu. Askotan ez dago oso argi beste adiera baten edo HAUL baten aurrean gauden. Mugimenduzko aditzetan bezala, ingelesezko ordaina topatzea ez da lan erraza izan, eta horrek are ageriago uzten du aditz-esapideen aurrean egon gaitezkeela.

Horrelako adibideak etiketatzean, HAULa ziurra bada (Urizar, 2012), besterik gabe etiketatu gabe utzi eta oharren eremuan “aditz-esapidea”¹²³ dela markatu dugu (*parte hartu*, esaterako). Baina zalantzazko kasuetan, hau da, HAULetara hurbiltze-bidean dauden horien kasuan, agerpena etiketatu egin dugu, baina oharretan horiei ere “aditz-esapidea” jarri diegu. Horrela, etorkizunean horrelako adibideak sakonago aztertu nahi izanez gero, identifikatuta izango ditugu.

Jarraian, ikus ditzagun esapideetarako joera duten hainbat aditz eta zalantzak sortu dizkiguten agerpenen adibideak:

- *Eman* aditzak datiboa (DAT) azpikategorizatzen du, maizen agertzen duen elkarraldatze-esanahian, baina adibide asko aurkitu ditugu non DATa ez den posible, eta, hain zuzen, horrelakoetan esanahia elkarraldatzetik urrundu egiten da. Horregatik iruditzen zaigu horrelakoetan HAULetatik hurbilago gaudela: *pausoak eman, amore eman, urratsak eman* etab.
- *Jarri* aditzaren agerpenetan honelakoak aurkitu ditugu: *kolokan jarri, zalantzan jarri, auzitan jarri...* Aditz honen lehen bi adierak “kokapen-aldaketa” eta “egoera-aldaketa” dira. Lehenengoan, inesiboa (INE) azpikategorizatzen da (*mahaian jarri*) eta bigarreanean, absolutiboa (ABS) edo adberbioa (*gorri jarri/gaizki jarri*), helburuko egoera adierazteko. Hasieran aipatutako kasuetan (*kolokan, zalantzan, auzitan*) INEa dugu, baina ez dirudi kokapen baten aurrean gaudenik, hau da,

¹²²Horrelako aditzak “aditz arinak” deitu izan ohi dira (Zabala, 2004; Urizar, 2012).

¹²³Ikusi oharrei buruzko informazio zehatza 4.3 atalean.

INE horrek ez du leku bat adierazten, baizik eta egoera bat: gai ukitua “nola” dagoen adierazten du eta ez “non” dagoen. Horrelakoak HAUL direla edo HAUL izateko bidean daudela iruditzen zaigu, eta gure erabakia izan da “egoera-aldaketa” bezala etiketatzea, INEa onartuta alternantzia gisa (alegia, ABS/INE arteko alternantzia planteatzea), betiere, “aditz-esapidea” oharra jarrita, horren bidez adieraziz INEa baino gehiago aditzondo lexikalizatua dela esapidearen lehen osagaia.

- *Hartu* aditzaren kasuan honelako esapideak aurkitu ditugu: *erabakia hartu, neurriak hartu, kaleak hartu, mendean hartu, kontuan hartu, indarra hartu...* Aditz honi bi balio semantiko definitzen zaizkio EADBN:

1. Entitate baten kokapen-aldaketa, ablatiboa (ABL) azpikategorizatzen duena (102):

(102) *Lasterketa buruan hartu zuen lekukoa Monique Hennagan kidearen eskutik, baina Hemmings jamaikarra gain-gainean zuela.*

2. Entitate baten edukitzea, INEa azpikategorizatzen duena (esanahi berezi samarra azken hau (103)):

(103) *Ibarretzek Lehendakaritzako aretoan hartu zuen Díez.*

Jakin badakigu aditz honek HAULEtan parte hartzeko joera nabarmena duela. Lehen *emanekin* ikusi dugunaren antzera, “leku-aldatze” esanahian azpikategorizatzen den ABLa ez da posible adibide askotan (*erabakia hartu, neurriak hartu, kaleak hartu*); eta, gainera, ABSa entitate abstraktua izaten da edo zentzu figuratiboan erabiltzen da. Horrek guztiak “leku-aldatze” esanahitik urrundu egiten ditu horrelako adibideak. Beste batzuk, aldiz, bigarren adieratik hurbilago daude, baina INEa ez da izaten kokapena edo lekua, eta horietan ere INEaren oinarrian dagoen izena abstraktua izan ohi da: *mendean hartu, kontuan hartu*, etab. Horrelako agerpenak aipatutako balio semantikoetako batean sailkatzea zaila izaten da, eta horrek erakusten digu, hain zuzen, HAULEtatik hurbilago daudela.

- *Etorri* aditza ere maiz agertzen da HAULEtan edo HAUL izateko bidean dauden esapideetan (*nor bere onera etorri, burura/gogora etorri, bat etorri, etab.*).

Bat etorriren kasua argia da, ez baititu onartzen aditz honetan hain ohiko diren ABLa (abiapuntuko kokapena), ez eta ALA-a (helburuko kokapena) ere. Hiztegietan ere esapide gisa markatzen da, eta [Urizarren \(2012\)](#) lanean HAUL ziurren zerrendan agertzen da; beraz, ez dugu etiketatu, eta “aditz-esapidea” oharrarekin markatu dugu.

Nor bere onera etorriren kasuan, ALAren bidez ez dirudi helburuko kokapena adierazten denik, eta *Elhuyar Hiztegian*, adibidez, *on* lema-*ren* sarreran agertzen da esapide gisa (*nor bere one(ta)ra etorri/itzuli*).

Etorrirekin bukatzeko, *burura/gogora etorri* gisako adibideak aurkitu ditugu. Kasu horietan iruditzen zaigu EADBko bigarren balio semantikoan sartzen direla, hau da, “sortze-prozesuan”. Hala ere, adiera honetan ez da batere ohikoa ALA-a (*inspirazioa etorri zaio, hilekoa etorri zaio, kakagura etorri zaio...*) agertzea, eta, agertzen denean, gunean gizakiaren gorputz-atal bat du sintagmak (*hitzak ezpainetara etorri, Mikel etorri zaio gogora, ideia bat etorri zaio burura...*). Adlatiboan duten kasu horiek *etorriren* “desplazamendu” adierarekin nahasgarri gertatzen dira, baina, gogoan izan behar da DATan doan entitatearen gorputz-atalak direla beti, eta orduan bakarrik agertzen direla (DATAren presentzian). Horretaz gain, corpusean *burura* eta *gogora* duten adibide asko agertu zaizkigu, eta fenomeno horrek HAUL izateko bidean jartzen dituela uste dugu.

- Azkenik, jakina da *egin* aditzarekin osaturiko HAULEn multzoa oso handia dela. Aditz horri adiera bakarra definitzen zaio EADBn, “sortze-prozesua”, hain zuzen ere; baina badira balio horretatik aldentzen diren hainbat eta hainbat adibide.

Ezaguna da, eta aipatu dugu [6.2.5](#) atalean ere, euskarak zenbait aditz-motatan material gutxiago lexikalizatzen duela aditzean ([Agirre et al., 2006b](#)). Hala, euskaraz joera dago predikatuari esanahia ematen dion elementua aditzetik kanpo geratzeko¹²⁴, hau da, aditzean material gutxiago lexikalizatzen da. Ingelesaren moduko hizkuntza batean, esaterako, *to bike* edo *to walk* gisako aditzak ditugu eta horiek euskaraz *bizikletaz ibili* eta *oinez ibili* izango lirateke. Euskarazko bi egitura horietako aditzondo-osagaiak (*bizikletaz* eta *oinez*) sintaktikoki modifikatzaileak dira. Alegia, ingelesez bi aditz desberdin erabiltzen dira eta

¹²⁴Mugimenduzko aditzekin ez bezala. Ikusi [6.2.2](#) atala.

euskaraz aditz berbera erabiltzen da, baina modifikatzaile batez lagunduta. Orobat, gaztelaniazko *traicionar*, *presionar*, *morder* moduko aditzekin, euskaraz *egin* aditza etiketatzean agertu zaizkigun *traizio egin*, *presioa egin* edo *koska egin* ditugu. Horrelako kasuetan, gaztelaniaz aditzean bertan lexikalizaturik dagoen elementua, euskaraz aditzetik kanpo adierazten da. Adibide horiek sortzezko esanahiarekin etiketatu ditugu, baina “aditz-esapidea” oharra jarri diegu (104).

- (104) a. *urrats eta pauso handiak egin zituzten*
 b. *traizio egin diola*
 c. *barrutik boikota egiteko*
 d. *zuzeneko emankizuna egingo dute*
 e. *errieta egingo digute*
 f. *koska egin*
 g. *presioa egin*

Bukatzeko, esan behar dugu, HAULEtan parte hartzeko joera handia duten aditzetan, zalantza nagusia adiera berri baten eta HAULaren arteko mugak bereiztean sortu zaigula (hots, EADBn landuta ez dagoen beste adiera bat den edo HAUL den jakitea). Jarraian, *bete*, *atera* eta *jo* aditzetan aurkitutako zalantzazko kasuak aipatuko ditugu.

- *Bete* aditzean *urteak bete* egiturarekin izan dugu zalantza. EADBn, egitura berezia dela aipatzen da:

“Bigarren balioko *bss* hauetan, ABS entitatea beti izaki abstraktu bat da. Zer abstraktu honek batzuetan denborari egiten dio erreferentzia, halako balio berezi bat emanez aditzari. Ez dugu bereizi nahi izan besteetatik, ordea, azken batean zer abstraktu gisa ikus daitekeelako” (Aldezabal, 2004: 214).

Argi dago, beraz, horrelako adibideak bereziak direla, baina oraingoz ez dugu adiera berririk sortu eta “aditz-esapidea” oharrarekin markatu ditugu.

- *Atera* aditzean, (105)-an ikusten ditugun adibideak topatu ditugu:

- (105) a. *Lana ondo atera zait.*
 b. *Armei probetxu atera die.*

EADBN, (105)-ean ikusten ditugun adibideak “sortze-prozesua” adierazten duen balio semantikoan sartuta daude. Erakusten duen kasuistika askotarikoaren aurrean, zalantza handiak izan ditugu aditz horren adierak desberdintzen, adiera berriak edo HAULak diren erabakitzen, etab. Aipatutako bi kasuetan, HAUL izateko bidean dauden esapideen aurrean gaudela uste dugu¹²⁵.

- *Jo* aditza etiketatzean (106)-an ikusten dugun adibidea aurkitu dugu:

(106) *Tiroa jo dio.*

(106) adibidea *joren* “kolpekatze-jarduera” adierazten duen balio semantikoan koka dezakegu, baina EADBN bertan esaten da adiera honetan datiboa agertzea ez dela batere ohikoa. Izan ere, adibide bakar bat ematen du EADBk (*ostiko bat jo zion aulkiari*), eta, gainera, badirudi adibide horretan *jo* aditzaren ordeztu, naturalagoa dela *eman* aditza agertzea. Gure kasuan ere, *tiroa jo* beharrez, normalagoa da *tiro egin/eman* agertzea (esapide horiek agertzen dira *Elhuyar Hiztegian*). Kontuan izanda corpusean agerpen bakarra aurkitu dugula, gure erabakia, oraingoz, “aditz-esapidea” oharrarekin etiketatzea izan da.

Gauzak horrela, ez dirudi adiera berrien aurrean gaudenik, baina argi dago horrelako adibideak bereziak direla, eta, beharbada, HAUL direla edo HAUL izateko bidean daudela; beraz, guk etiketatu egin ditugu eta oraingoz “aditz-esapidea” oharrarekin markatu ditugu, azterketa sakonagoaren zain.

6.3.2 EADBko aditzen aberastea: balio semantiko berriak

Zenbait kasutan lexikoietan agertzen ez diren adierak agertu zaizkigu corpusean, hau da, EADBk eta PB-VNk definituta ez dituzten adierak aurkitu ditugu. Etiketatze-lan horrek, beraz, EADBN landutako aditzen sarrerak aberasteko aukera eman digu. Jarraian, horrelako kasuen adibide aipagarri batzuk emango ditugu.

¹²⁵Esan beharra dago *Elhuyar Hiztegian* lokuzio gisa agertzen dela *probetxu ateraren* antzekoa den *aldea atera* esapidea.

- *Jokatu* aditzak, esaterako, bi adiera ditu EADBn:

1. Jarduera:

- esperimentatzailea [+giz] (ERG)
- jarduera (ABS/INE)

2. Gizaki baten jarrerazko jarduera:

- esperimentatzailea [+giz] (ERG)
- modua (ABS/INS)

Corpusean *apustua egin* esanahia duen eta bi adiera horietan sartzen ez den adibide bat aurkitu dugu (107), eta, zalantzak izan baditugu ere, adiera berria sortzea erabaki dugu.

(107) *Nahi duzuna jokatzan dizut hori lortuko dugula.*

EADBn honelako adibideak aditzaren lehen predikatu nagusian sartu dira, nahiz eta Aldezabalek (2004) HAUL gisa azter daitezkeela onartu:

“Adibideek erakusten digutenez, bss hauetan guztietan gizaki batek (edo hau ordezkatzan duen entitate batek: (1046), (1049)) orotariko jarduerak egiten ditu. Jarduera izanik, beti entitate abstraktu bat dute gunean jarduera hori adierazten duten sintagmek. [...]. Hala ere, esan behar da badirela hitz batzuk jardueratik urrun samar daudenak ((1045): diru asko), eta gure irudiz horrexegatik egokiago litzateke beharbada Hitz Anitzeko Unitate Lexikal gisa aztertzea” (Aldezabal, 2004: C eranskina, 230).

Gure kasuan, bi arrazoi nagusi izan ditugu adiera berria sortzeko erabakia hartzeko:

1. Adiera honek beste bi adierek baino argumentu bat gehiago du.
2. Beste bi adierek onartu ezin duten DAT postposizio-atzizkia onartzen du.

(107) adibidearentzat ingelesezko ordain egokia *bet.01* da (6.11 taula), eta horren argumentuak esleituko dizkiogu gure euskarazko adiera berriari ere (6.12 taula).

bet.01

vncls: 54.5, 94, 70	
Arg0	better (vnrole: 54.5-agent, 94-agent, 70-agent)
Arg1	amount of bet (vnrole: 54.5-asset, 70-theme)
Arg2	basis, proposition, bet-on (vnrole: 94-theme)
Arg3	co-better

6.11 taula: *bet.01* adiera PBn.

jokatu 3 - bet.01

Arg0	agent	esperimentatzailea (ERG)
Arg1	asset	neurria (ABS)
Arg2	theme	gaia (ABS/INE/KONPL)
Arg3	co-better	- (SOZ/DAT)

6.12 taula: *jokaturen* hirugarren adiera BVIn.

Lexikoian informazio hori izanda, (108)-an ikusten den bezala etiketatuko genuke (107) adibidea.

- (108) Arg1: asset, neurria (*nahi duzuna*)
 Arg2: theme, gaia (*hori lortuko dugula*)

- *Eman* aditzean ere, EADBn definituta ez dagoen adiera berri bat aurkitu dugu corpusean. EADBn 3 adiera ematen dira aditz honentzat:
 1. Elkarraldatzea.
 2. Denbora-epe baten barruan egiten den jarduera.
 3. Gertakizun baten inguruko ustea adierazten duen jarduera¹²⁶.

Corpusean (109) adibidea aurkitu dugu:

- (109) *Amak esan zidanean zure leihoak eraikuntzaren gibelean zegoen arkadira ematen zuela [...].*

¹²⁶Corpusean ez da agertu adiera horren adibiderik, horregatik ez dugu BVIn sartu.

Adiera honetan ALA-a agertzea ezinbestekoa da; EADBn definitutako hiru adieretan, aldiz, ALA postposizio-atzizkia onartu gabekoen¹²⁷ artean sartzen da. Horrek, beste adieretatik desberdintzeko ezaugarri bat ematen digu; beraz, adiera berria proposatu dugu BVIn, PBko *overlook.02* aditz-adierarekin bat egiten duen adiera, hain zuzen ere. PB-VNko eta BViko adiera horren informazioa 6.13 taulan ikus daiteke.

PB-VN: overlook.02	BVI: eman_3
Arg0: place ¹²⁸	Arg0: agent, esperimentatzailea (ERG)
Arg1: direction	Arg1: destination, helburua (ALA)

6.13 taula: PB-VNk eta BVik *overlook.02* eta *eman_3* adierentzat ematen duten informazioa.

- *Egon* aditzaren kasuan ere, EADBn definituta ez dagoen adiera berri bat aurkitu dugu corpusean. EADBk 2 adiera definitzen ditu aditz honentzat:

1. Entitate bat leku batean kokatuta egotea.
2. Entitate baten egoera.

Biek dituzte bi argumentu eta bietan ez da datiborik (DAT) onartzen. (110)-ean corpusean aurkitutako adibidea ikus dezakegu.

(110) *Eremu pribatua etxe barrenari dagokio.*

(110) adibidean DATA ezinbestekoa da: adiera honek DATA eskatu egiten du, eta, beraz, beste bi adieretatik bereizten da. Horregatik, proposatu dugu BVIn “ezaugarri/eginkizun bat esleitzea” esanahia duen adiera gehitzea. PB-VNko ordain egokia *correspond.02* da. Sortu dugun adiera berriak bi postposizio-atzizki azpikategorizatzen ditu, ABSa eta DATA, hain zuzen ere, eta PB-VN ereduari jarraiki, rolak *Theme*

¹²⁷Ikus Aldezabalaren (2004) C eranskineko 79., 81. eta 82. orrialdeak.

¹²⁸PBn agertzen da (*place/direction*), baina VN-n ez, beraz, ez ditu VNko rolak esleituta. VNko sistema jarraituz, guk gehitu ditugu ingelesezko rolak BVIn.

eta *Theme* izango dira¹²⁹. 6.14 taulan ikus daiteke adieraren sarrera BVIn.

egon 3 - correspond.02

Arg0	theme	gaia (ABS)
Arg1	theme	helburua (DAT)

6.14 taula: *egonen* hirugarren adiera BVIn.

- *Eraman* aditzean beste adiera berri bat topatu dugu (111):

(111) *Ezin dut zure laguna eraman.*

EADBN 3 adiera bereizten dira *eraman* aditzarentzat:

1. Leku-aldaketa:

- kausa_ERG, gai ukitua_ABS, helburuko kokapena_ALA, abiapuntuko kokapena/bidea_ABL

2. Edukitze bat:

- edukitzailea_ERG, edukia [-biz]_ABS, kokapena_INE

3. Denbora iragatezko jarduera bat:

- esperimentatzailea_ERG, denbora_ABS, jarduera_INE

Lehenengoan, ERGaz eta ABSaz gain, ABLa eta ALA-a azpikategorizatzen dira, eta bigarrenean eta hirugarrenean, aldiz, INEa. (111) adibidean ikusten dugu hiru postposizio horiek (ABL, ALA eta INEa) ezinezkoak direla adiera horretan; beraz, adiera berri baten aurrean gaude. Adiera honentzat PBko *stand.03* ordaina aukeratu dugu (6.15 taula):

¹²⁹*correspond.02* aditz-adieran ez da VNko klaserik agertzen, eta, beraz, ez du VNko rolik esleituta. Euskaraz Arg1 argumentuari *helburua* rola esleitu diogu, rol hau “ezaugarri-esleitze” adieran PBk ematen duen *Theme* rola baino ohikoagoa delako.

eraman 5¹³⁰- stand.03

Arg0	experiencer	esperimentatzailea (ERG)
Arg1	theme	gaia (ABS)

6.15 taula: *eramanen* bosgarren adiera BVIn.

Bukatzeko, *egin* aditzaren bi adibide aipatu behar ditugu (112).

- (112) a. *Ikerketek aldi hartan egin zuten gainbehera.*
 b. *Gorantz egin genuen.*

Aditz honek, aurretik ikusi dugun bezala, EADBko adiera bakarrean, bi argumentu eta rol ditu (*iturria*, ERGan gauzatzen dena, eta *gai sortua*, ABSan gauzatzen dena), eta ALA-a ez du onartzen. Horregatik, goiko adibideek zalantza sortu digute. Badirudi horrelakoetan *-ra(ntz) egin* moduko zerbait dugula eta *norabait joan* esanahiaren aurrean gaudela. Gainera, *joan* eta *etorri* aditzetan ezinezkoa den *Extent* argumentua posible da *eginen* balio honetan (*200 metro egin genituen gora*). Beraz, nahiko argi ikusten dugu adiera berri baten aurrean gaudela, eta horrela adierazi dugu BVIn ere, *egin* aditzari “kokapen-aldaketa” adiera berria gehituz.

Badira zalantzazko kasuak ere (113), eta, batzuetan, ez dago oso argi adiera berri baten aurrean, testu-mota jakin bateko erabilera berezi baten aurrean edo, besterik gabe, elipsi baten aurrean gauden.

- (113) *Epaileak txartela atera dio.*

(113) adibidean zer daukagu? Adiera berria, kirolaren eremuko adiera berezia ala elipsia? Adibide horretako adiera zer da? “Kokapen-aldaketa” ala “sortze-prozesua”? Badirudi gehiago dela “kokapen-aldaketa”, hau da, “txartela epailearen poltsikotik kanpora atera dela”, “jokalariaren aurrera atera dela”, baina abiapuntuko eta helburuko kokapen horiek ez direla inoiz esplizitu agertzen, baizik eta testuinguruagatik ulertzen ditugula. Kirolaren inguruko corpus batean horrelako zerbait ikusiz gero, denok ulertzen dugu “txartela poltsikotik atera diola”, baina inoiz ez dugu esplizitu idatzita ikusiko.

Horrelako adibideen aurrean, argumentuen izaera aztertu dugu eta hiru elipsi-mota bereizi beharrean aurkitu gara (Aldezabal, 2004: 129):

¹³⁰ *Eraman* aditzari *win.01* PBko ordainarekin lotzen den 4. adiera berri bat ere sortu diogu. Horregatik da hau 5. adiera.

- **Elipsi testuala:** Erreferentziakidetasunez testutik bertatik berreskura daitekeena (114).

(114) *Mikel **Indiara** joan zen oporretan eta [Indiara] iritsi bezain pronto, damutu egin zen.*

Aurreko predikatuaren argumentuak begiratuta, *iritsi* predikatuaren *destination* argumentua *Indiara* dela ondoriozta dezakegu (gauza bera predikatu bien *theme* argumentuarekin). Elipsia testuari dago lotuta, eta, beraz, eliditutako osagaia hortik berreskura daiteke.

- **Elipsi pragmatikoa:** Munduaz dugun ezagutzagatik ulertutzat ematen den argumentu bat ageri ez denean; hau da, argumentua ez da agertzen eta normalena da, gainera, ez inoiz agertzea (baina esanahi batetik abiatuta). Horrelakoetan, adiera bereziak ageri dira, adiera mugatuak, argumentuak ere jakinak direlako. Horrelakoak maiz agertzen dira testu-mota jakin bat aztertzen ari garenean, corpus berezituaren kasuan, hain zuzen ere (ikusi (113) eta (116) adibideak).
- **Infraespezifikazio-elipsia:** Hemen ez da argumentu bat ageri, baina besterik gabe esaldi horretan ez delako azpimarratu nahi beste zerbait gehiago azpimarratzeko. Esaterako, *abesturen* kasua (115):

(115) *Ez du ondo abesten.*

Badakigu doinuren bat izango dela, baina ez dugu zehazten ez zaigulako bereziki interesatzen, eta memento horretan ekintza bera azpimarratu nahiago dugulako.

Hori ikusita, hasieran jarri dugun (113) adibidean, corpus berezitu batean gertatzen den elipsi pragmatikoa izango genuke, munduaz dugun ezagutzagatik ulertutzat ematen den argumentu bat ageri ez denean; hau da, argumentua ez da agertzen eta normalena da, gainera, ez inoiz agertzea. Honenbestez, kirolari buruzko corpus berezituaren *atera* aditzaren “kokapen-aldaketa” adierak hartzen duen balio sintaktiko-semanticoko berezi baten aurrean egongo ginateke (Aldezabal, 2004: 128). Iruditzen zaigu gai hau sakonago aztertzekoa dela, eta euskarazko aditzentzat proposatzen ditugun adiera orokorrak corpus berezituetan nola gauzatzen diren eta elipsiak zer nolako eragina duen ikustea izan litekeela etorkizuneko lan interesgarri bat.

Gauza bera gertatzen da, elipsi pragmatikoa, hain zuzen ere, (116)-an ikusten ditugun *joan* eta *eraman* aditzen adibideetan:

- (116)
- *67 urtetan joan da, gaixotasun larri batek jota.*
 - *Antzerki saria Aitzpea Goenagak eraman zuen “Zu(T) gabe” obrarekin.*

Adibide horietan munduaz dugun ezagutzagatik ulertzen dugu pertsona bat joan dela [mundu terrenal honetatik] [beste norabait-edo] edo pertsona batek eraman duela saria [eman dioten lekutik] [bere etxera].

Adiera berriak aurkitzearena etenik ez duen fenomeno da, batetik, corpus berriak ustiatu ahala gertatuko baita, eta, bestetik, adiera berriak sortzea beharrezkoa baita hizkuntzak bizirik iraungo badu.

6.3.3 Aditz aspektualak

Zenbait aditzek euren balio semantikoa adierazi beharrean, aditzaren aspektuari buruzko informazioa ematen dute egitura jakin batzuetan. Horrelako aditzak “aditz-aspektualak” deitu izan ohi dira (Etxepare, 2003). EADBko aditzen artean *joan* eta *aritu* aditzak sartuko genituzke multzo honetan:

- *Ohitzen joan da/ugalduz joan da.*
- *Lanean aritu da/idazten aritu da.*

Bietan ikusten dugu aditzak, balio semantikoa adierazi beharrean, aurretik duen elementuaren nolabaiteko progresioa adierazten duela, hau da, jardueraren aspektua. *Joan* aditzaren kasuan, honelako adibideak ez ditugu etiketatu eta “aditz-perifraasi aspektuala” oharra jarri diegu, aspektu balio honek ez baitu inolako zerikusirik *joan* aditzak berezkoa duen kokapen-aldaketa adierarekin. *Aritu* aditzaren kasuan, aldiz, adibideak etiketatu egin ditugu, baina “aditz-perifraasi aspektuala” oharra ere gehitu diegu. Kasu honetan, aditzak berez duen balio semantikoa “zerbaitetan jardutea” da, hau da, ekin-tzaren aspektu progresiboa adierazten du. Aditzaren eduki semantikoaren falta dela eta, ingelesezko ordaina topatzean arazoak izan ditugu. Azkenean, *be.03* ordaina aukeratu dugu. Aurretik esan dugun bezala (ikusi 5.2.2 atala), PBN *to be* aditza ez dute etiketatu eta *be.03* adiera *auxiliary* dela esaten dute; beraz, ez diote argumenturik esleitzen. Horrek guztiak *aritu* aditzaren izaera aspektuala agerian uzten duela uste dugu, eta, horregatik, nahiz eta argumentu eta rolekin etiketatu, “aditz-perifraasi aspektuala” oharra ere jarri diegu.

EADBn landu diren aditzen artean, badira nolabaiteko aspektua adierazten duten bestelako aditzak ere, hala nola, *hasi* (*idazten hasi da/lanean hasi da*) eta *egon*¹³¹ (*lanean dago*). Hala ere, bi kasu horietan guk ez dugu “aditz-perifraasi aspektuala” oharrik jarri. Erabaki hori hartzeko bi arrazoi nagusi izan ditugu kontuan: i) aditzak aspektua adierazten duen balio semantikoa du; eta ii) ingelesezko ordaina topatzean, ez dugu inolako arazorik izan¹³².

6.4 Laburbilduz

Kapitulu honetan EADBn landutako aditzak etiketatzean agertutako kasuistika azaldu dugu, hiru azpiataletan banatuta:

1. EADB eta PB-VN ereduaren arteko adostasunak eta desadostasunak
2. Euskara eta ingelesaren arteko aldeak
3. Euskarazko aditzen inguruan agertutako fenomenoak

Hasteko, ereduaren arteko adostasunez eta desadostasunez aritu gara. Lehenbizi, ereduak bat egiten duten kasu argi bat aurkeztu dugu, eta, jarraian, bi ereduak bat etorri ez direnean hartu ditugun erabakiak edo egin ditugun moldaketak azaldu ditugu. Bi desadostasun nagusi aurkitu ditugu:

- **EADB eta PB-VN ez datoz bat argumentu/adjuntu bereizketan.** Normalean PB-VN-k EADBk baino argumentu gehiago esleitzen ditu, eta, gure irizpide nagusia PB-VN ereduari jarraitzea denez, kontrako arrazoirik egon ez denean, PB-VN-k ematen dituen argumentuak esleitu dizkiogu euskarazko aditzari.
- **PB-VN-n ez da EADBn agertzen den *Path* (bidea) rola agertzen.** Corpora etiketatzeko rol honen beharra izan dugu, eta, beraz, *Path* rola gehitu dugu PB-VNko rolen zerrendan.

Bigarren puntuan, euskararen eta ingelesaren artean aurkitu ditugun aldeez jardun dugu, eta hemen bost azpiatal bereizi ditugu:

¹³¹ *Egon* aditzaren kasuan, “aditz-perifraasi modala” oharra ere erabili dugu, *ez dakit hori botatzerik dagoen* eta *arrazoi duelakoan nago* gisako egiturak etiketatzeko.

¹³² Etxeparek beste aditz batzuk ere sartzen ditu multzo horretan, besteak beste, *ibili*, *bukatu*, *segitu* edo *jarraitu* (Etxepare, 2003). Aditz horiek ez daude EADBn landuta, eta, beraz, ez dugu oraindik horien gaineko azterketa sakonik egin.

- **Euskaraz posible ez den argumentu bat agertzen da PB-VN-n.** Ingeleseko mugimenduzko zenbait aditzetan (*come, go...*) agertzen den *Extent* argumentua ezinezkoa da euskaraz. Horrelako kasuetan, posible ez den argumentu hori ez dugu kontuan hartu BVIko sarrera definitzean, eta ondoren datozen argumentuak aurreratu egin ditugu.
- **Mugimenduzko aditzak: euskara aditz-hizkuntza vs ingelesa satellite-hizkuntza.** Ingelesa satellite-hizkuntza da eta euskara, aldiz, aditz-hizkuntza. Satellite-hizkuntzek aditzetik kanpo ematen dute mugimenduaren norabidea adierazten duen elementua, eta aditz-hizkuntzek, berriz, aditzean txertatuta dute norabide hori. Horrek arazoak sortzen ditu bi hizkuntza hauek parekatzean, eta, horien aurrean, kasuan kasuko erabakiak hartu behar izan ditugu.
- **Euskaraz aditz bat, ingelesez ordain bat baino gehiago.** Hau fenomeno arrunta da eta kontrako norabidean ere gertatzen da, baina euskarazko aditz batek ingelesezko ordain bat baino gehiago dituztenean, eta ordain horiek argumentu eta rol desberdinak dituztenean, arazo bat dugu. Oraingoan, eta PB-VN eredu jarraitzen ari garela kontuan izanda, agerpen bakoitzari ordainik egokiena jartzea erabaki dugu, baina jakin badakigu, euskara hutsezko lexikoia osatzen dugunean, adiera batek ezingo dituela argumentu eta rol desberdinak izan agerpenaren arabera.
- **Ingelesak eta euskarak ez dituzte kausatibo/inkoatibo egiturak berdin lexikalizatzen.** Zenbaitetan, DU-DA erabilera duten aditzetan, ez dago argi egitura horiek adiera beraren alternantzia ala adiera desberdineko egiturak ote diren. Euskarazko hiztegieta, batzuetan adiera desberdin gisa lantzen dira, eta beste batzuetan, aldiz, alternantzia gisa. Ingelesean eta gaztelaniaz, zenbaitetan, kausatibo/inkoatibo egiturak pieza lexikal desberdinekin gauzatzen dira, eta badirudi horrek eragina izan dezakeela euskaraz adiera desberdin gisa tratatzeko erabakian. Gure kasuan, ingelesezko eredu jarraitzen ari garenez eta euskarazko aditzari ingelesezko ordaina esleitzera behartuta gaudenez, ingelesez bi pieza lexikal daudenean, alternantzia bakoitzari ingelesezko aditz desberdin bat lotu behar izan diogu, nahiz eta, EADBri jarraiki, euskarazko adiera bakarra izan.
- **Euskarazko hainbat aditz ez dira ingelesez existitzen.** Esate-

rako, euskarazko *bazkaldu* eta *galdu* aditzek ez dute ordain egokirik PB-VN-n, ez direlako ingelesez pieza lexikal bakar batez adierazten, baizik eta *to have lunch* eta *to get lost* “aditz-esapideen” bitartez. Horrelakoetan, ingelesezko “aditz-esapideak” eman ditugu euskarazko aditzaren ordain gisa, eta EADBn agertzen diren argumentuak eta rolak esleitu dizkiegu.

Kapitulua bukatzeko, euskararen beraren ezaugarriak direla eta, euskarazko aditzen inguruan landu ditugun hiru gai aipatu ditugu:

- **HAULetan parte hartzeko joera duten aditzak.** Zenbait kasutan, horrelako aditzek zalantzak sortu dizkigute. HAULa ziurra den kasuetan (Urizar, 2012), ez da arazorik egon, etiketatu gabe utzi dugu eta oharren eremuan “aditz-esapidea” jarri dugu. Baina zalantzazko kasuetan, hots, HAUL izateko bidean egon daitezkeen esapideen kasuan, adibidea etiketatzen saiatu gara eta, gainera, “aditz-esapidea” oharra jarri dugu, etorkizunean gai honen inguruko azterketa sakonagoa egitea erabakitzen dugunean, horrelako adibideak markatuta izatea lagungarri izango baita. Zalantzak sortu dizkiguten hainbat adibide aurkeztu ditugu atal honetan (*eman, jarri, hartu...*). HAULetan parte hartzeko joera handia duten aditzetan, zailtasun handiena adiera berrien eta HAULen arteko mugak bereiztea izan da.
- **Balio semantiko berriak.** Corpus berriak ustiatu ahala, ohikoa da adiera berriak topatzea. Atal honetan, horrelako hainbat adibide aurkeztu ditugu, eta fenomeno horretan elipsiak eta testu berezituak duten erabateko eragina azaldu dugu.
- **Aditz aspektualak.** Aditzaren aspektuari buruzko informazioa ematen duten aditzak etiketatzea zaila izan da ez delako erraza izan horrelakoentzat ingelesezko ordain egokia aurkitzea. Aditz horien agerpenetan “aditz-perifraisi aspektuala” oharra jarri dugu.

Emaitzak

Hasierako helburuari jarraiki, tesi-lan honen emaitza nagusiak bi izan dira:

1. EPEC-RolSem corpora (predikatu-mailan etiketatutako EPEC corpora).
2. *Basque Verb Index* (BVI) lexikoia.

Jarraian, bi baliabide horien gaur egungo egoera zein den azalduko dugu labur.

7.1 EPEC-RolSem: Datuak eta gaur egungo egoera

7.1 taulan etiketatze-prozesuan lortutako emaitzak adierazten ditugu urrats-urats: etiketatutako aditzen eta agerpenen kopurua, erabilitako denbora eta lan-taldea¹³³.

7.1 taularen lehen zutabean, etiketatze-prozesuaren urratsak agertzen dira; 2. zutabean, urrats bakoitzean zenbat pertsona aritu diren etiketatze-lanean; 3. zutabean, urrats horretan landu den aditz-kopurua zehazten da; 4. zutabean, aditz-kopuru horri zenbat agerpen dagozkion agertzen da; 5. zutabean, agerpen horiek corpus osoaren zenbateko portzentajea osatzen duten zehaztu dugu; 6. zutabean, agerpen guztietatik zenbat etiketatu diren

¹³³Azken lerroko batuketa egitean batu behar ez diren zenbakiak grisez eman ditugu.

azaltzen da; 7. zutabean, urrats bakoitza egiteko zenbat denbora behar izan dugun zehazten da, ordutan neurtuta¹³⁴; eta bukatzeko, 8. zutabean, urrats bakoitzean etiketatu den corpusaren portzentajea agertzen da (corpus osoa kontuan hartuta).

Etiketatzean jarraitutako urratsak honako hauek dira:

1. urratsa: Lehen hurbilpenean etiketatutako aditzak.
2. urratsa: Metodologiaren oinarriak finkatzean etiketatutako aditzak (eskuz, 20-40 agerpen).
 2. 1. urratsa: Metodologiaren oinarriak finkatzean etiketatutako aditzak (metodologiaren ebaluazioa).
 2. 2. urratsa: Metodologiaren oinarriak finkatzean etiketatutako aditzak (gainontzeko agerpenak, erdi-automatikoki).
 2. 3. urratsa: [Levinen \(1993\)](#) klaseen baliagarritasuna aztertzean etiketatutako aditzak (erdi-automatikoa).
3. urratsa: Orain arte etiketatu gabe zeuden 30 agerpen baino gehiago dituzten aditzak (guztira 75). Gaur egun etiketatzen ari garen aditzak dira, eta orain arte horietatik 48 landu ditugu.
 3. 1. urratsa: Corpuseko gainontzeko aditz guztiak. 30 agerpen baino gutxiago dituzten aditzak dira eta ez dira eskuz etiketatuko¹³⁵.

¹³⁴Kontuan izanda orduko 13 agerpen etiketatzen ditugula batez beste.

¹³⁵Automatikoki etiketatuko dira garatzen ari garen SRL etiketatzalea erabiliz ([Salaberri et al., 2014](#)), eta gero eskuz errebisatuko da dena.

	Perts.	Ad.- kop.	Ager.- kop.	Corpus osoa %	Etik. ager.	Denbora	Etik. corpusa %
1.	3	3	1.007	3,27	150	11,53	0,48
2.	3	99	17.324	56,35	7.260	558,46	23,61
2.1	2	3	5.017	16,32	350	26,92	1,13
2.2	2	99	17.324	56,35	10.064	774,15	32,73
2.3	1	97	1.917	6,23	1.908	146,76	6,20
3.	1	75	7.105	23,11	3.544	272,61	11,52
3.1	1	940	4.394	14,29	0	0	0
Guztira		1.211	30.740	100	22.776	1.790,43 ¹³⁶	74,06

7.1 taula: Etiketatzeko-prozesuko urratsetan lortutako emaitzak.

7.1 taulan ikusten den bezala, memento honetan corpusaren % 74,06 etiketatua dugu, eta horretarako gutxi gorabehera urtebete behar izan dugu. Esan beharra dago 7.1 taulan etiketatze-lan soilari dagozkion datuak azaltzen direla. Ez dugu hemen sartu etiketatzen hasi aurretik egindako lan guztia: aditzen sarrerak prestatzea, etiketatzeko irizpideak finkatzea, eskuliburu osatzea edota etiketatzeko tresna prestatzea. Era berean, ez dago hemen adierazita prozedura automatikoak prestatzen eta inplementatzen emandako denbora.

7.2 Basque Verb Index (BVI)

EPEC-RolSem corpusa (predikatu-mailan etiketatutako EPEC corpusa) etiketatzeko, aditzen azterketa sakona egin behar izan dugu, eta azterketa horretan oinarrituta, *Basque Verb Index* (BVI) deitu duguna osatu dugu eskuz; hots, PB-VN ereduaren arabera landutako hainbat euskal aditzen lexikoa.

Etiketatzeko jarraitu dugun metodologian zehaztu bezala, aditz bat etiketatzen hasi aurretik, sarrera landu beharra dago. Hori izan da BVI lexikoaren oinarria. Baina ondoren, corpus etiketatua bera baliatuz, lexikoa aberasten joan gara.

Lexikoi honek, alde batetik, aditzaren argumentu-egitura, eta, bestetik, eduki semantikoa jasotzen ditu, eta, gainera, corpusean etiketatutako adibi-deekin lotura du. Aditzen lexikoi gehienetan ez da aditzaren semantika eta

¹³⁶256 egun, lanaldiak 7 ordu dituela kontuan hartuz.

sintaxia batera jasotzen. WordNet-en (Fellbaum, 1998), adibidez, aditzen adiera bakarrik lantzen da, nahiz eta zenbait azpiadiera sintaxi-bereizketaren batek sortua den batzuetan. Comlex-en (Grishman *et al.*, 1994), badira egitura sintaktikoak, baina ez da ezer esaten aditzen adiera edo predikatu orokorrei buruz, eta CoreLex¹³⁷ (Buitelaar, 1998), aldiz, izenetan dago oinarrituta aditzetan baino gehiago. LSC lexikoia (Dorr, 2001) aurrekoek dituzten arazoak gainditzen saiatzen da, baina ez ditu egitura sintaktiko guztiak eta adiera guztiak jasotzen, eta, gainera, ez dago loturarik inolako corpusetako adibide errealekin.

Guztira 244 aditz¹³⁸ eta 364 adiera daude landuta BVIn. 34.000 “postposizio-atzizki+rol” pare daude lexikoian. Horietatik 639 behin bakarrik agertzen dira, eta 475 pare 10 aldiz baino gutxiagotan agertzen dira. 7.2 taulan maizen agertu diren “postposizio-atzizki+rol” pareak ikus daitezke. Rol bakoitzarekin gehien agertzen diren bi postposizio-atzizkiak jarri ditugu taulan. Adibidez, 7.2 taulan *Agent* rolaekin batera gehien agertzen den postposizio-atzizkia ERGa da (3.423), eta hurrengoa ABSa, 165 agerpenekin. Horrelako informazioa oso baliagarria da rolak automatikoki etiketatzeko sistementzat¹³⁹.

¹³⁷ *The Generative Lexicon* (Pustejovsky, 1995) lanean oinarritzen da.

¹³⁸ Memento honetan etiketatzen ari garen 30 agerpenetik gorako aditz guztiak bukatzean, 271 aditz egongo dira BVI lexikoian.

¹³⁹ 7.2 taulan “-” ikurra agertzen da postposizio-atzizkirik ez duen argumentu bat denean.

Postposizio-atzizkia	Rola	Agerpen-kopurua
ERG	Agent	3.423
ABS		165
ABS	Attribute	1.686
-		398
KAUS	Cause	381
MOT		53
ALA	Destination	449
INE		125
ABS	Patient	583
DAT		89
ABS	Product	1.133
PAR		47
INE	Location	508
ALA		129
DAT	Recipient	394
INE		40
ABL	Source	198
DAT		38
ABS	Theme	5.287
KONPL		572
-	ADV	648
INE		13
INE	LOC	1.718
-		131
-	MNR	925
MOD		622
NEG	NEG	1.057
-		26
HELB	PRP	392
ALA_DES		35
-	TMP	1.298
INE		1.144

7.2 taula: "Postposizio-atzizki+rol" pareen agerpen-kopurua EPEC-Rol-Sem corpusean.

Esan beharra daukagu EPEC-DEP corpusean 1.211 aditz desberdin ager-
tzen direla, baina horietatik 151 bakarrik direla 30 agerpen baino gehiago
dituztenak. Maiztasun handieneko aditz guztiak etiketatuta izateko 27 fal-
ta zaizkigu¹⁴⁰, eta horiek dira memento honetan lantzen ari garenak. Hona
hemen sarrera kanoniko horietan definitzen diren eremuak:

- Aditzaren adiera edo predikatu nagusia (1).
- Adiera bakoitzaren PropBank-eko ordaina (2).
- Adiera bakoitzaren argumentu-kopurua: balentzia (3).
- Argumentu bakoitzaren gauzatze sintaktikoak edo postposizio-atziz-
kiak (4).
- Argumentuei esleitzen zaizkien rol semantikoak: VNk ematen duena
(5) eta EADBn agertzen dena (6).
- Zenbaitetan, argumentuen hautapen-murriztapenak (HM) (7)¹⁴¹.

Adibide gisa, 7.3 eta 7.4 tauletan *ekarri* eta *jan* aditzen sarrera kanoni-
koak ikus daitezke.

¹⁴⁰Beste hiru aditz daude EPEC-DEP corpusean 30 agerpen baino gehiago dutenak
(*behar izan*, *ari izan* eta *edin*), baina horiek ez ditugu kontuan hartu.

¹⁴¹HMak ez ditugu sistematikoki markatu aditz guztietan.

ekarri ¹⁴² bring.01 (1 + 2)
Arg0 (3): agent (5), kausa (6) (erg (4))
Arg1(3): theme (5), gai ukitua (6) (abs/par (4))
Arg2 (3): destination (5), helburuko kokapena (6) (ala (4))
Arg3 (3): source (5), abiapuntuko kokapena (6) (abl (4))
ekarri2 cause.01 (1 + 2)
Arg0 (3): theme (5), iturria (6) (erg (4))
Arg1 (3): theme (5), gai sortua (6) (abs/par/konpl (4))
ekarri3 have.03 (1 + 2)
Arg0 (3): theme (5), edukitzailea (6) (erg (4))
Arg1 (3): theme (5), edukia (6) (abs (4))

7.3 taula: *ekarri* aditzaren sarrera BVIn.

jan1 eat.01 (1 + 2)
Arg0 (3): agent (5), esperimentatzailea (6) [+biz] (7) (erg) (4)
Arg1 (3): patient (5), gaia (6) [-biz] (7) (abs) (4)

7.4 taula: *jan* aditzaren sarrera BVIn.

Hori dena ikusgai dago, hurrengo atalean azalduko dugun *e-ROLda* kontsulta-tresnan.

7.3 Laburbilduz

Kapitulu honetan tesi-lanaren 2 ekarpen nagusiak aurkeztu ditugu:

1. **EPEC-RolSem corpora**: Predikatu-mailan etiketatutako EPEC corpora. 244 aditz etiketatu ditugu maiztasun handieneko aditzak barne hartuta, eta horrekin corpusaren % 74,06 etiketatu dugu.
2. **Basque Verb Index (BVI) lexikoa**: Memento honetan 244 aditz sarrera ditu eta 364 adiera. Lexikoian, euskarazko aditz bakoitzaren sarreran, adierak, adiera bakoitzaren PB-VNko ordain egokiak, adiera bakoitzaren argumentu-kopurua eta argumentuei esleitutako rolak

¹⁴²Zenbaki hau euskarazko aditzak BVIn duen adieraren zenbakia da.

(VNkoak eta EADBkoak), argumentu bakoitzaren gauzatze sintaktikoak (postposizio-atzizkiak) eta, zenbaitetan, hautapen-murriztapenak ere jasotzen dira.

e-ROLda kontsulta-tresnaren deskribapena

Corpuseko nahiz lexikoiko informazioa modu errazean kontsultatu, eta BVI zuzendu, aldatu edota moldatu ahal izateko, *e-ROLda* tresna (Estarrona *et al.*, 2014b) garatu dugu¹⁴³. Tresna horrek bi atal ditu: bata publikoa eta bestea pribatua. Atal publikoan corpuseko eta lexikoiko informazioa kontsulta daiteke, eta pribatuan, lexikoia gainean aldaketak egin daitezke, alegia, BVIko sarrerak edita daitezke.

Esan behar dugu tresnak BVI lexikoiko aditz-adiera bakoitzaren PBko, FrameNet-eko eta Euskal WordNet-eko loturak ere eskaintzen dizkigula.

Kapitulu honetan, baliabide horren deskribapen zehatza egingo dugu. Lehenik eta behin, tresnaren ezaugarri teknikoak buruzko zenbait xehetasun emango ditugu. Ondoren, bertan aurki daitezkeen informazio linguistikoa bistaratzeko modua deskribatu eta informazio horren gainean bilaketak nola egiten diren azalduko dugu.

8.1 Ezaugarri teknikoak

Esan bezala, *e-ROLda* kontsulta- eta edizio-tresna bat da. BVI lexikoa eta EPEC-RolSem corpusa kontsultatzeko balio du, eta, lexikoia gainean, aditz-sarreraren edizioa ere ahalbidetzen du.

Oinarrian, teknologia hauek erabiltzen ditu:

¹⁴³<http://ixa2.si.ehu.es/e-rola/index.php>

- XML lengoia: BVI lexikoa formatu honetan dago deskribatuta.
- MySQL datu-base erlazionala: EPEC-RolSem corpuseko datuak gordetzeko, bai eta BVI lexikoiak beste baliabide semantikoekin (PB, FN eta EusWN) dituen loturak gordetzeko ere.
- PHP lengoia: web-aplikazioa implementatzeko (XML eta datu-baseko datuak irakurri eta idazteko).

Tresnak BVI eta EPEC-RolSem corpusean bilaketa konplexuak egiteko aukera ematen du. AJAX teknikak erabili dira EPEC-RolSem corpusaren baitan egindako bilaketen emaitzak erakusteko, interfaze eraginkor eta erabilerraza implementatuz.

8.2 Informazio linguistikoa eta bilaketak

Erabiltzaileak, *e-ROLdan* sartzen denean, tresnan ikusgai dauden baliabideen (BVI lexikoa eta EPEC-RolSem corpusa) gaineko deskribapen laburrak aurkituko ditu, bai eta tresnari berari buruzko oinarritzko informazioa ere.

Tresna horrek gordetako baliabideen gainean bilaketak egiteko aukera ematen digu eta bilaketa horiek eremu desberdinak kontuan izanda egin daitezke:

- Euskarazko aditza
- Euskarazko aditzaren adiera jakin bat
- PB-VNko ingelesezko ordaina

Bilaketa zehatzagoak ere egin daitezke; esaterako, argumentu jakin baten adibideak (Arg0, Arg1...), rol semantiko zehatz bat duten adibideak, postposizio-atzizki jakin bat duten adibideak, bai eta HM jakin bat dutenak ere. Jarraian, adibideak erabiliz, xeheago azalduko ditugu tresnaren eta bilaketen ezaugarri horiek guztiak.

8.2.1 Bilaketa orokorrak

Erabiltzaileak aditz bat bilatzen duenean, hiru motatako informazioa jasotzen du:

1. BVI lexikoian aditz horrentzat dagoen informazioa, hau da, aditzaren BVIko sarrera.
2. Adiera bakoitzeko, PropBank (PB), FrameNet (FN) eta Euskal WordNet (EusWN)¹⁴⁴ baliabideetarako estekak (Laparra eta Rigau, 2010)¹⁴⁵.
3. EPEC-RolSem corpusean agertzen diren aditz horren agerpen etiketatuak adieraka antolatuta¹⁴⁶. Adibide bakoitzeko, informazio hau ematen digu sistemak:
 - Fitxategiaren izena
 - Delako esaldiak fitxategian duen zenbakia
 - Aditza eta adiera-zenbakia
 - PB-VNko ordaina
 - Agerpen jakin horretan agertzen diren argumentu esplizituen kopurua, kontuan hartuta aditz-adiera horrek sarrera kanonikoan dituen argumentuak (adibidez, 2/2k esan nahi du BVIko sarrera kanonikoan bi argumentu daudela, eta adibide jakin horretan biak agertzen direla esplizitu; 1/2k, ordea, BVIko sarreran agertzen diren 2 argumentuetatik, agerpen horretan bakarra agertzen dela adierazten du).
 - Adibidearen esteka

8.1 irudian, erabiltzaileak *jan* aditza bilatzen duenean tresnak bueltatzen duen informazioa ikus dezakegu.

¹⁴⁴e-ROLda tresnan BWN laburtzapenarekin adierazita dago.

¹⁴⁵AnCora (<http://clic.ub.edu/corpus/es/ancora>) proiektuan ere lotu dira gaztelaniazko aditzak PBra, eta hortik, *SemLink* erabilia, baita VN, FN, WN eta OntoNotes-era ere (Taulé *et al.*, 2010). Gauza bera egiten ari dira NewsReader (<http://www.newsreader-project.eu/files/2012/12/NewsReader-316404-D4.2.pdf>) proiektuan ere, baliabideen arteko elkarreagina errazteko eta PB ereduiko SRLa VN, FN eta WN moduko baliabideetako informazio semantikoarekin osatzeko (López de Lacalle *et al.*, 2014).

¹⁴⁶Oraingoz *e-ROLda* tresnan argumentu esplizituak dituzten adibideak baino ez dira erakusten.

Aditza: PB ordaina: Adiera:

bilaketa zehatza: Arg: VN rola: EADB rola: Kasua: HM:

BVI

jan

1 eat_01

arg0: agent, esperimentatzailea [+biz] (erg)

arg1: patient, gaia [-biz] (abs)

eat_01: PB · FN: (Ingestion) · BWN₁ · BWN₂ · BWN₃

EPEC-RolSem

jan 1 eat_01

eabs.451222524.txt sent94 jan 1 eat_01 2/2 ikusi

eabs.450730033.txt sent72 jan 1 eat_01 1/2 ikusi

8.1 irudia: BVIko sarrera, PB, FN eta EusWNekeo estekak, eta EPEC-RolSem corpuseko adibideak *jan* aditzarentzat.

8.1 irudian ikusten den bezala, *jan* aditzak adiera bakarra du BVI lexikoian (*eat.01* PB-VN-n). Aditzak 2 argumentu ditu sarrera kanonikoan:

1. Arg0, *Agent* PB-VN-n, *esperimentatzailea* EADBn, ERGan azaltzen da eta [+biz] HMa du.
2. Arg1, *Patient* PB-VN-n, *gaia* EADBn, ABSan agertzen da eta [-biz] HMa du.

8.1 taulan *jan* aditzaren EPEC-RolSem corpuseko adibide etiketatuen informazio zehatza ikus daiteke. Jarraian agerpen bati buruzko informazioa zehazten dugu (kasu honetan 8.1 taulako hirugarrena da):

- Fitxategiaren izena (eeps.450540276.txt)
- Esaldiak fitxategian duen zenbakia (sent59)
- Euskarazko aditza eta adiera zenbakia (*jan* 1)
- PB-VNko ordaina (*eat.01*)
- Agerpen jakin horretan agertzen diren argumentu esplizituen kopurua, kontuan hartuta aditz-adiera horrek sarrera kanonikoan dituen argumentuak (1/2)
- Adibidearen esteka (*ikusi*)

Fitxategia	Esaldia	Aditza	PB-VN	Arg.-kop.	Esteka
eeps.451222524.txt	sent94	jan 1	eat.01	2/2	ikusi
eeps.450730033.txt	sent72	jan 1	eat.01	1/2	ikusi
eeps.450540276.txt	sent59	jan 1	eat.01	1/2	ikusi
eeps.450530228.txt	sent11	jan 1	eat.01	1/2	ikusi
eeps.450640531.txt	sent11	jan 1	eat.01	1/2	ikusi
eeps.450540276.txt	sent15	jan 1	eat.01	1/2	ikusi

8.1 taula: *jan* aditzaren adibideen zerrenda.

Ikusi dioen estekaren gainean klik eginez gero, adibidea ikusi ahal izango dugu, bere dependentzia-etiketekin eta *arg_info* etiketaren bitartez adierazten den informazio semantikoarekin (ikusi 8.2 irudia).

BVI	
jan	
1 eat_01	
arg0: agent, esperimentatzailea [+biz] (erg)	
arg1: patient, gaia [-biz] (abs)	
eat_01: PB - FNr (Ingestion) - BWN1 - BWN2 - BWN3	
EPEC-RolSem	
jan 1 eat_01	
ebs.451222524.txt	sent94 jan 1 eat_01 2/2 ikusi
ebs.450730033.txt	sent72 jan 1 eat_01 1/2 ikusi
	itxi
ebs.450540276.txt	sent59 jan 1 eat_01 1/2
	<div style="border: 1px solid black; padding: 5px;"> <p>Houdaren familiak asko jaten du.</p> <p>ncmod (gen, familiak-[w706], Houdaren-[w705], Houdaren-[w705]) ncsubj (erg, jaten-[w708], familiak-[w706], familiak-[w706], subj) arg_info (eat_01, jaten-[w708], familiak-[w706], arg0, Agent, esperimentatzailea, +biz) ncmod (-, jaten-[w708], asko-[w707], asko-[w707]) arg_info (eat_01, jaten-[w708], asko-[w707], argM, ADV, -, -) auxmod (-, jaten-[w708], du-[w709])</p> </div>
egun.07-1-p0003000.2000-07-01.kirola	sent4 jan 1 eat_01 1/2 ikusi
ebs.450530228.txt	sent111 jan 1 eat_01 1/2 ikusi

8.2 irudia: *jan* aditzaren adibide baten dependentzia-harremanak eta etiketa semantikoak.

Bilaketa orokorrekin jarraituz, aditz baten adiera jakin bat ere eska dezakegu. 8.3 irudian ikus dezakegu *eman* aditzaren 1. adiera eskatuz gero tresnak bueltatzen duen informazioa.

The screenshot shows the e-ROLda web application interface. At the top, there is a navigation bar with the language 'EN' and the application name 'e-ROLda'. Below this, there is a menu with options: 'Etxera', 'Bilatzaila', 'BVI', 'EPEC-RolSem', 'IXA taldea', 'Kontaktua', and 'Editatu'. The 'Etxera' option is highlighted in green.

Below the menu, there are search filters: 'Aditza: eman', 'PB ordaina: [input]', and 'Adiera: 1'. A search bar contains the text 'bilaketa zehatza: [checkbox] Arg: [dropdown] VN rola: [dropdown] EADB rola: [dropdown] Kasua: [dropdown] HM: [dropdown]'. Below the search bar is a 'Bilatu' button.

The results section is titled 'BVI' and shows the role 'eman'. Below this, there is a list of results for '1 give_01'. The first result is '1 give_01' with the following arguments: 'arg0: agent, abiapuntua (erg)', 'arg1: theme, gaia (abs)', and 'arg2: recipient, helburua (dat)'. Below this, there is a list of roles: 'give_01: PB - FN1 (Giving) - BWN1 - BWN2 - BWN3 - BWN4 - BWN0 - BWN0 - BWN0 - BWN0'.

Below the results, there is a section titled 'EPEC-RolSem' which shows the role 'eman 1 give_01'. Below this, there is a list of roles: 'egun.12-2-p00008002.2000-12-02.politika sent9 eman 1 give_01 3/3 ikusi' and 'egun.06-2-p4202.2000-06-02.kirola sent2 eman 1 give_01 3/3 ikusi'.

8.3 irudia: *eman* aditzaren 1. adiera.

Atal honekin bukatzeko, ingelesezko PB-VNko aditz-adiera jakin bati buruzko bilaketa ere egin dezakegu. 8.4 irudian, PB-VNko *call.01* ordaina eskatuta tresnak bueltatzen duen informazioa ikus dezakegu.

The screenshot shows the e-ROLda web application interface. At the top, there is a navigation bar with the language 'EN' and the application name 'e-ROLda'. Below this, there is a menu with options: 'Etxera', 'Bilatzailea', 'BVI', 'EPEC-RolSem', 'IXA taldea', 'Kontaktua', and 'Editatu'. The main content area features a search form with the following fields: 'Aditza:' (empty), 'PB ordaina:' (filled with 'call_01'), and 'Adiera:' (empty). Below the search form, there is a row of filters: 'bilaketa zehatza:' (checkbox), 'Arg:' (dropdown), 'VN rola:' (dropdown), 'EADB rola:' (dropdown), 'Kasua:' (dropdown), and 'HM:' (dropdown). A 'Bilatu' button is located below the filters. The results are displayed under the heading 'BVI' and are organized into three sections: 'deitu', 'esan', and 'EPEC-RolSem'. The 'deitu' section shows 3 results for 'call_01' with arguments: arg0: agent, abiapuntua [+giz] (erg); arg1: theme, helburua (abs/dat); arg2: predicate, ezaugarria (abs). The 'esan' section shows 2 results for 'call_01' with arguments: arg0: agent, abiapuntua [+giz] (erg); arg1: theme, helburua (dat); arg2: predicate, ezaugarria (abs). The 'EPEC-RolSem' section shows 3 results for 'call_01'. At the bottom, there is a list of files and their corresponding results: 'eebs.450540276.txt' (sent194 deitu 3 call_01 3/3 ikusi), 'eebs.450540276.txt' (sent78 deitu 3 call_01 3/3 ikusi), and 'egun.09-2-p2712.2000-09-02.kirola' (sent3 deitu 3 call_01 2/3 ikusi).

8.4 irudia: *call.01* PB-VNko ordaina eskatuta, *e-ROLdak* ematen duen informazioa.

8.4 irudian ikusten den bezala, *call.01* aditz-adiera euskarazko *deitu* aditzaren hirugarren adieraren eta *esan*en bigarrenaren ordain gisa etiketatu da.

8.2.2 Bilaketa zehatzak

Orain artean, BVI lexikoian aditz baten bilaketa egiteko tresnak eskaintzen dituen aukera orokorrez aritu gara, baina, lehen esan bezala, *e-ROLda* tresnak ezaugarri zehatzagoak ere kontuan hartuz bilaketak egiteko aukera ematen digu. Jarraian, adibideen bitartez azalduko dugu bilaketa zehatz horiek egiteko modua.

Honako bost eremu hauen gaineko bilaketak egin daitezke:

- **Arg:** argumentu zenbakitua
- **VN rola:** VerbNet-eko rola
- **EADB rola:** EADBN agertzen den rola
- **Kasua:** Postposizio-atzizkia
- **HM:** Hautapen-murriztapena

8.5 irudian *eman* aditzaren lehen adierako adibideak ikus ditzakegu, baina soilik datiboaren postposizio-atzizkia dutenak.

The screenshot shows the e-ROLda web interface. At the top, there is a navigation bar with the language 'EN' and the title 'e-ROLda'. Below the title, there are several menu items: 'Ettxera', 'Bilatzailea', 'BVI', 'EPEC-RolSem', 'IXA taldea', 'Kontaktua', and 'Editatu'. The main content area has a search filter section with the following options: 'Aditza: eman', 'PB ordaina: [input]', 'Adiera: 1'. Below this, there is a search criteria section with 'bilaketa zehatza: [checked]', 'Arg: [dropdown]', 'VN rola: [dropdown]', 'EADB rola: [dropdown]', 'Kasua: dat', and 'HM: [dropdown]'. A 'Bilatu' button is located below the search criteria. The results are displayed in two columns, showing a list of text files and their corresponding verb forms and arguments.

egun.06-2-p0003000.2000-06-02.europa	sent10 eman 1 give_01 ikusi
eebs.450622443.txt	sent28 eman 1 give_01 ikusi
eebs.4715027003.txt	sent25 eman 1 give_01 ikusi
egun.11-2-p00033002.2000-11-02.kirola	sent11 eman 1 give_01 ikusi
egun.06-2-p0003200.2000-06-02.mundua	sent7 eman 1 give_01 ikusi
egun.06-2-p3902.2000-06-02.kirola	sent2 eman 1 give_01 ikusi
egun.11-2-p00020001.2000-11-02.europa	sent28 eman 1 give_01 ikusi
egun.06-2-p0003000.2000-06-02.europa	sent12 eman 1 give_01 ikusi
egun.08-1-p2701.2000-08-01.kirola	sent5 eman 1 give_01 ikusi
egun.06-2-p4202.2000-06-02.kirola	sent2 eman 1 give_01 ikusi
egun.06-2-p0003400.2000-06-02.kirola	sent17 eman 1 give_01 ikusi
eebs.450650153.txt	sent87 eman 1 give_01 ikusi
egun.08-1-p0001911.2000-08-01.europa	sent2 eman 1 give_01 ikusi
egun.06-2-p4006.2000-06-02.kirola	sent29 eman 1 give_01 ikusi
eebs.450322049.txt	sent25 eman 1 give_01 ikusi
egun.10-2-p1002.2000-10-03.politika	sent3 eman 1 give_01 ikusi
eebs.450520426.txt	sent18 eman 1 give_01 ikusi
eebs.4716005507.txt	sent19 eman 1 give_01 ikusi
eebs.450520392.txt	sent56 eman 1 give_01 ikusi
eebs.451240686.txt	sent12 eman 1 give_01 ikusi

8.5 irudia: *eman* aditzaren 1. adierako datibodun adibideak.

8.6 irudian, aldiz, *hil* aditzean Arg0 argumentua esplizitu agertzen duten adibideak ikus ditzakegu.

EN e-ROLda

Etxera Bilatzailera BVI EPEC-RolSem IXA taldea Kontaktua Editatu

Aditza: PB ordaina: Adiera:

bilaketa zehatza: Arg: VN rola: EADB rola: Kasua: HM:

Bilatu

egun.07-1-p2501.2000-07-01.mundua sent2 hil 1 kill_01 ikusi
 egun.10-2-p601.2000-10-03.politika sent13 hil 1 kill_01 ikusi

8.6 irudia: *hil* aditzean Arg0 argumentua esplizitu duten adibideak.

Adibideekin jarraituz, 8.7 irudian *ekarri* aditzean Arg1 balentzia eta AB-Sa duten adibideak agertzen dira.

EN e-ROLda

Etxera Bilatzaila BVI EPEC-RolSem IXA taldea Kontaktua Editatu

Aditza: ekarri PB ordaina: Adiera:

bilaketa zehatza: Arg: arg1 VN rola: EADB rola: Kasua: abs HM:

Bilatu

eebs.4716005510.txt	sent24 ekarri 1 bring_01 ikusi
egun.09-2-p0003401.2000-09-02.kirola	sent11 ekarri 1 bring_01 ikusi
eebs.4715028709.txt	sent13 ekarri 1 bring_01 ikusi
egun.06-1-p3701.2000-06-01.kirola	sent7 ekarri 1 bring_01 ikusi
egun.11-1-p00009000.2000-11-01.politika	sent6 ekarri 1 bring_01 ikusi
eebs.451240713.txt	sent135 ekarri 1 bring_01 ikusi
eebs.450640593.txt	sent214 ekarri 1 bring_01 ikusi
eebs.451240713.txt	sent24 ekarri 1 bring_01 ikusi
egun.06-2-p0004000.2000-06-02.kirola	sent27 ekarri 1 bring_01 ikusi
egun.12-2-p00034003.2000-12-02.kirola	sent131 ekarri 1 bring_01 ikusi
eebs.450640531.txt	sent110 ekarri 1 bring_01 ikusi
egun.06-2-p4006.2000-06-02.kirola	sent20 ekarri 1 bring_01 ikusi
epecegun.29.2000-03-01.gizartea1.txt	sent29 ekarri 1 bring_01 ikusi
eebs.451240713.txt	sent183 ekarri 1 bring_01 ikusi
eebs.450740168.txt	sent49 ekarri 1 bring_01 ikusi

8.7 irudia: *ekarri* aditzean Arg1 argumentua eta ABSa duten adibideak.

Eta azkenik, 8.8 irudian, *ekarri* aditzaren 2. adieran ERGan gauzatzen den *Theme* rol semantikoa duten adibideak azaltzen dira.

The screenshot shows the e-ROLda web interface. At the top, there is a navigation bar with the logo 'e-ROLda' and a menu with options: Etxera, Bilatzaillea, BVI, EPEC-RolSem, IXA taldea, Kontaktua, and Editatu. Below the navigation bar, there are search filters: 'Aditza: ekarri', 'PB ordaina: [input]', and 'Adiera: 2'. A search bar contains 'bilaketa zehatza: [input]', 'Arg: [input]', 'VN rola: theme', 'EADB rola: [input]', 'Kasua: erg', and 'HM: [input]'. A 'Bilatu' button is located below the search bar. The results are displayed in a table with two columns: the first column contains file names and the second column contains search results.

eebs.451150179.txt	sent10 ekarri 2 cause_01 ikusi
egun.07-1-p0003400.2000-07-01.kirola	sent7 ekarri 2 cause_01 ikusi
egun.11-1-p00027001.2000-11-01.europa	sent5 ekarri 2 cause_01 ikusi
epecegun.43.2000-05-10.europa1.txt	sent29 ekarri 2 cause_01 ikusi
eebs.450640633.txt	sent5 ekarri 2 cause_01 ikusi
eebs.4716009708.txt	sent18 ekarri 2 cause_01 ikusi
egun.11-1-p00011000.2000-11-01.politika	sent2 ekarri 2 cause_01 ikusi
epecegun.42.2000-03-01.politika10.txt	sent15 ekarri 2 cause_01 ikusi
eebs.450630737.txt	sent3 ekarri 2 cause_01 ikusi
eebs.451230789.txt	sent24 ekarri 2 cause_01 ikusi
egun.08-2-p0002411.2000-08-02.kirola	sent2 ekarri 2 cause_01 ikusi
egun.12-1-p00016007.2000-12-01.politika	sent5 ekarri 2 cause_01 ikusi
egun.06-1-p0002900.2000-06-01.mundua	sent13 ekarri 2 cause_01 ikusi
egun.11-1-p00023002.2000-11-01.ekonomia	sent6 ekarri 2 cause_01 ikusi

8.8 irudia: *ekarri* aditzean *Theme* rola ERGan duten adibi-deak.

Bukatzeko, ezin aipatu gabe utzi, tresnak lexikoia editatzeko aukera ematen digula. Kapitulu honen hasieran aipatu bezala, tresnaren atal pribatuan editorea integratu da, eta, horrela, hizkuntzalariak sarrerak gehitu edo aldatzeko aukera du, modu errazean.

Tresna hau hainbat azterketa linguistiko egin ahal izateko oso baliagarri izan daiteke. Izan ere, tesi-lan honetan garatutako bi baliabideetan gordetzen den aditzaren gaineko informazio sintaktiko-semantiko guztia hizkuntzalarien esku jartzen du.

Ondorioak, zabaldutako ikerlerroak eta etorkizuneko lanak

Tesi-lan hau hasi genuenean, bi helburu nagusi markatu genituen:

1. EPEC corpusa predikatu-mailan eskuz etiketatzea.
2. Euskal aditzen lexikoa sortzea.

7. kapituluan aurkeztu ditugu xehe bi helburu nagusi horien inguruan lortutako emaitza nagusiak. Corpus etiketatua eta lexikoa erdiesteko, tarteko bestelako helburuak ere planteatu genituen, besteak beste, **etiketatzeko metodologia egokia garatzea** eta etiketatze-lana aurrera eramateko irizpide argiak bilduko zituen **eskuliburua** osatzea. Ikerlan honetan, hori guztia erdiesteko jarraitu dugun ibilbidea azaldu dugu, eta bertatik, eta bereziki 6. kapituluan azaldu dugun aditzen azterketa zehatzetik, hainbat ondorio atera ditugu:

- Tesi-lan honetan landutako aditz-kopuruaren, corpuseko aditz-kopuruaren eta aditzek corpusean duten maiztasunen arteko harremana desorekatua da.
- Aditzek HAULetan portaera berezia erakusten dute.
- Aditzek, corpus-motaren arabera, adiera eta portaera sintaktiko desberdinak erakusten dituzte.
- Argumentu eta rol asko eta asko ez dira testuan esplizitu agertzen.

Ondorio horiek azalduko ditugu datozen lau azpiataletan. Eta bukatzeko, 9.5 azpiatalean, gure ustez etorkizunean aztertzeke interesgarriak diren bestelako lan batzuk aipatuko ditugu.

Hasteko, azal ditzagun banan-banan aipatutako lau ondorioak eta horietatik irekitzen diren ikerlerro berriak.

9.1 Aditz-kopuruaren eta aditzen maiztasunen arteko harremana

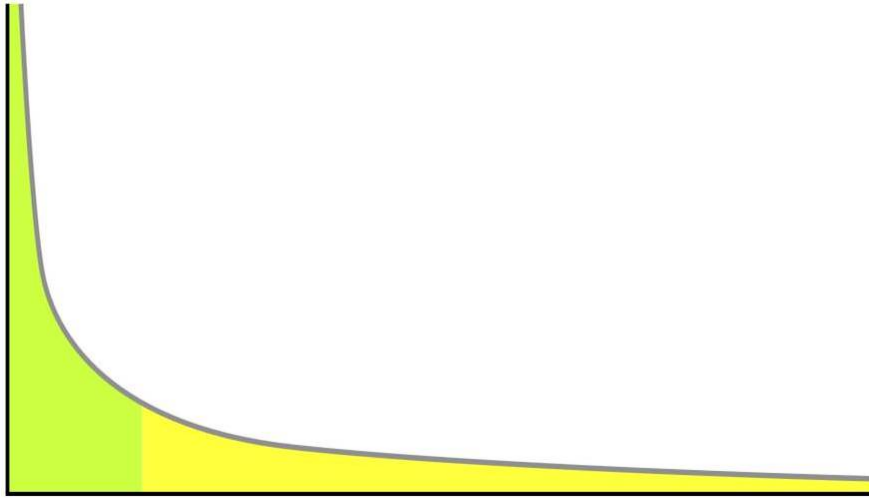
EPEC-DEP corpusean 1.211 aditz desberdin daude, 7.2 atalean esan bezala, eta ikerlan honetan, horietatik 244 aditz landu ditugu. Hasiera batean, eta bestelako azterketarik egin ezean, landutako aditz-kopurua oso txikia dela pentsa genezake, baina kontuan izanda 1.211 aditz horietatik 151 bakarrik direla corpusean 30 aldiz baino gehiagotan agertzen direnak, eta, beraz, 244 horiek landuta corpusaren % 75 barne hartzen dugula, hasierako irudipena aldatu egiten da erabat. Gainera, esan beharra dago aditz-kopuruaren eta maiztasunaren arteko harreman desorekatu hori beste hizkuntzetan ere gertatzen dela; esaterako, PropBank corpusean 3.101 aditz desberdin agertzen dira, baina horietatik 485 aditzek bakarrik dituzte 30 agerpen baino gehiago. EPEC-DEP corpusaren kasuan, 30 agerpen baino gehiago dituzten aditzen portzentajea % 12,71 da eta PBren kasuan, % 15,64¹⁴⁷. Aditzen maiztasunaren banaketa honek Zipfen legea jarraitzen du (Zipf, 1949) eta grafiko batean jarriko bagenu, 9.1 irudian¹⁴⁸ ikusten denaren antzeko zerbait azalduko litzaiguke. Banaketa-mota hori *the long tail* izenez ezagutzen da eta fenomeno aski arrunta da edozein hizkuntzako corpusetan¹⁴⁹.

Maiztasunen gaiarekin bukatzeko, beste datu harrigarri bat aipatu nahiko genuke. EPEC-DEP corpusean agerpen guztien % 43,58 10 aditzi dagokie, eta PBn, aldiz, % 20,67 baino ez. 6.3.1 atalean ikusi dugun bezala, jakina da euskarak zenbait aditz-motatan material gutxiago lexikalizatzen duela

¹⁴⁷Esan beharra dago PropBank corpusak milioi bat hitz dituela, eta, 4.1.2 atalean aipatu bezala, EPEC-DEP corpusak 200.000.

¹⁴⁸Irudia Wikipediatik aterea da: http://en.wikipedia.org/wiki/Long_tail

¹⁴⁹PBn ere aipatzen dute banaketa-mota hori etiketatu behar duten corpusa deskribatzen dutenean: “These verbs are arrayed in a classic Zipfian distribution, with a few verbs occurring very often (*say*, for example, is the most common verb, with over 10,000 instances in its various inflectional forms), and most verbs occurring two or fewer times” (Palmer *et al.*, 2005: 13).



9.1 irudia: *The long tail* fenomenoaren irudikapena.

aditzean (Agirre *et al.*, 2006b). Beharbada hortik etor daitezke aipatu berri ditugun aditzen maiztasunen portzentajeen desberdintasunak, baina ondorio garbiak atera ahal izateko azterketa sakonagoa egin beharko litzateke.

9.2 Hitz Anitzeko Unitate Lexikaletan aditzek erakusten duten portaera berezia

Corpuseko aditzak aztertzean, Hitz Anitzeko Unitate Lexikaletan (HAUL) parte hartzen duten aditzek portaera sintaktiko edo/eta semantiko berezia dutela ikusi dugu. Linguistikoki oso bide interesgarria litzateke HAULetan eta hitz anitzeko esaeretan agertzen diren aditzak sakon aztertzea. 6.3.1 atalean ikusi dugun bezala, aditz batek hainbat adiera edo predikatu orokor izan ditzake, baina aditz hori hitz anitzeko unitate baten partaide denean, bere esanahia edota portaera sintaktikoa berezko dituen predikatu horietatik aldentzen da, esaterako, aditzaren adiera batean beharrezkoa den postposizio-atzizki edo rol bat ezinezkoa denean (DATik ez: *pausoak eman*, *amore eman*), edota kokapenari buruzko adiera batean, inesiboak lekua ez den beste zerbait adierazten duenean (*kolokan jarri*, *zalantzan jarri*). Hala, rol semantikoak HAULetan nola gauzatzen diren aztertzea izango litzateke, tesi-lan honen ondoren, etorkizunerako irekia geratzen den ikerlerroetako bat.

9.3 Aditzek corpus-motaren arabera erakusten duten portaera desberdina

Gure azterketan aurrera egin ahala, argi eta garbi ikusi dugu corpusaren izaerak eragin zuzena duela aditzek erakusten dituzten adiera eta portaera sintaktikoekin. Adibidez, EPEC corpora gehienbat kazetaritza-corpora izanik, ikusi dugu zenbait aditzek kazetaritza-testuetan agertzeko joera handiagoa dutela, edo mota horretako testuetan aditz jakin batek adieraz ditzakeen predikatu orokor guztien artean bakararra baino ez dela agertzen. Hala, gure BVIk corpus berezitetan izan dezakeen baliagarritasuna aztertu nahiko genuke. Oso interesgarria dirudi aditzek testu-motaren arabera erakusten dituzten portaera desberdinak aztertzea, eta horiekin gure BVI aberastea eta linguistikoki koherentea izango den moduan antolatzea.

9.4 Esplizitu agertzen ez diren argumentuak eta rolak

Tesi-lan honetan, testuan modu esplizituan agertzen diren predikatuen parte-hartzaileak aztertu ditugu, baina jakina da beste parte-hartzaile asko modu implizituan edo elidituta agertzen direla testuan.

Corpusa etiketatzean eta argumentuen izaera aztertzean, elipsia behin eta berriz agertu zaigun gaia izan da, eta, 6.3.2 atalean ikusi dugun bezala, elipsi testuala, elipsi pragmatikoa eta infraespezifikazio-elipsia bereizi beharrean aurkitu gara.

Eliditutako edo implizitu dagoen informazioan bi motako entitateak bereizten dira (Palmer *et al.*, 1986): i) elidituta dauden osagai sintaktikoak (adibidez, subjektuak euskaraz: *sagarra jan dut*) eta ii) gauzatu gabe gertatzen diren rol semantikoak (adibidez, infraespezifikazio-elipsia dela eta gauzatzen ez diren rolak: *kupeltxo bat aldatuko dute* [beste berri batez] *zartatuta baitago*). Orain artean, implizitu dagoen informazio hori azalratzea eta etiketatzea korreferentzia-ebazpenaren atal berezi gisa tratatu izan da (Palmer *et al.*, 1986; Whittemore *et al.*, 1991; Tetreault, 2002; Gerber eta Chai, 2010; Silberer eta Frank, 2012). IXA taldean ere hasiak gara argumentu eta rol implizituak automatikoki etiketatzeko bideak aztertzen (Laparra eta Rigau, 2013a, b).

Etorkizunean, hizkuntzalaritzak HPko ataza berri honetan zertan eta nola

lagundu dezakeen aztertzeke asmoa dugu.

Tesi-lan honetatik ateratako ondorio nagusiez eta horietatik irekitako ikerlerro berriez jardun ondoren, etorkizunean egin nahi ditugun beste lan batzuk aipatu nahiko genituzke.

9.5 Etorkizunerako beste lan batzuk

- **Rolen etiketatzailer automatikoa garatzea**

Hasteko, SRL (*Semantic Role Labeling*) alorrean egiten ari garen eta egin behar dugun lana azpimarratu behar dugu. 1.2 atalean azaldu bezala, dagoeneko egin ditugu rol semantikoak automatikoki etiketatzeko lehen urratsak, baina oraindik ere bide luzea geratzen zaigu euskararako baliagarri izango den etiketatzailer automatiko sendo bat lortzeko. Alor horretan egin dugun lehenengo gauza EPEC corpuseko aditz guztiak etiketatzeko balio duen etiketatzailer sortzea izan da (Salaberri *et al.*, 2014). Etiketatzailer hori EPEC-RolSem osorik etiketatzeko erabiliko dugu, hau da, 30 agerpen baino gutxiago dituzten aditzak automatikoki etiketatuko ditugu; ondorengo lana EPEC-RolSem urre-patroi (*gold standard*) gisa erabili ahal izateko, automatikoki etiketatutako gutzia eskuz errebisatzea izango da.

Lehenengo urrats hori beteta, hurrengo edozein testu etiketatzeko gai izango den etiketatzailer sortzea izango da. Hain zuzen, eginkizun horietan ari gara lanean memento honetan.

- **Semantika osoa etiketatzea**

Etorkizunean jorratu beharreko ikerlerroekin jarraitzeko, semantika osoa etiketatzearen gaia aipatu behar dugu. 1.1 atalean azaldu bezala, IXA taldean EPEC corpora tresna automatikoak garatzeko erreferentziazko corpora izan dadin, hainbat mailatan eskuz etiketatzen ari gara, eta beste hizkuntza batzuetan, eta batez ere ingelesean, egin denari jarraiki, predikatu-mailan etiketatu ondoren, egitura semantiko osoa etiketatzea izango litzake hurrengo pausoa; hots, sintaxiaren eremutik at, esaldien esanahia adieraziko duen irudikapen semantikoa lortzea. Bide horretan kokatzen dira Abstract Meaning Representation (AMR, (Banasescu *et al.*, 2013)), Universal Networking Language (UNL, (Uchida *et al.*, 1996)) eta UCCA (A Semantics-based grammatical annotation

scheme, (Abend eta Rappoport, 2013)) ekimenak. Ekimen horietan, esaldien esanahiaren irudikapen abstraktua sortzea da helburua, eta horretarako hainbat egitura semantiko etiketatzen dira eskuz:

- Rol semantikoak
- Aditzen eta izenen argumentuak eta adjuntuak
- Esaldi barruko korreferentzia
- Entitateak
- Modalitatea
- Ukapena
- Galderak
- Kantitateak
- Egitura kopulatiboen semantika
- Perpausen arteko erlazio semantikoak
- Adjektiboak

Orain artean, zerrendan agertzen diren elementuak modu isolatuan etiketatu izan dituzte, eta, sintaxiaren eredua jarraituz, dena batera etiketatzea proposatzen dute orain. Ikerlerro horren helburu nagusia *semantic parsera* lortzea da, hau da, analizatzaile sintaktiko osoak dauden bezala, semantikan ere analizatzaile orokorrak izatea, hizkuntzaren ulermen orokorra hobetzeko asmoz.

- **BVI lexikoia aditz berriekin aberastea**

e-ROLda tresnak eskaintzen digun editatzeko ingurunea aprobetxatuz, BVI lexikoian dauden aditzen kopurua handitzea da gure asmoa. EPEC-RolSem corpora erabat etiketatua dugunean, EPEC-DEP corpuseko aditz guztiak (1.211) BVI lexikoian gehituko ditugu. Horretarako, corpus etiketatutik ateratako informazioa aztertu eta aditzen sarrerak editatu beharko dira eskuz.

- **Rol semantikoaren informazioa analisi-katean txertatzea**

Etorkizunean, lexikoian dagoen informazioa IXA taldearen analisi-katean txertatzeko moduak aztertu nahi ditugu. Tesi-lanean zehar, eta, batez ere 1.2 atalean aipatu dugunez, informazio sintaktiko-semantikoa

analisi-katean gehituz gero, tresnen emaitzak hobetu egiten dira; beraz, ezinbestekoa ikusten dugu txertatze hori katearen zein puntutan eta nola egingo dugun aztertzea.

- **BVIn ditugun WNko loturak EusWN aberasteko erabili**

1.1 atalean ikusi dugun bezala, taldean orain arte Euskal WordNet-en izenak bakarrik landu ditugu. Etorkizun hurbilean, BVIn landutako aditzek WNko *synsetekin* dituzten loturak erabili nahi ditugu Euskal WordNet-eko aditzak lantzeko.

Euskarazko aditz bakoitza WNko *synsetekin* lotuta izateak (PBko ordainari esker), ingelesezko *synset* horietan euskarazko aditza gehitzeko aukera ematen digu. Beraz, Euskal WordNet-eko aditzak osatzeko bidea ematen digu e-ROLDan eskaintzen den baliabideen arteko lotura horrek.

- Bukatzeko, gure tesi-lanak ireki ditzakeen bide berriez den bezainbestean, ezin dugu ahaztu BVI lexikoa sortzeak PropBank, VerbNet, FrameNet eta WordNet baliabideetan agertzen den informazio guztia euskarazko aditzei lotuta izateak abiapuntua ematen digula eredu horien gaineko bestelako azterketak eta esperimenduak egiteko. Lotura horiek oso baliagarriak dira, esaterako, baliabideen arteko elkarreragina errazteko, horrela eredu desberdinetan etiketatutako baliabideak batera erabil daitezkeelako. Hori da NewsReader proiektuan¹⁵⁰ gaztelania, ingelesa, nederlandera eta italierarako egiten ari direna (López de Lacalle *et al.*, 2014), eta etorkizunean, gure baliabideari esker, euskararako ere egin ahal izango duguna.

¹⁵⁰<http://www.newsreader-project.eu/files/2012/12/NewsReader-316404-D4.2.pdf>

Bibliografia

- Abend O. eta Rappoport A. (2013). UCCA: A semantics-based grammatical annotation scheme. *Proceedings of IWCS*.
- Aduriz I. (2000). *EUSMG: Morfologiatik sintaxira Murriztapen Gramatika erabiliz. Euskararen desanbiguazio morfologikoaren tratamendua eta azterketa sintaktikoaren lehen urratsak*. Doktore-tesia, Filologia eta Historia-Geografia Fakultatea, Euskal Herriko Unibertsitatea, Gasteiz.
- Aduriz I., Aranzabe M., Arriola J., Atutxa A., Díaz de Ilarraza A., Ezeiza N., Gojenola K., Oronoz M., Soroa A. eta Urizar R. (2006). Methodology and steps towards the construction of EPEC, a corpus of written Basque tagged at morphological and syntactic levels for the automatic processing. In Wilson A., Rayson P. eta Archer D. editoreak, *Corpus Linguistics Around the World*, 56 lib. in serie *Language and Computers*, 1–15. Rodopi, Holanda.
- Agirre E., Aldezabal I., Etxeberria J. eta Pociello E. (2006a). A Preliminary Study for Building the Basque PropBank. *Proceedings of the 5th International Conference on Language Resources and Evaluations (LREC)*, Genova, Italia.
- Agirre E., Aldezabal I. eta Pociello E. (2006b). Lexicalization and multiword expressions in the Basque WordNet. *Proceedings of Third International WordNet Conference*, Jeju, Korea.

- Aldabe I., Gónzalez I., López-Gazpio I., Madrazo J. eta Maritxalar M. (2013). Two Approaches to Generate Questions in Basque. *Procesamiento del Lenguaje Natural*, 51:101–108.
- Aldezabal I. (1998). Levin’s verb classes and basque. A comparative approach. Oral presentation. UMIACS Departmental Colloquia. University of Maryland.
- Aldezabal I. (2004). *Aditz-azpikategorizazioaren azterketa sintaxi partzialetik sintaxi osorako bidean. 100 aditzen azterketa, Levin-en (1993) lana oinarri hartuta eta metodo automatikoak baliatuz*. Doktore-tesia, Euskal Filologia Saila, Euskal Herriko Unibertsitatea.
- Aldezabal I. (2008). Aditz-azpikategorizazioa. In Artiagoitia X. eta Lakarra J.A. editoreak, *Gramatika Jaietan. Patxi Goenagaren omenez. ASJUren gehigarria*, 29–46. Gipuzkoako Foru Aldundia-EHU, Donostia.
- Aldezabal I. (2010). Basis for the annotation of EPEC-RolSem. *Interdisciplinary Workshop on Verbs. The identification and Representation of Verb Features*, 92–97, Pisa, Italia. Università di Pisa, Dipartimento de Linguistica.
- Aldezabal I., Aranzabe M., Arriola J. eta Díaz de Ilarraza A. (2009a). Syntactic annotation in the Reference Corpus for the Processing of Basque (EPEC): Theoretical and practical issues. *Corpus Linguistics and Linguistic Theory*, 5-2:245–274.
- Aldezabal I., Aranzabe M., Arriola J., Díaz de Ilarraza A., Estarrona A., Fernández K., Iruskieta M. eta Uria L. (2007a). EPEC (Euskararen Prozesamendurako Erreferentzia Corpora) dependentzietan etiketatze esku-liburua. Barne-txostena UPV/EHU/LSI/TR 12-2007, Euskal Herriko Unibertsitatea.
- Aldezabal I., Aranzabe M., Díaz de Ilarraza A. eta Estarrona A. (2010a). Building the Basque PropBank. In Calzolari N., Choukri K., Maegaard B., Mariani J., Odjik J., Piperidis S., Rosner M. eta Tapias D. editoreak, *Proceedings of the Seventh Conference on International Language Resources and Evaluation, LREC 2010*, 1414–1417. European Language Resources Association (ELRA).

- Aldezabal I., Aranzabe M., Díaz de Ilarraza A. eta Estarrona A. (2011). Preliminary evaluation of EPEC-RolSem, a Basque corpus labelled at predicate level. *Procesamiento del Lenguaje Natural*, 47 lib., 1–9. Universidad de Huelva.
- Aldezabal I., Aranzabe M., Díaz de Ilarraza A. eta Estarrona A. (2013). A methodology for the semiautomatic annotation of EPEC-RolSem, a Basque corpus labeled at predicate level following the PropBank-VerbNet model. Barne-txostena UPV/EHU/LSITR01-2013, UPV-EHU, Donostia.
- Aldezabal I., Aranzabe M., Díaz de Ilarraza A., Estarrona A., Ezeiza N. eta Uria L. (2009b). Corpusen etiketatze linguistikoa. *Beñat Oihartzabali Gorazarre. Festschrift for Bernard Oyharçabal. Anuario del Seminario de Filología Vasca Julio de Urquijo*, 37–50. Euskal Herriko Unibertsitatea.
- Aldezabal I., Aranzabe M., Díaz de Ilarraza A., Estarrona A., Fernández K. eta Uria L. (2010b). EPEC-RS: EPEC (Euskararen Prozesamendurako Erreferentzia Corpora) rol semantikoekin etiketatzeko eskuliburua [Guidelines to tag semantic roles in the EPEC corpus (the Reference Corpus for the Processing of Basque)]. Barne-txostena, UPV-EHU.
- Aldezabal I., Aranzabe M., Díaz de Ilarraza A., Estarrona A. eta Uria L. (2010c). EusPropBank: Integrating semantic information in the Basque Dependency Treebank. In Gelbukh A. editorea, *Lecture Notes in Computer Science (LNCS) n^o 6008*, 11th International Conference, CICLing-2010 lib. in serie *Computational Linguistics and Intelligent Text Processing*, 60–73. Springer.
- Aldezabal I., Ceberio K., Esparza I., Estarrona A., Iruskieta M., Izagirre E. eta Uria L. (2007b). EPEC (Euskararen Prozesamendurako Erreferentzia Corpora) segmentazio-mailan etiketatzeko eskuliburua. Barne-txostena UPV/EHU/LSI/TR11-2007, Euskal Herriko Unibertsitatea.
- Aparicio J. (2007). Clasificación semántica de los predicados en español. Doktoretza-ikastaroetako defentsa-lana, Universitat de Barcelona.
- Aparicio J., Taulé M. eta Martí M. (2008). AnCora-Verb: A Lexical Resource for the Semantic Annotation of Corpora. *Proceedings of 6th International Conference on Language Resources and Evaluation (LREC)*, Marrakech, Maroko.

- Aranzabe M. (2008). *Dependentzia-ereduan oinarritutako baliabide sintaktikoak: zuhaitz-bankua eta gramatika konputazionala*. Doktore-tesia, Euskal Filologia Saila. Euskal Herriko Unibertsitatea, Donostia.
- Artiagoitia X. (2002). *Hatsarreak eta Parametroak*. Arabako Foru Aldundia.
- Artola X., Díaz de Ilarraza A., Ezeiza N., Gojenola K., Sologaitoa A. eta Soroa A. (2004). Eulia: a graphical web interface for creating, browsing and editing linguistically annotated corpora. *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC)*, Lisboa, Portugal.
- Babko-Malaya O. (2005). *PropBank Annotation Guidelines*. <http://verbs.colorado.edu/~mpalmer/projects/ace/PBguidelines.pdf>.
- Babko-Malaya O., Bies A., Taylor A., Yi S., Palmer M., Marcus M., Kulick S. eta Shen L. (2006). Issues in synchronizing the English Treebank and PropBank. *Proceedings of the Workshop on Frontiers in Linguistically Annotated Corpora, A Merged Workshop with 7th International Workshop on Linguistically Interpreted Corpora (LINC-2006) and Frontier in Corpus Annotation III, Coling/ACL*, Sydney, Australia.
- Baker C., Fillmore C. eta Lowe J. (1998). The Berkeley FrameNet Project. *Proceedings of COLING-ACL '98*.
- Banarescu L., Bonial C., Cai S., Georgescu M., Griffitt K., Hermjakob U., Knight K., Koehn P., Palmer M. eta Schneider N. (2013). Abstract Meaning Representation for Sembanking. *Proceedings of the 7th Linguistic Annotation Workshop*, Sofia, Bulgaria.
- Bhatt R., Narasimhan B., Palmer M., Rambow O., Sharma D. eta Xia F. (2009). A Multi-Representational and Multi-Layered Treebank for Hindi-Urdu. *Proceedings of the Third Linguistic Annotation Workshop. ACL-IJCNLP*.
- Boas H. (2009). *Multilingual FrameNets: Methods and Applications*. Mouton de Gruyter.
- Bonial C., Brown S., Corvey W., Palmer M., Petukhova V. eta Bunt H. (2011a). Exploratory comparison of thematic roles in VerbNet and LL-RICS. *Proceedings of the Sixth Joint ISO - ACL SIGSEM Workshop on Interoperable Semantic Annotation (ISA-6)*.

- Bonial C., Corvey W., Palmer M., Petukhova V. eta Bunt H. (2011b). A Hierarchical Unification of LIRICS and VerbNet Semantic Roles. *Proceedings of the ICSC Workshop on Semantic Annotation for Computational Linguistic Resources (SACL-ICSC 2011)*.
- Buitelaar P. (1998). CoreLex: An Ontology of Systematic Polysemous Classes. *Proceedings of Formal Ontology in Information Systems, FOIS'98*, Trento, Italy.
- Bunt H., Petukhova V. eta Schiffrin A. (2007). Lirics deliverable d4.4. multilingual test suites for semantically annotated data. Barne-txostena, <http://lirics.loria.fr>.
- Bunt H. eta Romary L. (2002). Towards multimodal content representation. *International standards of terminology and language resources management (LREC)*, Las Palmas, Espainia.
- Burchardt A., Erk K., Frank A., Kowalsky A. eta Padó S. (2009). *Multilingual FrameNets in Computational Lexicography: Methods and Applications*, chapter FrameNet for the semantic analysis of German: Annotation, representation and automation, 209–244. De Gruyter Mouton, Berlin/New York.
- Carletta J. (1996). Assessing agreement on classification tasks: The Kappa statistic. *Computational Linguistics*, 22(2):249–254.
- Castellón I., Fernández A., Vázquez G., Alonso L. eta Capilla J. (2006). The Sensem Corpus: a Corpus Annotated at the Syntactic and Semantic Level. *Fifth International Conference on Language Resources and Evaluation (LREC)*, Génova, Italy.
- Chomsky N. (1957). *Syntactic Structures*. Mouton de Gruyter, Berlin/New York.
- Chomsky N. (1965). *Aspects of the Theory of Syntax*. MIT Press, Cambridge (Massachusetts).
- Civit M., Aldezabal I., Pociello E., Taulé M., Aparicio J. eta Márquez L. (2005). 3LBLEX: léxico verbal con frames sintáctico-semánticos. *XXXI Congreso de la SEPLN*, Granada, Spain.

- Díaz de Ilarraza A., Garmendia A. eta Oronoz M. (2004). Abar-hitz: An annotation tool for the basque dependency treebank. *Proceedings of the fourth international conference on Language Resources and Evaluation (LREC)*, 251–254, Lisboa, Portugal.
- Dorr B. (2001). LSC Verb Database, Online Software Database of Lexical Conceptual Structures and Documentation. Barne-txostena, University of Maryland.
- Dowty D. (1991). Thematic protores and argument selection. *Language*, 67:547–619.
- Duran M. eta Aluísio S. (2012). Propbank-Br: a Brazilian treebank annotated with semantic role labels. *8th International Conference on Language Resources and Evaluation (LREC)*, 1862–1867, Istanbul, Turkia.
- Elhuyar (2000). *Euskal Hiztegi Modernoa*. Elhuyar. Elkarlanean, Donostia.
- Elhuyar (2006). *Elhuyar Hiztegia. Euskara/Gaztelania - Castellano/Vasco*. Elhuyar, Usurbil.
- Estarrona A., Aldezabal I., Díaz de Ilarraza A. eta Aranzabe M. (2014a). EPEC-RolSem: Ingelesezko PropBank-VerbNet eredura etiketatutako euskarazko corpusa. erabakiak, egokitzapenak eta berezitasunak. In Gomez R. eta Ezeizabarrena M. editoreak, *Eridenen du zerzaz kontenta. Sailkideen omenaldia Henrike Knörr irakasleari (1947-2008)*. UPV/EHuren Argitalpen Zerbitzua (argitaratzeke), Bilbo.
- Estarrona A., Aldezabal I., Díaz de Ilarraza A. eta Aranzabe M. (2014b). Methodology of construction of the corpus-based Basque Verb Index (BVI) Lexicon. *Language Resources and Evaluation*, (submitted).
- Etxepare R. (2003). Valency and argument structure in the Basque verb. In Hualde J.I. eta Ortiz de Urbina J. editoreak, *A Grammar of Basque*, 363–426. Mouton de Gruyter, Berlin / New York.
- Euskaltzaindia (1987). *Euskal Gramatika. Lehen Urratsak - II*. Euskaltzaindia.
- Fellbaum C. (1998). *WordNet, An Electronic Lexical Database*. MIT Press, Cambridge.

- Fernández M. eta Anula R. (1995). *Sintaxis y cognición. Introducción al conocimiento, el procesamiento y los déficits sintácticos*. Síntesis.
- Fillmore C. (1968). The case for case. In Bach E. eta Harms R. editoreak, *Universals in Linguistic Theory*, 1–88. Holt, Rinehart and Winston, New York.
- Fillmore C. (1976). Frame semantics and the nature of language. *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language*, 280:20–32.
- Fillmore C. eta Baker C. (2001). FrameNet: Frame semantics meets the corpus. *Proceedings of WordNet on Other Lexical Resources Workshop*, Pittsburgh.
- Friberg K. eta Toporowska M. (2012). The Rocky Road towards a Swedish FrameNet. *Proceedings of the Eighth conference on International Language Resources and Evaluation (LREC)*, Istanbul, Turkia.
- García-Miguel J. eta Albertuz F. (2005). Verbs, Semantic Classes and Semantic Roles in the ADESSE Project. *Proceedings of the Interdisciplinary Workshop on the Identification and Representation of Verb Features and Verb Classes*, Saarbrücken, Alemania.
- Gardent C. eta Cerisara C. (2010). Semi-Automatic Propbanking for French. *Proceedings of the ninth international workshop on Treebanks and Linguistic Theories*, Tartu, Estonia.
- Gerber M. eta Chai J. (2010). Beyond NomBank: A Study of Implicit Arguments for Nominal Predicates. *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 1583–1592, Uppsala, Suecia.
- Gràcia L., Azkarate M., Cabré M. eta Varela S. (2000). *Configuración morfológica y estructura argumental: léxico y diccionario*. Euskal Herriko Unibertsitatea.
- Grimshaw J. (1979). Complement Selection and the Lexicon. *Linguistic Inquiry*, 10:279–326.

- Grishman F., Macleod C. eta Meyers A. (1994). Complex Syntax: building a computational lexicon. *Proceedings of the 15th conference on Computational linguistics, COLING'94*, 1 lib., 268–272, Kyoto, Japonia.
- Gruber J. (1965). *Studies in Lexical Relations*. Bloomington: Indiana University Linguistics Club.
- Hajic J. (1998). *Building a Syntactically Annotated Corpus: The Prague Dependency Treebank*, 106–132. Charles University Press.
- Hajic J., Panevová J., Uresová Z., Bémová A., Kolárová V. eta Pajas P. (2003). PDT-VALLEX: Creating a Large-coverage Valency Lexicon for Treebank Annotation. *Proceedings of the Second Workshop on Treebanks and Linguistic Theories*, 57–68, Vaxjo, Suedia.
- Hajicova E. (2000). Dependency-based underlying-structure tagging of a very large czech corpus. *T.A.L.*, 41:47–66.
- Hajicova E., Panevová J. eta Sgall P. (2000). A manual for tectogrammatic tagging of the prague dependency treebank. Barne-txostena, Charles University, Txekiar Errepublikak.
- Hualde J.I. eta Ortiz de Urbina J. (2003). *A Grammar of Basque*. Mouton de Gruyter, Berlin/New York.
- Ibarretxe-Antuñano I. (2001). An overview of Basque locational cases: Old descriptions, new approaches. Barne-txostena, University of California at Berkeley.
- Ibarretxe-Antuñano I. (2003). Semantic motivation in Basque locational cases. *Euskal gramatikari eta literaturari buruzko ikerketak XXI. mendaren atarian. Gramatika gaiak, Iker-14*. Euskaltzaindia.
- Jackendoff R. (1972). *Semantic Interpretation in Generative Grammar*. MIT Press, Cambridge (Massachusetts).
- Jaka A. (2009). Mugimenduzko ekintzak ingelesez eta euskaraz, Sarrionandiaren itzulpen baten azterketatik abiatuta. *Uztaro*, 69:53–76.
- Karlsson F. (2006). *Constraint Grammar - A Language-Independent System for Parsing Unrestricted Text*. Mouton de Gruyter, Berlin/New York.

- Kingsbury P. eta Palmer M. (2002). From Treebank to PropBank. *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC)*, Las Palmas, Espainia.
- Kingsbury P. eta Palmer M. (2003). PropBank: The Next Level of Treebank. *Proceedings of Treebanks and Lexical Theories*, Vaxjo, Suedia.
- Kipper K. (2005). *VerbNet: A Broad-coverage, Comprehensive Verb Lexicon*. Doktore-tesia, University of Pennsylvania.
- Kipper K., Dang H. eta Palmer M. (2000). Class-based construction of a verb lexicon. *Seventeenth National Conference on Artificial Intelligence*, Austin, Texas.
- Kipper K., Korhonen A., Ryant N. eta Palmer M. (2006). Extending VerbNet with novel classes. *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC)*, Genova, Italia.
- Kipper K., Korhonen A., Ryant N. eta Palmer M. (2008). A large-scale classification of english verbs. *Language Resources and Evaluation*, 1:21–40.
- Laparra E. eta Rigau G. (2010). eXtended WordFrameNet. *Proceeding of the 7th International Conference on Language Resources and Evaluation (LREC)*, Valletta, Malta.
- Laparra E. eta Rigau G. (2013a). Impar: A deterministic algorithm for implicit semantic role labelling. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL)*, 33–41, Sofia, Bulgaria.
- Laparra E. eta Rigau G. (2013b). Sources of Evidence for Implicit Argument Resolution. *Proceedings of 10th International Conference on Computational Semantics (IWCS)*, 155–166, Potsdam, Alemania.
- Levin B. (1993). *English Verb Classes and Alternations. A Preliminary Investigation*. The University of Chicago Press, Chicago/Londres.
- Levin B. (2009). Lexical semantics of verbs VI: Assessing semantic determinants of argument realization.

- Loper E., Yi S. eta Palmer M. (2007). Combining Lexical Resources: Mapping Between PropBank and VerbNet. *Proceedings of the Seventh International Workshop on Computational Semantics (IWCS-7)*, Tilburg, Alemania.
- López de Lacalle M., Laparra E. eta Rigau G. (2014). First steps towards a Predicate Matrix. *Proceedings of the 7th Global WordNet Conference (GWC)*, Tartu, Estonia.
- Marcus M. (1994). The Penn Treebank: A revised corpus design for extracting predicate argument structure. *Proceedings of the ARPA Human Language Technology Workshop*.
- Marcus M., Kim G., Marcinkiewicz M., MacIntyre R., Bies A., Ferguson M., Katz K. eta Schasberger B. (1994). The Penn Treebank: Annotating predicate argument structure. *Proceedings of the ARPA Human Language Technology Workshop*.
- Marcus M., Santorini B. eta Marcinkiewicz M. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19 (2):313–330.
- Merlo P. eta Van der Plas L. (2009). Abstraction and Generalisation in Semantic Role Labels: PropBank, VerbNet or both? *Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP*, Suntec, Singapur.
- Miller G. (1985). WordNet: A dictionary browser. *Proceedings of the First International Conference on Information in Data*, Waterloo, Kanada.
- Monachesi P., Stevens G. eta Trapman J. (2007). Adding semantic role annotation to a corpus of written Dutch. *Proceedings of the Linguistic Annotation Workshop*, 77–84, Praga, Txekiar Errepublika.
- Narayanan S. eta Harabagiu S. (2004). Question answering based on semantic structures. *Proceedings of the 20th COLING*. Association for Computational Linguistics.
- Ohara K. (2012). Semantic Annotations in Japanese FrameNet. *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC)*, Istanbul, Turkia.

- Oregi J. (1975). Euskal aditzaren sailkapen jator baten bila. *Euskera*, 20:359–372.
- Padó S. (2007). *Cross-lingual Annotation Projection Models for Role-Semantic Information*. Doktore-tesia, Saarland University.
- Palmer M. (2009). SemLink: Linking PropBank, VerbNet and FrameNet. *Proceedings of the Generative Lexicon Conference*, Pisa, Italia.
- Palmer M., Babko-Malaya O., Bies A., Diab M., Maamouri M., Mansouri A. eta Zaghouni W. (2008). A pilot arabic propbank. *Proceedings of the sixth International Conference on Language Resources and Evaluation (LREC)*, Marrakech, Maroko.
- Palmer M., Dahl D., Schiffman R., Hirschman L., Linebarger M. eta Dowding J. (1986). Recovering implicit information. *Proceedings of the 24th annual meeting on Association for Computational Linguistics*, 10–19. Association for Computational Linguistics.
- Palmer M., Gildea D. eta Kingsbury P. (2005). The Proposition Bank: A Corpus Annotated with Semantic Roles. *Computational Linguistics*, 31:1:71–106.
- Palmer M., Ryu S., Choi J., Yoon S. eta Jeon Y. (2006). *Korean PropBank*. Linguistic Data Consortium, Philadelphia.
- Petukhova V. eta Bunt H. (2008). LIRICS semantic role annotation: Design and evaluation of a set of data categories. *Proceedings of the sixth International Conference on Language Resources and Evaluation (LREC)*, Marrakech, Maroko.
- Pociello E. (2007). *Euskararen ezagutza-base lexikala: Euskal WordNet*. Doktore-tesia, Euskal Filologia Saila. Euskal Herriko Unibertsitatea.
- Pociello E., Agirre E. eta Aldezabal I. (2011). Methodology and Construction of the Basque WordNet. *Language Resources and Evaluation*, 45:2:121–142.
- Ponzetto S. eta Strube M. (2006). Exploiting Semantic Role Labeling, WordNet and Wikipedia for Coreference Resolution. *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*.

- Pradhan S., Hovy E., Marcus M., Palmer M., Ramshaw L. eta Weischedel R. (2007). OntoNotes: A Unified Relational Semantic Representation. *International Journal of Semantic Computing*, 1(4):405–419.
- Pustejovsky J. (1995). *The Generative Lexicon*. MIT Press, Cambridge.
- Sagarna A. (2007). Euskara eta informazioaren teknologiak. egungo egoeratik etorkizunera begira. *Euskera*, 49-1:43–55.
- Salaberri H., Arregi O. eta Zafirain B. (2014). First approach toward Semantic Role Labeling for Basque. *Proceedings of the 9th Language Resources and Evaluation Conference (LREC)*, Reykjavik, Islandia.
- Salomão M. (2009). Framenet brasil: um trabalho em progresso. *Calidoscópio*, 7(3):171–182.
- Sarasola I. (1984-1995). *Hauta Lanerako Euskal Hiztegia*. Gipuzkoako Aurrezki Kutxa Probintziala. Kutxa Fundazioa, Donostia.
- Schiffrin A. eta Bunt H. (2007). Lirics deliverable d4.3. document compilation of semantic data categories. Barne-txostena, <http://lirics.loria.fr>.
- Shen D. eta Lapata M. (2007). Using Semantic Roles to Improve Question Answering. *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, Praga, Txekiar Errepublika.
- Siegel S. eta Castellan N. (1988). *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill, New York, second edition.
- Silberer C. eta Frank A. (2012). Casting implicit role linking as an anaphora resolution task. *SEM2012: The First Joint Conference on Lexical and Computational Semantics*, 1–10, Montreal, Canada.
- Subirats C. eta Petruck M. (2003). Surprise: Spanish FrameNet. *International Congress of Linguists. Workshop on Frame Semantics*, Praga, Txekiar Errepublika.
- Talmy L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In Shopen T. editorea, *Language typology and syntactic description: Vol 3. Grammatical categories and the lexicon*, 3 lib., 36–149. Cambridge University Press, Cambridge.

- Taulé M., Castellví J., Martí M. eta Aparicio J. (2006). Fundamentos teóricos y metodológicos para el etiquetado semántico de CESS-CAT y CESS-ESP. *XXII Congreso de la SEPLN*, Zaragoza, España.
- Taulé M., Martí M. eta Borrega O. (2010). AnCorra-Net: Mapping the Spanish Ancora-Verb lexicon to VerbNet. *Workshop on Verbs. The Identification and Representation of Verb Features*, Pisa, Italia.
- Taulé M., Martí M. eta Borrega O. (2011). Ancora 2.0: Argument structure guidelines for catalan and spanish. Barne-txostena, Centre de Llenguatge y Computació (CLiC). Universitat de Barcelona.
- Taulé M., Martí M. eta Recasens M. (2008). Ancora: Multilevel Annotated Corpora for Catalan and Spanish. *Proceedings of 6th International Conference on Language Resources and Evaluation*, Marrakech, Maroko.
- Tesnière L. (1959). *Éléments de Syntaxe Structurale*. Paris.
- Tetreault J. (2002). Implicit Role Reference. *International Symposium on Reference Resolution for Natural Language Processing*, 109–115, Alicante, España.
- Uchida H., Zhu M. eta Della Senta T. (1996). UNL: Universal Networking Language - an electronic language for communication, understanding and collaboration. Barne-txostena, IAS/UNU, Tokyo, Japonia.
- Urizar R. (2012). *Euskal lokuzioen tratamendu konputazionala*. Dokortesia, Euskal Herriko Unibertsitatea, Donostia.
- Urkia M. (2008). Euskararen erreferentzia-copusaren beharraz. *Euskalgintza XXI. mendeari buruz. XV. Biltzarra. Azkue eta Urkixoren omenez, Bilbon, 2001-9-17/19an*, Iker 19:307–312.
- Van Der Plas L., Samardžić T. eta Merlo P. (2010). Cross-lingual validity of PropBank in the manual annotation of French. *Proceedings of the Fourth Linguistic Annotation Workshop (LAW IV '10)*, 113–117, Uppsala, Suedia.
- Van Valin R. (1993). *Advances in Role and Reference Grammar*. John Benjamin Publishing Company.

- Vázquez G., Alonso L., Capilla J., Castellón I. eta Fernández A. (2006). Sensem: Sentidos verbales, semántica oracional y anotación de corpus. *Procesamiento del Lenguaje Natural*, 37:113–120.
- Vázquez G., Fernández A. eta Martí M. (2000). *Clasificación Verbal. Alternancias de diátesis*. Quaderns de Sintagma 3. Edicions de la Universitat de Lleida, Lleida.
- Whittemore G., Macpherson M. eta Carlson G. (1991). Event-building through Role-filling and Anaphora Resolution. In Appelt D.E. editorea, *29th Annual Meeting of the Association for Computational Linguistics (ACL)*, 17–24. University of California (Berkeley).
- Wiecheteck L. eta Arriola J. (2011). An Experiment of Use and Reuse of Verb Valency in Morphosyntactic Disambiguation and Machine Translation of Euskara and North Sámi. *Constraint Grammar Workshop at NODALIDA 2011. NEALT Proceedings*, 14 lib. Tartu University Library.
- Xue N. (2008). Labeling Chinese predicates with semantic roles. *Computational Linguistics*, 34(2):225–255.
- Xue N. eta Palmer M. (2003). Annotating propositions in the Penn Chinese Treebank. *The Proceedings of the 2nd SIGHAN Workshop on Chinese Language Processing*, Sapporo, Japonia.
- Xue N. eta Palmer M. (2009). Adding semantic roles to the Chinese Treebank. *Natural Language Engineering*, 15(1):143–172.
- You L. eta Liu K. (2005). Building Chinese FrameNet database. *Proceedings of 2005 IEEE International Conference on Natural Language Processing and Knowledge Engineering*, 301–306, Wuhan, Txina.
- Zabala I. (2004). Los predicados complejos en vasco. In Pérez Gaztelu E., Zabala I. eta Gràcia L. editoreak, *Las fronteras de la composición en lenguas románicas y en vasco*, 445–534. Deustuko Unibertsitatea, Donostia.
- Zipf G. (1949). *Human behavior and the principle of least effort*. Addison-Wesley Press.

A

BVI lexikoa

Aditza	Adiera	PB-VN ordaina	Balantzia	VN rola	EADB rola	Postposizio-atzizkia	HM
abestu	1	sing.01	Arg0 Arg1 Arg2	agent topic recipient	esperimentatzailea gaia	erg abs dat	+giz -biz/+konkr
abiatu	1	leave.01	Arg0 Arg1 Arg2	theme source destination	gai ukitua abiapuntuko kokapena helburuko kokapena esperimentatzailea	abs abl ala/abz	
adierazi	1	state.01	Arg0 Arg1 Arg2	agent topic recipient	esperimentatzailea gaia	erg abs/konpl dat	-biz/-konkr
afaldu	1	express.01	Arg0 Arg1 Arg2	agent theme recipient	esperimentatzailea gaia	erg abs/konpl dat	-biz/-konkr
agertu	1	dine.01	Arg0 Arg1 Arg2	agent patient theme	esperimentatzailea gaia	erg abs abs/par	+biz -biz/+konkr
	1	appear.01/merge.01	Arg1 Arg2	theme location	kokapena	abs/par ine	
	2	appear.02	Arg1 Arg2	theme predicate	gaia egoera	abs abs/soz/mod	
	3	show.01	Arg0 Arg1 Arg2	agent topic recipient	esperimentatzailea gaia	erg abs dat	
agindu	1	order.01	Arg0 Arg1 Arg2	agent patient proposition	esperimentatzailea gai ukitua gaia	erg dat dat abs	
	2	promise.01	Arg0 Arg1 Arg2	agent recipient theme	esperimentatzailea helburua gaia	erg dat abs/konpl	
ahaztu	1	forget.01	Arg0 Arg1	agent theme	esperimentatzailea gaia	erg/abs/dat abs/ins/konpl/gen/soz	
alpatu	1	cite.01	Arg0 Arg1 Arg2	agent topic recipient	esperimentatzailea gaia	erg abs/konpl dat	
aitortu	1	avow.01	Arg0 Arg1 Arg2	agent topic recipient	esperimentatzailea gaia	erg abs/konpl dat	
akabatu	1	kill.01	Arg0 Arg1 Arg2	agent patient instrument	kausa gai ukitua	erg abs/gen ine/soz/ins	+biz
	2	finish.01	Arg0 Arg1	agent theme		erg abs/soz	
aldatu	1	alter.01/change.01	Arg0 Arg1 Arg2 Arg3	agent patient product material	kausa gai ukitua helburua/kokapena/egoera abiapuntua	erg abs/par ala/ins abl	
	2	replace.01	Arg0 Arg1 Arg2	agent theme theme	esperimentatzailea gaia gaia	erg abs ins	
amaitu	1	end.01/finish.01	Arg0 Arg1 Arg2	agent theme instrument	kausa gai ukitua	erg abs/soz soz/ins	
	2	end.03	Arg1	theme	gai ukitua	erg	

Jarraipena hurrengo orrialdean

Aditza	Adiera	PB-VN ordaina	Balantzia	VN rola	EADB rola	Postposizio-atzizkia	HM
antolatatu	1	organize.01	Arg2	attribute	helburuko kokapena	abs/ine/	
			Arg0	agent		erg	
			Arg1	product		abs	
			Arg2	material		soz/ins	
antzman	1	recognize.01	Arg3	beneficiary		des	
			Arg4	attribute		-	
			Arg0	agent	esperimentatzailea	erg	
			Arg1	theme	gaia	abs/konpl	
arakatu	1	scrutinize.01	Arg0	experimenter	esperimentatzailea	erg	
			Arg1	stimulus	gaia	abs	
			Arg0	agent	abiapuntua	erg	
			Arg1	theme	helburua	abs	
arbuia	1	scorn.01	Arg2	predicate		mot	
			Arg0	cause	kausa	erg/soz/ins	
			Arg1	experimenter	gai ukitua	abs	
			Arg0	agent	kausa	erg	
arduratu	1	preoccupy.01	Arg0	patient/source	kausa	abs/par	
			Arg1	theme	gai ukitua	abs	
			Arg0	agent	esperimentatzailea	abs	
			Arg1	theme	jarduera	ine/mod	
argitu	1	light.02/clear.04	Arg0	product	esperimentatzailea	erg	
			Arg1	agent	gaia	abs	
			Arg0	agent	esperimentatzailea	erg	
			Arg1	theme	gaia	abs/konpl	
aritu	2	dawn.02	Arg0	agent	esperimentatzailea	erg	
			Arg1	theme	gaia	abs	
			Arg0	agent	kausa	erg	
			Arg1	theme	gaia	abs/konpl	
asmatu	1	invent.01	Arg0	agent	esperimentatzailea	erg	
			Arg1	product	gaia	abs	
			Arg0	agent	esperimentatzailea	erg	
			Arg1	theme	gaia	abs/konpl	
atera	1	come.out.01/extract.01/take.out.01/take.01	Arg0	agent	esperimentatzailea	erg	
			Arg1	theme	gaia	abs/ine/soz	
			Arg2	agent	kausa	erg	
			Arg3	source	gaia	abs	
atxilotu	2	come.09/emerge.01/get.01/publish.01	Arg0	destination	abiapuntuko kokapena	abl/dat	
			Arg1	theme	helburuko kokapena	ala	
			Arg0	location	gai sortua	abs	
			Arg1	location	iturria	erg/abl/dat	
aukeratu	1	choose.01	Arg0	agent		erg	
			Arg1	patient		abs	
			Arg2	predicate		mot	
			Arg0	agent		erg	
aurkeztu	1	present.01/submit.01	Arg1	theme		abs/par	
			Arg2	beneficiary		abl/-en artean	
			Arg3	attribute		des	
			Arg4	attribute		abs/gisa/-t(z)eko/-rako	
antolatatu	2	introduce.01	Arg0	agent		erg	
			Arg1	theme		abs	
			Arg2	recipient		dat/ine	
			Arg0	agent		erg	
antzman	1	file.01	Arg1	patient		abs	
			Arg2	patient		dat	
			Arg0	agent		erg	
			Arg1	theme		abs	
arduratu	3	stand.02	Arg2	location		ine/dat	
			Arg3	recipient		dat/en aurka	
			Arg0	agent		erg	
			Arg1	theme		abs	

Jarraipena hurrengo orrialdean

Aditza	Adiera	PB-VN ordaina	Balantzia	VN rola	EADB rola	Postposizio-atzizkia	HM
aurkitu	1	find.01	Arg2	destination		ala	
			Arg0 Arg1 Arg2	agent theme beneficiary	esperimentatzailea gaia kokapena	erg abs/par ine/dat	+giz
auzipetu	2	feel.02/be.01	Arg0 Arg2	agent theme predicate	esperimentatzailea gaia egoera	erg abs abs/ine	
	1	prosecute.01	Arg0 Arg1 Arg2	agent theme predicate	abiapuntua helburua gaia	erg abs mot	+giz +giz -biz
azaldu	1	explain.01/state.01/demonstrate.01	Arg0 Arg1 Arg2	agent topic recipient	esperimentatzailea gaia	erg abs/konpl dat	
	2	appear.01	Arg1 Arg2	theme location	gaia kokapena	abs ine/ala	
azpimarratu	3	appear.02	Arg1 Arg2	theme predicate	gaia egoera	abs abs/soz	
	1	underline .01	Arg0 Arg1	agent theme		erg abs/konpl	
aztertu	1	examine.01/study.01	Arg0 Arg1	experiencer stimulus		erg abs/konpl	
babestu	1	protect.01	Arg0 Arg1 Arg2	agent theme theme		erg abs abl	
	2	sponsor.01	Arg0 Arg1	agent theme		erg abs	
baieztatu	1	confirm.01	Arg0 Arg1 Arg2	agent predicate theme	esperimentatzailea gaia	erg abs/konpl dat	-biz/-konkr
	1	separate.01/divide.02	Arg0 Arg1 Arg2	agent patient patient	kausa gai ukitua gai ukitua	erg abs abl	
banatu	1	distribute.01	Arg0 Arg1 Arg2	agent theme recipient		erg abs dat/ine/soz	
	2	divide.02	Arg1 Arg2	agent patient patient		erg abs ine	
barkatu	3	separate.01	Arg0 Arg1 Arg2	agent theme source		erg abs abl	
	1	forgive.01	Arg0 Arg1 Arg2	agent theme predicate	abiapuntua helburua gaia	erg dat/abs abs	+giz +giz -biz
bazkaldu	1	have.lunch.01	Arg0 Arg1	agent patient	esperimentatzailea gaia	erg abs	+biz +konkr
	1	bless.01	Arg0 Arg1 Arg2	agent theme predicate	abiapuntua helburua gaia	erg abs mot	+giz +giz -biz
begiratu	1	look.01	Arg0	experiencer		erg	

Jarraipena hurrengo orrialdean

Aditza	Adiera	PB-VN ordaina	Balantzia	VN rola	EADB rola	Postposizio-atzizkia	HM
behatu	1	observe.01	Arg1	stimulus	experimentatzailea	abs/konpl/dat	
			Arg0	experiencer	gaia	erg	+giz
bereizi	1	split.01/distinguish.01	Arg0	agent	kausa	erg	
			Arg1	patient	gai ukitua	abs	
berotu	1	heat.01/wallop.01/warm.02	Arg0	agent	experimentatzailea	erg	
			Arg1	patient	gaia/helburua	abs	
besar-katu	1	hug.01/embrace.02	Arg0	instrument		soz//ins	
			Arg1	theme	experimentatzailea	erg	+giz
bete	1	fill.01	Arg0	theme	experimentatzailea	abs	+giz
			Arg1	agent	kausa	erg	
bidali	1	send.01	Arg0	destination	helburuko kokapena	abs	
			Arg2	theme	ins	abs	
bilakatu	2	fulfill.01/obey.01	Arg0	agent	experimentatzailea	erg	
			Arg1	theme	gaia	abs	-konkr
bilakatu	1	send.01	Arg0	agent		erg	
			Arg1	theme	erg	abs	
bilatu	2	expel.01	Arg0	destination	dat/ala/helb	erg	
			Arg1	agent		abs	
bilakatu	1	convert.01/become.01/change.01	Arg0	theme		erg	
			Arg2	source		abs	
bilatu	1	look.05	Arg0	agent	kausa	erg	
			Arg1	patient	gai ukitua	abs	
bildu	1	meet.03	Arg0	product	helburuko egoera	abs	
			Arg2	theme	experimentatzailea	erg	
bilakatu	2	compile.01/collect.01	Arg0	source	gaia	abs/konpl/par	
			Arg1	beneficiary		abl/dat	
bisitatu	3	wrap.01/envelop.01/bandage.01	Arg0	agent		abs	
			Arg1	destination		soz//ins	
bombardatu	4	join.01	Arg0	theme		erg	
			Arg1	agent		abs	
bota	5	bend.01	Arg0	patient		soz//ala	
			Arg2	patient		erg	
bota	6	gather.01	Arg0	patient		abs	
			Arg1	agent		erg	
bisitatu	1	visit.01	Arg0	theme		abs	+giz
			Arg1	actor	gai ukitua	erg	
bombardatu	1	bombard.01	Arg0	actor	helburuko kokapena	abs	
			Arg1	agent	experimentatzailea	erg	+giz
bota	1	throw.01	Arg0	theme	gaia	abs	
			Arg1	agent		erg	

Jarraipena hurrengo orrialdean

Aditza	Adiera	PB-VN ordaina	Balantzia	VN rola	EADB rola	Postposizio-atzizkia	HM
	2	expel.04	Arg3 Arg0 Arg1 Arg2	destination agent theme source		ala/dat erg abs abl	
	3	throw.07	Arg0 Arg1	agent theme		erg abs	
	4	throw.04	Arg0 Arg1 Arg2	agent topic recipient		erg abs dat	
	5	put.01	Arg0 Arg1 Arg2	agent theme location		erg abs ine/abl	
	6	knock.02	Arg0 Arg1 Arg2 Arg3 Arg4	agent theme source destination instrument		erg abs abl ala soz/ins	
	7	kick.01	Arg0 Arg1	agent theme		erg abs	
bukatu	1	end.01/finish.01	Arg0 Arg1 Arg2	agent theme instrument		erg abs/soz soz/ins	
	2	end.03	Arg1 Arg2	theme attribute		erg abs/ine	
bultzatu	1	push.01/push.02	Arg0 Arg1 Arg2	agent patient destination	kausa gaia helburuko kokapena	erg abs ala	
	2	promofe.02	Arg0 Arg1	agent theme	kausa gaia	erg abs	
dantzatu	1	dance.01	Arg0 Arg1 Arg2	agent theme theme	esperimentatzailea gaia gaia	erg abs soz	+giz
dastatu	1	savor.01	Arg0 Arg1	experientcer stimulus	esperimentatzailea gaia	erg abs	+giz
dedikatu	1	spend.02	Arg0 Arg1 Arg2	agent theme recipient	abiapuntua gaia/denbora helburua	erg abs dat/ala	
deitu	1	call.02	Arg0 Arg1 Arg2	agent theme location	esperimentatzailea helburua jarduera	erg abs abs/ala/par	
	2	call.02	Arg0 Arg1 Arg2 Arg3	agent beneficiary location location	abiapuntua helburua kokapena kokapena	erg dat ala abl	+giz +giz
	3	call.01	Arg0 Arg1 Arg2	agent theme predicate	abiapuntua helburua ezaugarria	erg abs/soz abs	+giz
deklaratu	1	declare.01	Arg0 Arg1 Arg2	agent topic recipient	esperimentatzailea gaia	erg abs/kompl dat	
desagertu	1	disappear.01	Arg0	theme	gai ukitua	abs	Jarraitzen jarraitzen orrialdean

Aditza	Adiera	PB-VN ordaina	Balantzia	VN rola	EADB rola	Postposizio-atzizkia	HM
diseinatu	1	design.01	Arg1	source		abl	
			Arg0 Arg1 Arg2	agent product beneficiary	esperimentatzailea gaia	erg abs dat/des	
ebaki	1	cut.01	Arg0 Arg1 Arg2 Arg3	agent patient patient instrument	esperimentatzailea gaia	erg abs dat soz/ins	
			Arg0	agent	esperimentatzailea gaia	erg abs/konpl	
			Arg1 Arg0 Arg1 Arg2	theme agent patient source	esperimentatzailea gaia	erg abs abl	+biz -biz
egiaztatu	1	prove.01	Arg0 Arg1 Arg2	agent topic recipient	esperimentatzailea gaia	erg abs/konpl dat	
			Arg0 Arg1 Arg2 Arg3	agent product instrument beneficiary	iturria gai sortua	erg abs ins dat/des	-biz
egin	1	make.01/build.01/construct.01	Arg0	agent	iturria	erg	
			Arg1 Arg2	product instrument beneficiary	gai sortua	abs ins//soz dat	
			Arg0 Arg1 Arg2	agent topic recipient	iturria gai sortua	erg abs dat	
			Arg0 Arg1 Arg2	agent product beneficiary	iturria gai sortua	erg abs dat/des	
			Arg0 Arg1 Arg2	agent theme instrument	iturria gai sortua	erg abs ins//soz	
			Arg0 Arg1 Arg2	agent product beneficiary	iturria gai sortua	erg abs dat/des	
			Arg0 Arg1 Arg2	agent theme instrument	iturria gaia	erg abs ins//soz	
			Arg0 Arg1 Arg2	agent product beneficiary	iturria gai sortua	erg abs dat/des	
			Arg0 Arg1 Arg2	agent asset goal	iturria gai sortua	erg abs dat	
			Arg0 Arg1 Arg2	agent theme predicate	esperimentatzailea neurria jarduera	erg abs ine	
			Arg0 Arg1 Arg2	agent theme source	esperimentatzailea gaia ezaugarria	erg abs abs	
			Arg0 Arg1 Arg2	agent source destination	gai ukiftua kokapena helburuko kokapena	erg abl ala/abu	

Jarraitzen jarraitzen

Aditza	Adiera	PB-VN ordaina	Balantzia	VN rola	EADB rola	Postposizio-atzizkia	HM
egokitu	1	adapt.01	Arg0 Arg1 Arg2	agent patient destination	kausa gai ukitua helburuko egoera	erg abs ala/dat/des abs	
	2	happen.02	Arg1	theme	gaia	abs	
egon	1	be.02	Arg0 Arg1	theme location	gaia kokapena	abs/par ine/ala/abl	
	2	be.01	Arg0 Arg1	topic attribute	gaia egoera	abs/erg abs/erg	
	3	correspond.02	Arg0 Arg1	theme theme	gaia helburua	abs/abl/ala/dat/emen/espl/ins/mod abs dat	
ehizatu	1	hunt.01	Arg0 Arg1	agent theme	esperimentatzailea gaia	erg abs abs	+biz +biz
	1	bring.01	Arg0 Arg1 Arg2 Arg3	agent theme destination source	kausa gai ukitua helburuko kokapena abiapuntuoko kokapena	erg abs/par ala abl	
elkartu	2	cause.01	Arg0 Arg1	theme theme	iturria gai sortua	erg abs/par/konpl	
	3	have.03	Arg0 Arg1	theme theme	edukitzailea edukia	erg abs	
	1	meet.03/join.01	Arg0 Arg1 Arg2	actor patient	kausa gai ukitua	erg abs dat/soz	
eman	2	unite.01	Arg0 Arg1 Arg2	agent patient patient	kausa gai ukitua helburuko egoera	erg abs ine	
	1	give.01	Arg0 Arg1 Arg2	agent theme recipient	agent abiapuntua gaia helburua	erg abs abs dat	
	2	spend.02	Arg0 Arg1 Arg2	agent theme recipient	esperimentatzailea denbora jarduera	erg abs abs ine	
entrenatu	3	overlook.02	Arg0 Arg1	agent destination	esperimentatzailea helburua	erg ala	
	1	coach.01/train.01	Arg0 Arg1	agent beneficiary	agent beneficiary	erg abs	
entzun	1	hear.01/listen.01	Arg0 Arg1 Arg2	agent theme source	esperimentatzailea gaia	erg abs/konpl/par abl/dat	
	1	judge.01	Arg0 Arg1 Arg2	agent theme predicate	esperimentatzailea gaia	erg abs mot	
erabaki	1	decide.01	Arg0 Arg1 Arg2	agent theme beneficiary	esperimentatzailea gaia helburuko kokapena	erg abs/konpl des	
	1	use.01	Arg0 Arg1 Arg2	agent theme predicate	esperimentatzailea gaia	erg abs helb	
eragin	1	cause.01	Arg0 Arg1	theme theme	iturria gai sortua	erg abs/par	-konkr
	2	affect.01	Arg0	cause	kausa	erg	

Jarraipena hurrengo orrialdean

Aditza	Adiera	PB-VN ordaina	Balantzia	VN rola	EADB rola	Postposizio-atzizkia	HM
	3	activate.01	Arg1	experiencier	iturria	dat/ine	+biz
			Arg2	instrument		ins/soz	
			Arg0	agent		erg	
erail	1	murder.01	Arg1	patient		dat	+biz
			Arg0	agent	kausa gai ukitua	erg abs/gen ine/soz/ins	
erakutsi	1	show.01	Arg0	agent		erg	
			Arg1	topic		abs/par/konpl/t(z)en	
eraman	1	transport.01	Arg2	recipient		dat	
			Arg0	agent	kausa gai ukitua	erg	
	2	carry.01	Arg1	theme	helburuko kokapena	abs ala/abu	
			Arg2	destination	abiapuntuko kokapena	abl	
	3	spend.02	Arg0	agent	edukitzailea	erg	
			Arg1	theme	edukia	abs	
	4	win.01	Arg2	location	kokapena	ine	
			Arg0	agent	esperimentatzailea	erg	
	5	stand.03	Arg1	theme	denbora	abs	
			Arg2	recipient	jarduera	ine	
erantsi	1	glue.01	Arg0	Theme	gai.	erg	
			Arg1	experiencier	esperimentatzailea	abs	
erantzun	2	add.01	Arg0	theme	gai.	erg	
			Arg1	agent	esperimentatzailea	abs	
erosi	1	buy.01	Arg0	agent	esperimentatzailea	erg	
			Arg1	topic	gaia	abs/konpl	
erran	1	say.01/tell.01	Arg2	recipient		dat	-konkr
			Arg0	agent	kausa gai ukitua	erg abs/ine dat mot/ine des/dat	
erre	2	burn.01	Arg1	theme	gaia	abs/konpl	
			Arg2	recipient	gaia	dat -ri buruz	
erreparatu	1	notice.01	Arg0	agent	esperimentatzailea	erg	+giz
			Arg1	theme	kausa gai ukitua	abs	
esan	1	say.01/tell.01	Arg0	patient	esperimentatzailea	erg	+giz -konkr
			Arg1	experiencier	gaia	abs/konpl/ine/dat	
			Arg2	stimulus	gaia	erg	
			Arg0	agent	esperimentatzailea	erg	
			Arg1	topic	gaia	abs/konpl	
			Arg2	recipient		dat	

Jarraipena hurrengo orrialdean

Aditza	Adiera	PB-VN ordaina	Balantzia	VN rola	EADB rola	Postposizio-atzizkia	HM
	2	call.01	Arg3 Arg0 Arg1 Arg2	attribute agent theme predicate	abiapuntua helburua ezaugarria	-ri buruz/ins erg dat abs	+giz
eskaini	1	offer.01	Arg0 Arg1 Arg2	agent theme recipient	abiapuntua gaia helburua	erg abs/par dat/des	
		dedicate.01	Arg0 Arg1 Arg2	agent theme proposition	abiapuntua gaia helburua	erg abs/par dat	
eskatu	1	ask.02	Arg0 Arg1 Arg2	agent proposition patient	esperimentatzailea gaia	erg abs/par/konpl dat	
		order.02	Arg0 Arg1 Arg2 Arg3	agent theme beneficiary source	esperimentatzailea gaia	erg abs/konpl/gen des dat	
		demand.01	Arg0 Arg1 Arg2	agent proposition patient	esperimentatzailea gaia	erg abs/konpl dat	
	2	require.01	Arg0 Arg1 Arg2	theme theme source	esperimentatzailea gaia	erg abs ine	
eskertu	1	thank.01	Arg0 Arg1 Arg2	agent theme predicate	abiapuntua helburua gaia	erg/dat abs abs/mot	+giz -biz
	1	give.01	Arg0 Arg1 Arg2	agent theme recipient		erg abs dat	
etorri	2	obtain.01/receive.01/get.01	Arg0 Arg1 Arg2	agent theme recipient		erg abs abl	
	1	come.01	Arg0 Arg1 Arg2	theme source destination	gai ukitua abiapuntuko kokapena helburuko kokapena	abs abl ala	
	2	come.03	Arg0 Arg1	theme source	gai sortua iturria	abs abl/dat	-konkr +biz
	3	be.02	Arg0 Arg1	theme location	edukia edukitzailea	abs ine	-biz
etzan	4	be.01	Arg0 Arg1	topic attribute	gaia ezaugarria	abs abs	
	1	lay.01	Arg0 Arg1 Arg2	agent theme destination	kausua gai ukitua helburuko kokapena	erg abs ine	
eutsi	2	lie.01	Arg1 Arg2	theme location	helburuko kokapena	abs ine	
	1	hold.01/hold.03/hold.07	Arg0 Arg1	agent theme	esperimentatzailea gaia	erg abs/dat	-biz

Jarraitzen duen orrialdean

Aditza	Adiera	PB-VN ordaina	Balantzia	VN rola	EADB rola	Postposizio-atzizkia	HM
existitu	1	exist.01	Arg0	theme	gaia	abs/erg	
ezagutu	1	recognize.02	Arg0 Arg1	agent theme	esperimentatzailea gaia	erg abs/konpl	+biz
ezarri	1	lay.01	Arg0 Arg1 Arg2	agent theme destination	kausa gai ukitua helburuko kokapena	erg abs ine	
ezkondu	1	marry.01/wed.01	Arg0 Arg1 Arg2	agent patient patient	kausa gai ukitua gai ukitua	erg abs soz	+giz +giz +giz
fixatu	1	sign.03	Arg0 Arg1 Arg2	agent theme attribute		erg abs/par abs	+giz +giz
flotatu	1	float.01	Arg0	theme	gaia	erg	
frogatu	1	prove.01	Arg0 Arg1 Arg2	location agent topic	kokapena esperimentatzailea gaia	ine erg abs/konpl dat	
gainditu	1	overcome.01/defeat.01/exceed.01/pass.07/pass.02	Arg0 Arg1	recipient theme	esperimentatzailea	erg abs/dat	
gaineratu	1	add.02	Arg0 Arg1 Arg2	agent patient patient	gaia kausa gai ukitua gai ukitua	erg abs/konpl dat	
gaitzetsi	1	censure.01	Arg0 Arg1 Arg2	agent theme predicate	esperimentatzailea helburua	erg abs/konpl/dat mot	
galdu	1	spoil.01/corrupt.01	Arg0 Arg1	agent patient		erg abs	
	2	lose.02/miss.02	Arg0 Arg1 Arg2	agent theme recipient		erg abs	
	3	get.lost.01/disappear.01 develop.02	Arg1	theme		abs	
garatu	1		Arg0 Arg1 Arg2 Arg3	agent product material beneficiary	kausa gai ukitua	erg abs soz/ins dat/des	
gehitu	1	add.02	Arg0 Arg1 Arg2	agent patient patient	kausa gai ukitua gai ukitua	erg abs/konpl dat	
geratu	1	stop.01	Arg0 Arg1 Arg2	agent theme instrument		erg abs soz/ins	
	2	stay.01	Arg1 Arg2	theme location		abs ine/ala	
	3	be.01	Arg1 Arg2	topic attribute		abs abs/ine/soz	
	4	remain.01	Arg1	theme		abs	
	5	meet.03	Arg0 Arg1	agent agent		abs soz	
gertatu	1	happen.01/occur.01	Arg0 Arg1	theme location	gaia kokapena	abs/par ine/soz	
	2	be.01	Arg0	topic	gaia	abs	Jarraipena hurrengo orrialdean

Aditza	Adiera	PB-VN ordaina	Balantzia	VN rola	EADB rola	Postposizio-atzizkia	HM
gidatu	1	guide.01	Arg1	attribute	egoera	abs/ine	
			Arg0 Arg1 Arg2 Arg3	agent theme destination source	kausa gai ukitua helburuko kokapena abiapuntuko kokapena	erg abs ala abl	
gogoratu	1	remember.01	Arg0	agent		erg/abs	
			Arg1 Arg2	theme predicate		abs/soz/ins/konpl gisa/bezala	
gogoratu	2	remind.01	Arg0	agent		erg	
			Arg1 Arg2	topic recipient		abs/konpl dat	
gogoratu	1	praise.01/laud.01	Arg0	agent	abiapuntua	erg	+giz
			Arg1 Arg2	theme predicate	helburua gaia	abs/dat abs/mot	+giz -biz
gosaldu	1	breakfast.01	Arg0	agent	esperimentatzailea	erg	
			Arg1	patient		abs	
gozatu	1	savor.01	Arg0	agent	esperimentatzailea	erg	+giz
			Arg1	theme	gaia	abs/ine/soz/ins	
grabatu	1	tape.02	Arg0	agent	esperimentatzailea	erg	+biz
			Arg1	theme	gaia	abs abs abs	-biz
gustatu	1	like.01	Arg0	experiencer		dat	
			Arg1	stimulus		abs/konpl	
gutxietsi	2	like.02	Arg0	pivot		dat	
			Arg1	theme		abs/konpl	
gutxietsi	1	disparage.01	Arg0	agent	abiapuntua	erg	+giz
			Arg1 Arg2	theme predicate	helburua gaia	abs mot	+giz -biz
gutxitu	1	decrease.01/diminish.01	Arg0	agent	gai ukitua	erg	
			Arg1 Arg2	patient instrument		abs soz/ins	
handitu	1	swell.01	Arg0	agent	gai ukitua	erg	
			Arg1	patient		abs	
harilkatu	1	wind.01	Arg0	agent		erg	
			Arg1 Arg2	theme location		abs ine	
hartu	1	take.01	Arg0	agent	kausa	erg	
			Arg1 Arg2	theme source	gai ukitua abiapuntuko kokapena	abs abl	
take.01	2	take.01	Arg0	agent	edukitzailea	erg	
			Arg1 Arg2	theme location	edukia kokapena	abs ine	
take.02/tolerate.01	3	take.02/tolerate.01	Arg0	experiencer	esperimentatzailea	erg	
			Arg1	cause	gaia	abs	
haserretu	1	anger.01	Arg0	cause	kausa	soz/erg	
			Arg1	experiencer	gai ukitua	abs	+giz
hasi	1	begin.01/start.01	Arg0	agent	kausa	erg	+giz
			Arg1 Arg2	theme instrument	gai ukitua	abs ins/soz	-biz
begin.01/start.01	2	begin.01/start.01	Arg0	agent	esperimentatzailea	abs	
			Arg1	theme	jarduera	ine/soz	

Jarraipena hurrengo orrialdean

Aditza	Adiera	PB-VN ordaina	Balantzia	VN rola	EADB rola	Postposizio-atzizkia	HM
haustu	1	molder.01	Arg2 Arg0 Arg1	instrument agent patient		ins/soz erg abs	
hautatu	1	select.01	Arg0 Arg1 Arg2 Arg3	agent theme source attribute	gai ukitua esperimentatzailea gaia	erg abs abl abs	-biz/ +konkr
		elect.01	Arg0 Arg1 Arg2	agent theme predicate	esperimentatzailea gaia	erg abs abs	
hauteman	1	perceive.01	Arg0 Arg1	experiencer stimulus	esperimentatzailea gaia	erg abs/konpl	+giz
hautsi	1	break.01	Arg0 Arg1 Arg2	agent patient instrument	kausa gai ukitua	erg abs soz/ins	
		break.01/fracture.01	Arg0	experiencer patient	gai ukitua gai ukitua	erg abs	
hazi	1	grow.01/increase.01	Arg0 Arg1 Arg2	agent patient extent	kausa gai ukitua	erg abs abs	
		rear.01	Arg0 Arg1	agent patient	kausa gai ukitua	erg abs	
		grow.03	Arg0 Arg1 Arg2	agent product material	kausa gai ukitua	erg abs ins/soz	
haztatu	1	paw.01	Arg0 Arg1 Arg2	agent experiencer instrument	esperimentatzailea gaia	erg abs soz/ins	
herdoildu	1	rust.01	Arg0 Arg1	agent patient	gai ukitua	erg abs	-biz/ +konkr
hil	1	kill.01	Arg0 Arg1 Arg2	agent patient instrument	kausa gai ukitua	erg abs soz/ine/ins	
	2	die.01	Arg1 Arg2	theme instrument	gai ukitua	abs soz/ine/ins	
hondatu	1	deteriorate.01	Arg0 Arg1	agent patient	gai ukitua	erg abs	
hornitu	1	furnish.01/provide.01/supply.01	Arg0 Arg1 Arg2	agent theme recipient	kausa gai ukitua helburuko kokapena	erg ins abs	
hurbildu	1	approach.01/bring.closer.01	Arg0 Arg1 Arg2 Arg3	agent theme destination source		erg abs ala/dat abl	
hustu	1	empty.01	Arg0 Arg1 Arg2	agent source theme	kausa abiapuntuko kokapena/gai ukitua	erg abs ins/abs	
ibili	1	walk.01/flow.01/be.02/travel.01	Arg1 Arg2	theme location		abs/erg abs/ala/abu/abz	
	2	be.01	Arg1 Arg2	topic attribute		abs abs/ine/ins	Jarraipena hurrengo orrialdean

Aditza	Adiera	PB-VN ordaina	Balentzia	VN rola	EADB rola	Postposizio-atzizkia	HM
	3	work.09	Arg1	theme		abs	
	4	go.22	Arg1 Arg2	theme theme		abs soz	
	5	wear.01	Arg0 Arg1	agent theme		erg abs	
	6	use.01	Arg0 Arg1 Arg2	agent theme predicate		erg abs -rako	
	7	play.01	Arg0 Arg1	agent theme		abs ine/ins	
	8	ride.01	Arg0 Arg1	agent theme		abs ins	
idatzi	1	write.01	Arg0 Arg1 Arg2	agent theme destination	esperimentatzailea gaia helburua	erg abs dat	
igarri	1	notice.01	Arg0 Arg1	experiencer stimulus	esperimentatzailea gaia	erg abs	+giz
igo	1	rise.01/raise.01	Arg0 Arg1 Arg2 Arg3 Arg4	agent theme/patient extent source location	kausa gai ukitua abiapuntua helburua ala	erg abs abs abl ala	
		climb.01/board.01/ascend.01	Arg1 Arg2	theme location	gai ukitua helburua	abs ala/abu	
		promote.01	Arg0 Arg1 Arg2 Arg3	agent theme destination source	kausa gai ukitua helburua abiapuntuko kokapena	erg abs ala abl	
		wither.01	Arg0 Arg1	agent patient	gai ukitua patient	erg abs	
ihartu	1	learn.01/study.01	Arg0 Arg1 Arg2	agent topic source	esperimentatzailea jarduera	erg abs/konpl/ine/mod/par abl	+giz
ikusi	1	see.01	Arg0 Arg1	agent theme	esperimentatzailea gaia	erg abs/konpl/par	
ikuskatu	1	examine.01	Arg0 Arg1	experiencer stimulus	esperimentatzailea gaia	erg abs	+giz
interesatu	1	interest.01	Arg0 Arg1	cause experiencer	kausa gai ukitua	abs dat	+giz
irabazi	1	win.01/earn.01/gain.02/gain.01	Arg0 Arg1 Arg2 Arg3	agent theme beneficiary source		erg abs dat/des abl	
iragarri	1	announce.01	Arg0 Arg1 Arg2	agent topic recipient		erg abs/konpl dat	
	2	predict.01/presage.01	Arg0 Arg1	cause topic		erg abs/konpl	
iraindu	1	insult.01	Arg0 Arg1 Arg2	agent theme predicate	abiapuntua helburua	erg abs mot	+giz +giz

Jarraipena hurrengo orrialdean

Aditza	Adiera	PB-VN ordaina	Balantzia	VN rola	EADB rola	Postposizio-atzizkia	HM
irakin irakurri	1	boil.01	Arg0	patient	gai ukitua	erg	
	1	read.01	Arg0 Arg1 Arg2	agent theme recipient	esperimentatzailea gaia	erg abs/konpl/par dat	+giz -biz
iraun	1	last.01	Arg0 Arg1	theme attribute	esperimentatzailea denbora	erg abs/ins	
	2	linger.01	Arg0 Arg1	agent location	esperimentatzailea egoera	erg ine	
ireki	1	open.01	Arg0 Arg1 Arg2 Arg3	agent patient instrument recipient		erg abs soz/ins dat	
	2	open.02	Arg0 Arg1 Arg2	agent theme location		erg abs ala/abz/abu	
iritsi	1	arrive.01	Arg1 Arg2 Arg3	theme source location	gai ukitua abiapuntuko kokapena helburuko kokapena	abs abl ala/abu/abz	
	2	reach.01	Arg0 Arg1	agent theme	gai ukitua helburua	erg abs	
irten	1	leave.01/come.03	Arg0 Arg1	theme source	gai ukitua abiapuntuko kokapena	abs/erg abl	
	2	come.09	Arg0 Arg1	theme attribute		abs abs	
isildu	1	silencie.01	Arg0	cause	kausa	erg	
isuri	1	flow.01	Arg0 Arg1	patient location	gai ukitua iturria	abs erg/abl/ala	-biz
	2	flow.01/pour.01/fill.01	Arg0 Arg1 Arg2 Arg3	theme agent theme source location	gai sortua kausa gai ukitua abiapuntuko kokapena helburuko kokapena	abs erg abs abl ala	+biz -biz
itsatsi	1	gum.01	Arg0 Arg1 Arg2	agent patient patient	kausa gai ukitua gai ukitua	erg abs dat/soz	
	1	close.01	Arg0 Arg1 Arg2	agent patient instrument	kausa gai ukitua	erg abs soz/ins	
itzuli	1	return.01	Arg1 Arg2 Arg3	theme source destination		abs abl ala	
	2	return.02	Arg0 Arg1 Arg2	agent theme recipient		erg abs dat/ala	
izan	3	translate.01	Arg0 Arg1 Arg2 Arg3	agent patient product material		erg abs ala soz/ins	
	1	be.02	Arg0 Arg1	theme location	gaia kokapena	abs ine	
	2	be.01	Arg0	topic	gaia	abs/konpl	

Jarraipena hurrengo orrialdean

Aditza	Adiera	PB-VN ordaina	Balentzia	VN rola	EADB rola	Postposizio-atzizkia	HM
	3	have.03	Arg1 Arg0 Arg1	attribute theme theme	ezaugarria edukitzailea edukia	abs erg abs	
izendatu	1	mention.01	Arg0 Arg1 Arg2	agent topic recipient		erg abs/konpl dat	
	2	appoint.01/name.01	Arg0 Arg1 Arg2	agent theme attribute		erg abs abs/par/pro	
izoztu	1	freeze.01	Arg0 Arg1	agent patient	kausa gai ukitua	erg abs	
	1	descend.01/lower.01	Arg0 Arg1 Arg2 Arg3 Arg4	agent theme extent source location	kausa gai ukitua neurria abiapuntua helburua	erg abs abs abl ala/abu	
jaitsi	1	fall.01	Arg1 Arg2 Arg3 Arg4	patient extent source location	gai ukitua neurria abiapuntua helburua	abs abs abl ala	
	1	know.01	Arg0 Arg1 Arg2	agent theme predicate		erg abs/konpl/t(z)en ins/ri buruz	
jakinarazi	1	notify.01/inform.01	Arg0 Arg1 Arg2	agent topic recipient		erg abs/konpl dat	
	1	eat.01	Arg0 Arg1	agent patient	esperimentatzailea gaia	erg abs	+biz -biz
jarri	1	put.01	Arg0 Arg1 Arg2	agent theme destination	kausa gai ukitua helburuko kokapena	erg abs ine/abl	
	2	become.01/get.03	Arg0 Arg1 Arg2	agent destination patient	kausa helburuko egoera	erg abs abs/ine	
	3	give.01/put.01	Arg0 Arg1 Arg2	agent theme recipient	abiapuntua ezaugarria helburua	erg abs dat	+giz
jasan	1	stand.03/undergo.01/suffer.01	Arg0 Arg1	experimenter cause	esperimentatzailea gaia	erg/dat abs/par	
	1	lift.01/raise.01	Arg0 Arg1 Arg2 Arg3	agent theme source destination	kausa gai ukitua abiapuntuko kokapena helburuko kokapena	erg abs abl ala	
jaso	2	receive.01	Arg0 Arg1 Arg2	agent theme source	helburua gaia abiapuntua	erg abs abl	+giz -biz
	3	include.01	Arg0 Arg1 Arg2	agent theme location	edukitzailea edukia kokapena	erg abs ine	
		erect.01	Arg0 Arg1	agent product	esperimentatzailea gaia	erg abs	

Jarraipena hurrengo orrialdean

Aditza	Adiera	PB-VN ordaina	Balantzia	VN rola	EADB rola	Postposizio-atzizkia	HM
jazo	1	occur.01	Arg0 Arg1	theme location	gaia kokapena	abs ine	
jo	1	play.01/ring.01	Arg0 Arg1 Arg2	agent theme instrument	esperimentatzailea gaia	erg abs ins/soz	-biz
	2	hit.01/strike.01/knock.01	Arg0 Arg1 Arg2	agent patient instrument	esperimentatzailea gaia	erg abs/ine/dat ins/soz	
joan	3	go.01	Arg0 Arg1 Arg2	theme source destination	gai ukitua abiapuntu kokapena helburuko kokapena	erg abl ala	
	4	consider.01	Arg0 Arg1 Arg2	agent theme predicate	abiapuntua helburua ezaugarria	erg abs abs/pro/des	+giz
	5	shoot.01	Arg0 Arg1 Arg2	agent patient destination	esperimentatzailea gai ukitua helburua	erg abs dat	
	1	go.01	Arg0 Arg1 Arg2	theme source destination	gai ukitua abiapuntu kokapena helburuko kokapena	abs abl ala/abu	
	1	play.01	Arg0 Arg1 Arg2	agent theme agent	esperimentatzailea jarduera	erg abs/ine/ala/dat soz	
jokatu	2	act.01	Arg0 Arg1	agent predicate	esperimentatzailea modua	erg ins/abs	
	3	bet.01	Arg0 Arg1 Arg2 Arg3	agent asset theme co-better	esperimentatzailea neurria gaia	erg abs abs/ine/konpl soz/dat	
	1	play.01/ring.01	Arg0 Arg1 Arg2 Arg3	agent theme agent instrument	esperimentatzailea jarduera esperimentatzailea	abs/erg ine/abs/ala soz	+biz +biz
justifikatu	1	warrant.01/justify.01	Arg0 Arg1 Arg2	agent topic recipient	esperimentatzailea gaia	erg abs/konpl dat	
kargatu	1	load.01	Arg0 Arg1 Arg2	agent destination theme	kausa helburuko kokapena gai ukitua	erg abs soz/ins	
kateatu	1	chain.01	Arg0 Arg1 Arg2	agent patient patient	kausa gai ukitua gai ukitua	erg abs dat/soz	
kendu	1	remove.01	Arg0 Arg1 Arg2	agent theme source	gai ukitua theme	erg abs abl	
	2	take.01/steal.01	Arg0 Arg1 Arg2	agent theme source	gai ukitua theme source	erg abs dat	
kezkatu	1	worry.01	Arg0 Arg1	cause experiencier	kausa gai ukitua	ins/soz/erg abs	+giz
kokatu	1	place.01/situate.01	Arg0	agent	kausa	erg	

Jarraipena hurrengo orrialdean

Aditza	Adiera	PB-VN ordaina	Balantzia	VN rola	EADB rola	Postposizio-atzizkia	HM
			Arg1 Arg2	theme destination	gai ukitua helburuko kokapena	abs/par ine	
		stand.04	Arg0 Arg1 Arg2	agent theme destination	kausa gai ukitua helburuko kokapena	erg abs ine	
kondenatu	1	condemn.01	Arg0 Arg1 Arg2 Arg3	agent theme predicate destination	abiapuntua helburua gaia	erg abs/dat mot ala	+giz +giz
konektatu	1	connect.01	Arg0 Arg1 Arg2	agent patient patient	kausa gai ukitua gai ukitua	erg abs dat/soz	
konparatu	1	compare.01	Arg0 Arg1 Arg2	agent patient patient	esperimentatzailea gaia	erg abs soz	
kontatu	1	narrate.01	Arg0 Arg1 Arg2	agent topic recipient	esperimentatzailea gaia	erg abs/konpl dat	
	2	rely.01	Arg0 Arg1 Arg2	agent theme	esperimentatzailea gaia	erg soz/ine	
	3	count.01	Arg0 Arg1	agent theme	esperimentatzailea gaia	erg abs	
konturatu	1	realize.01	Arg0 Arg1	agent theme	esperimentatzailea gaia	abs ins/konpl	
korritu	1	run.02	Arg0 Arg1	theme location		erg abs/ine	
kritikatu	1	criticize.01	Arg0 Arg1 Arg2	agent theme predicate	abiapuntua helburua gaia	erg dat/abs/konpl abs/mot	+giz
lagundu	1	help.01	Arg0 Arg1 Arg2	agent theme beneficiary		erg ine/t(z)en abs/dat	
	2	accompany.01	Arg0 Arg1 Arg2 Arg3	agent theme source destination		erg abs/dat abl ala	
landatu	1	plant.01	Arg0 Arg1 Arg2	agent theme destination	kausa gai ukitua helburuko kokapena	erg abs/ins abs/ine	
landu	1	work.01/cultivate.01/develop.02/discuss.01/sculpt.01	Arg0 Arg1 Arg2 Arg3 Arg4	agent product material beneficiary instrument	esperimentatzailea gaia	erg abs abl dat/des soz/ins	+giz -biz
laudatu	1	commend.01	Arg0 Arg1 Arg2	agent theme predicate	abiapuntua helburua gaia	erg abs mot	+giz +giz -biz
laztandu	1	caress.01/stroke.01	Arg0 Arg1 Arg2	agent experientier instrument	esperimentatzailea gaia	erg abs ins/soz	+giz
lehiatu	1	compete.01	Arg0	actor	gai ukitua	abs	

Jarraipena hurrengo orrialdean

Aditza	Adiera	PB-VN ordaina	Balantzia	VN rola	EADB rola	Postposizio-atzizkia	HM
leitu	1	read.01	Arg1 Arg0 Arg1 Arg2	actor agent theme recipient	gai ukitua esperimentatzailea gaia	dat/soz erg abs/konpl dat	+giz -biz
lortu	1	obtain.01	Arg0 Arg1 Arg2	agent theme source	esperimentatzailea gaia	erg abs/par/gen abl	
miatu	1	win.01	Arg0 Arg1	agent theme	esperimentatzailea gaia	erg abs	
mintzatu	1	scrutinize.01	Arg0 Arg1	experiencer stimulus	esperimentatzailea gaia	erg abs	+giz
modelatu	1	speak.01	Arg0 Arg1 Arg2	actor topic	abiapuntua gaia	abs ins/inc soz/dat	
moldatu	1	model.01	Arg0 Arg1 Arg2	agent product material	kausua gai ukitua	erg abs soz/ins	
moztu	1	mold.01/sculpt.01/make.01	Arg0 Arg1 Arg2	agent product material	kausua gai ukitua	erg abs soz/ins	
mugitu	2	manage.02	Arg1 Arg2	agent theme		abs	
	3	adapt.01	Arg1 Arg2	patient destination		soz/t(z)eko/t(z)en abs ala	
nabarmendu	1	cut.01/interrupt.01/block.01	Arg0 Arg1 Arg2 Arg3	agent patient patient instrument	esperimentatzailea gaia	erg abs dat ins/soz	-biz
	1	move.01	Arg0 Arg1 Arg2 Arg3	agent theme source destination	kausua gai ukitua abiapuntuoko kokapena helburuko kokapena	erg abs/par abl ala	
nahasi	1	stand.06	Arg1 Arg2	theme topic		abs	
nahastu	1	emphasize.01	Arg0 Arg1 Arg2	recipient		abs/konpl dat	
negoziatu	1	jumble.01	Arg0 Arg1 Arg2	agent patient patient	kausua gai ukitua gai ukitua	erg abs soz	
	1	mix.01	Arg0 Arg1 Arg2	agent patient patient	kausua gai ukitua gai ukitua	erg abs soz	
	2	confuse.01	Arg0 Arg1	agent patient	kausua gai ukitua	erg abs	
neurtu	1	negotiate.01	Arg0 Arg1 Arg2	agent actor theme	kausua gai ukitua gai ukitua	erg soz abs	
	1	measure.01	Arg0 Arg1 Arg2	agent theme recipient	esperimentatzailea gaia	erg abs dat	+giz -biz
	2	contend.02	Arg1	actor		abs	

Jarraipena hurrengo orrialdean

Aditza	Adiera	PB-VN ordaina	Balantzia	VN rola	EADB rola	Postposizio-atzizkia	HM
ohoratu	1	honor.01	Arg2 Arg0 Arg1 Arg2	actor agent theme predicate	abiapuntua helburua gaia	soz/ren aurka/kontra erg abs mot	+giz +giz -biz
	1	bawl.01	Arg0 Arg1 Arg2	agent topic recipient	esperimentatzailea gaia	erg abs dat	+biz -biz
onartu	1	accept.01/allow.01/approve.01	Arg0 Arg1 Arg2	agent theme source	esperimentatzailea gaia	erg abs/konpl/mod/par dat	
ordaindu	1	pay.01	Arg0 Arg1 Arg2 Arg3	agent asset beneficiary theme		erg abs dat abs/mot	
	1	remember.01	Arg0 Arg1 Arg2	agent theme predicate	esperimentatzailea gaia	abs/erg ins/abs/konpl/soz post	
osatu	1	complete.01	Arg0 Arg1	agent theme		erg abs	
	2	form.02/constitute.01	Arg0 Arg1	agent theme		erg/ins abs	
oroitu	3	cure.01/recover.01	Arg0 Arg1 Arg2	agent source theme		erg dat/abs abs/abl	
	4	assemble.01	Arg0 Arg1 Arg2	agent product material		erg abs soz/ins	
ofu	1	occur.01	Arg1 Arg2	theme beneficiary	gaia	konpl dat	+giz
pagatu	1	pay.01	Arg0 Arg1 Arg2	agent theme recipient	esperimentatzailea abiapuntua gaia helburua	erg abs dat	
	1	pass.19	Arg0 Arg1 Arg2 Arg3	agent theme source/path destination	kausa gai ukitua abiapuntua/bidea helburua	erg abs abl/me ala	
pasatu	2	pass.08	Arg0 Arg1	theme path	gai ukitua bidea	erg abs	
	3	pass.05 happen.01/suffer.01 pass.03 spend.02	Arg0 Arg1 Arg2 Arg1 Arg2	agent theme recipient experiencer cause theme	abiapuntua gaia helburua esperimentatzailea gaia gaia	erg abs dat erg abs abs	+giz -biz +giz +giz -konkr -konkr
pausatu	1	perch.01	Arg0 Arg1 Arg2	agent theme recipient	kausa gai ukitua	erg abs ine/soz	+giz -konkr
	1	think.01	Arg0	agent	helburuko kokapena	erg abs ine	+giz -konkr

Jarraipena hurrengo orrialdean

Aditza	Adiera	PB-VN ordaina	Balantzia	VN rola	EADB rola	Postposizio-atzizkia	HM
pitzatu	1	split.01	Arg1	theme		abs/konpl	
			Arg2	attribute		ine/ins/ri buruz	
prestatu	1	prepare.01	Arg0	agent	kausa	erg	
			Arg1	patient	gai ukitua	abs	
proposatu	2	prepare.02	Arg2	instrument		soz/ins	
			Arg0	agent		erg	
puskatu	1	propose.01	Arg1	beneficiary		abs	
			Arg2	agent		dat/des	
saiatu	1	propose.01	Arg0	theme		erg	
			Arg1	agent		abs	
sailkatu	1	break.01	Arg0	topic		erg	
			Arg2	recipient		abs/konpl	
salatu	1	try.01	Arg0	agent	kausa	erg	
			Arg1	patient	gai ukitua	abs	
salkatu	1	classify.01/class.01	Arg2	instrument		soz/ins	
			Arg1	theme		abs	
saldu	1	reward.01	Arg2	predicate		t(z)en/t(z)era	
			Arg0	agent		erg	
sarraskitu	1	massacre.01	Arg0	agent	abiapuntua	erg	
			Arg1	theme	helburua	abs/konpl	
sartu	1	put.01/include.01	Arg2	predicate	gai	mot	
			Arg0	agent	abiapuntua	erg	
sentitu	1	sign.01	Arg1	theme	gai	abs	
			Arg2	recipient	helburua	dat	
sinatu	2	sign.02	Arg0	agent	abiapuntua	erg	
			Arg1	theme	helburua	abs	
sarraskitu	1	massacre.01	Arg0	agent	kausa	erg	
			Arg1	patient	gai ukitua	abs/gen	
sartu	1	enter.01	Arg2	instrument		ine/soz/ins	
			Arg1	theme	gai ukitua	abs	
sentitu	1	feel.01	Arg0	location	helburuko kokapena	ine/ala	
			Arg1	agent	kausa	erg	
sinatu	1	sign.01	Arg0	theme	gai ukitua	abs	
			Arg1	source	abiapuntuko kokapena	abl	
sartu	2	score.01	Arg2	location	helburuko kokapena	ala/ine	
			Arg0	agent	kausa	erg	
sentitu	1	sign.01	Arg1	theme	gai ukitua	abs/par	
			Arg0	experiencer		erg/abs	
sinatu	2	sign.02	Arg1	stimulus		abs/par/konpl/ine/soz	
			Arg0	agent		erg	
sartu	1	enter.01	Arg1	destination		abs	
			Arg2	theme		soz/ins	
sartu	1	enter.01	Arg0	agent		erg	
			Arg1	theme		abs	

Jarraipena hurrengo orrialdean

Aditza	Adiera	PB-VN ordaina	Balantzia	VN rola	EADB rola	Postposizio-atzizkia	HM
sortu	1	create.01	Arg2	agent		soz	
			Arg0	agent		erg	
			Arg1	product		abs	
sumatu	1	perceive.01	Arg2	beneficiary		soz/ins	
			Arg3	material		dat/des	
			Arg0	experiencer	esperimentatzailea	erg	abs/konpl
topatu	1	find.01	Arg2	stimulus	gaia	abs	
			Arg0	predicate		abs/erg	
			Arg1	theme	esperimentatzailea	abs/soz/par	
transkribatu	1	transcribe.01	Arg2	beneficiary	gaia	dat	
			Arg0	agent		erg	
			Arg1	theme	esperimentatzailea	abs	
ukitu	1	touch.01	Arg0	experiencer	gaia	erg	
			Arg1	instrument	esperimentatzailea	abs	
			Arg2	instrument	gaia	ins/soz	
ulertu	1	understand.01	Arg0	agent	gaia	erg	
			Arg1	proposition		abs/konpl/gen/par	
			Arg2	predicate	gaia	-ri buruz...	
zabaldtu	1	open.01	Arg0	agent		erg	
			Arg1	patient		abs/par	
			Arg2	instrument		soz/ins	
zarpatu	2	enlarge.01	Arg3	recipient		dat	
			Arg1	agent		erg	
			Arg1	patient		abs	
zartatu	3	spread.02	Arg0	agent		erg	
			Arg1	theme		abs/konpl	
			Arg2	location		ine/ala	
zauritu	1	wound.01	Arg0	agent	esperimentatzailea	erg	
			Arg1	patient	gai ukitua	abs	
			Arg2	instrument		soz/ins	
zeharkatu	1	crack.01	Arg0	agent	kausa	erg	
			Arg1	patient	gai ukitua	abs	
			Arg2	instrument		soz/ins	
zehaztu	1	specify.01/detail.01	Arg0	stimulus		erg	
			Arg1	experiencer		abs	
			Arg2	instrument		soz/ins	
zigortu	1	condemn.01/punish.01	Arg0	theme	gai ukitua	erg	
			Arg1	location	bidea	abs	
			Arg1	agent		erg	
zimeldu	1	wither.01	Arg0	theme		abs/konpl	
			Arg1	agent	abiapuntua	erg	
			Arg2	theme	helburua	abs	
zintzilikatu	1	hang.01/suspend.02	Arg3	predicate	kausa	mot	
			Arg2	destination	gaia	ala/soz	
			Arg0	agent	gai ukitua	erg	
zorroztu	1	sharpen.01	Arg1	patient		abs	
			Arg0	agent	kausa	erg	
			Arg1	theme	gai ukitua	abs	
zorroztu	1	sharpen.01	Arg2	destination	helburuko kokapena	abs	
			Arg1	theme	kausa	abl/ine	
			Arg0	agent	kausa	erg	

Jarraipena hurrengo orrialdean

Aditza	Adiera	PB-VN ordaina	Balantzia	VN rola	EADB rola	Postposizio-atzizkia	HM
zuzendu	1	manage.01/supervise.01/lead.01/conduct.01	Arg1 Arg2	patient instrument	gai ukitutua	abs soz/ins	
	2	address.01/direct.01	Arg0 Arg1	agent theme	esperimentatzailea gaia	erg abs	
	3	correct.01	Arg0 Arg2 Arg1	agent theme destination agent patient	esperimentatzailea gaia helburuko kokapena esperimentatzailea gai ukitutua	erg abs dat/ala erg abs	



EPEC-DEPeko aditzak eta maiztasunak

Aditza	Agerpenak	Portzentajea %
izan	4.832	15,7189
egin	2.471	8,0384
egon	1.533	4,9870
ukan	1.529	4,9740
esan	642	2,0885
eman	602	1,9584
jokatu	508	1,6526
hartu	505	1,6428
joan	407	1,3240
lor্তু	370	1,2036
irabazi	358	1,1646
hasi	350	1,1386
ikusi	349	1,1353
sartu	302	0,9824
jakin	257	0,8360
utzi	257	0,8360
jarri	247	0,8035
etorri	229	0,7450
sortu	215	0,6994
eduki	212	0,6897

(jarraipena hurrengo orrialdean)

Aditza	Agerpenak	Portzentajea %
gertatu	204	0,6636
adierazi	199	0,6474
hil	190	0,6181
eskatu	190	0,6181
jo	187	0,6083
jaso	187	0,6083
atera	185	0,6018
azaldu	185	0,6018
erabili	182	0,5921
galdu	180	0,5856
onartu	172	0,5595
behar_izan	167	0,5433
ibili	163	0,5303
aritu	159	0,5172
eraman	148	0,4815
aurkeztu	145	0,4717
bildu	144	0,4684
jarraitu	143	0,4652
osatu	136	0,4424
geratu	136	0,4424
erakutsi	134	0,4359
agertu	134	0,4359
erabaki	132	0,4294
amaitu	124	0,4034
bete	121	0,3936
aipatu	115	0,3741
heldu	111	0,3611
aztertu	107	0,3481
iritsi	103	0,3351
iruditu	101	0,3286
gelditu	99	0,3221
saiatu	94	0,3058
ezagutu	92	0,2993
itzuli	91	0,2960
pasatu	89	0,2895
edun	88	0,2863

(jarraipena hurrengo orrialdean)

Aditza	Agerpenak	Portzentajea %
aurkitu	87	0,2830
aldatu	84	0,2733
eragin	84	0,2733
zabaldu	78	0,2537
ekin	78	0,2537
entzun	77	0,2505
ekarri	77	0,2505
salatu	77	0,2505
kendu	76	0,2472
pentsatu	74	0,2407
eutsi	73	0,2375
antolatu	72	0,2342
erantzun	72	0,2342
deitu	72	0,2342
eskaini	72	0,2342
azpimarratu	70	0,2277
itxi	70	0,2277
idatzi	69	0,2245
irten	68	0,2212
sailkatu	67	0,2180
jardun	66	0,2147
atxilotu	65	0,2115
lagundu	64	0,2082
ari_izan	62	0,2017
prestatu	61	0,1984
eskuratu	60	0,1952
ikasi	60	0,1952
bukatu	60	0,1952
bilatu	59	0,1919
bota	57	0,1854
abiatu	55	0,1789
jakinaraz	55	0,1789
ezarri	53	0,1724
iragarri	52	0,1692
aukeratu	48	0,1561
sentitu	48	0,1561

(jarraipena hurrengo orrialdean)

Aditza	Agerpenak	Portzentajea %
banatu	48	0,1561
gainditu	48	0,1561
entrenatu	48	0,1561
luzatu	47	0,1529
jasan	47	0,1529
nabarmendu	47	0,1529
igo	46	0,1496
ireki	46	0,1496
zehaztu	46	0,1496
edin	46	0,1496
lotu	45	0,1464
sinatu	45	0,1464
proposatu	45	0,1464
zauritu	45	0,1464
begiratu	45	0,1464
bultzatu	44	0,1431
piztu	43	0,1399
bihurtu	42	0,1366
iragan	42	0,1366
ulertu	42	0,1366
nagusitu	42	0,1366
landu	42	0,1366
bidali	42	0,1366
fitxatu	41	0,1334
irakurri	41	0,1334
gogoratu	41	0,1334
bereizi	41	0,1334
aitortu	41	0,1334
baztertu	40	0,1301
erori	39	0,1269
ahaztu	39	0,1269
zuzendu	38	0,1236
leporatu	38	0,1236
jaitsi	38	0,1236
mintzatu	38	0,1236
jantzi	37	0,1204

(jarraipena hurrengo orrialdean)

Aditza	Agerpenak	Portzentajea %
izendatu	36	0,1171
iraun	35	0,1139
desagertu	35	0,1139
deklaratu	35	0,1139
egokitu	34	0,1106
babestu	34	0,1106
baliatu	34	0,1106
kokatu	34	0,1106
korritu	34	0,1106
konpondu	33	0,1074
agindu	33	0,1074
armatu	33	0,1074
zelairatu	33	0,1074
aurreratu	32	0,1041
elkartu	32	0,1041
ordaindu	31	0,1008
aldarrikatu	31	0,1008
gustatu	31	0,1008
gauzatu	31	0,1008
asmatu	31	0,1008
argitu	31	0,1008
igaro	31	0,1008
hautsi	31	0,1008
hurbildu	30	0,0976
eraiki	30	0,0976
erosi	30	0,0976
hobetu	29	0,0943
mantendu	29	0,0943
segitu	29	0,0943
galdetu	29	0,0943
ukatu	29	0,0943
iritzi	29	0,0943
zaindu	28	0,0911
kontatu	28	0,0911
handitu	28	0,0911
kanporatu	27	0,0878

(jarraipena hurrengo orrialdean)

Aditza	Agerpenak	Portzentajea %
bermatu	27	0,0878
ziurtatu	27	0,0878
oinarritu	27	0,0878
harrapatu	27	0,0878
kezkatu	27	0,0878
behartu	27	0,0878
egotzi	26	0,0846
finkatu	26	0,0846
baieztatu	25	0,0813
mugitu	25	0,0813
burutu	25	0,0813
mugatu	25	0,0813
erretiratu	25	0,0813
jaio	24	0,0781
argitaratu	24	0,0781
gorde	24	0,0781
bizitu	24	0,0781
frogatu	24	0,0781
hautatu	24	0,0781
gaitzetsi	24	0,0781
garatu	24	0,0781
erran	24	0,0781
debekatu	23	0,0748
ikertu	23	0,0748
lehiatu	23	0,0748
aurreikusi	23	0,0748
bilakatu	23	0,0748
zigortu	22	0,0716
berdindu	22	0,0716
eseri	22	0,0716
gehitu	22	0,0716
murriztu	22	0,0716
kutsatu	22	0,0716
hazi	22	0,0716
eratu	22	0,0716
ezkutatu	22	0,0716

(jarraipena hurrengo orrialdean)

Aditza	Agerpenak	Portzentajea %
jan	21	0,0683
ordezkatu	21	0,0683
ahalegindu	21	0,0683
berreskura	21	0,0683
bateratu	21	0,0683
neurtu	20	0,0651
eztabaidatu	20	0,0651
saldu	20	0,0651
bereganatu	20	0,0651
atzeratu	20	0,0651
erantsi	20	0,0651
gozatu	20	0,0651
berritu	19	0,0618
batu	19	0,0618
tratatu	19	0,0618
ohitu	19	0,0618
arduratu	18	0,0586
ezkondu	18	0,0586
askatu	18	0,0586
jolastu	18	0,0586
nahastu	18	0,0586
gaineratu	18	0,0586
kontrolatu	18	0,0586
harritu	18	0,0586
kritikatu	18	0,0586
hitzartu	18	0,0586
antzeman	18	0,0586
buelatu	17	0,0553
inguratu	17	0,0553
lapurtu	17	0,0553
ipini	16	0,0520
ospatu	16	0,0520
konturatu	16	0,0520
aireratu	16	0,0520
errespetatu	16	0,0520
eraso	16	0,0520

(jarraipena hurrengo orrialdean)

Aditza	Agerpenak	Portzentajea %
sustatu	16	0,0520
sinetsi	16	0,0520
jakiN+araz	16	0,0520
eragotzi	15	0,0488
jaurti	15	0,0488
aprobetxatu	15	0,0488
berretsi	15	0,0488
erre	15	0,0488
moldatu	15	0,0488
akusatu	15	0,0488
bahitu	15	0,0488
plazaratu	15	0,0488
nabariu	14	0,0455
topatu	14	0,0455
moztu	14	0,0455
eten	14	0,0455
epaitu	14	0,0455
estali	14	0,0455
ohartu	14	0,0455
errepikatu	14	0,0455
altxatu	14	0,0455
beR+eskura	14	0,0455
ebaki	14	0,0455
exijitu	14	0,0455
kaleratu	14	0,0455
indartu	14	0,0455
botatu	14	0,0455
ohartaraz	14	0,0455
sendotu	14	0,0455
saihestu	14	0,0455
nekatu	13	0,0423
areagotu	13	0,0423
bozkatu	13	0,0423
sumatu	13	0,0423
argudiatu	13	0,0423
bideratu	13	0,0423

(jarraipena hurrengo orrialdean)

Aditza	Agerpenak	Portzentajea %
hedatu	13	0,0423
atxiki	13	0,0423
gutxitu	13	0,0423
jaiki	13	0,0423
gailendu	13	0,0423
gonbidatu	13	0,0423
indargabetu	12	0,0390
konformatu	12	0,0390
definitu	12	0,0390
nahasi	12	0,0390
atzeman	12	0,0390
markatu	12	0,0390
garbitu	12	0,0390
itzali	12	0,0390
defenditu	12	0,0390
lasaitu	12	0,0390
galdegin	12	0,0390
alderatu	12	0,0390
orekatu	12	0,0390
zeharkatu	12	0,0390
lehertu	12	0,0390
berotu	11	0,0358
kolpatu	11	0,0358
gezurtatu	11	0,0358
oztopatu	11	0,0358
pairatu	11	0,0358
nahi_izan	11	0,0358
grabatu	11	0,0358
justifikatu	11	0,0358
arakatu	11	0,0358
agortu	11	0,0358
egiaztatu	11	0,0358
etsi	11	0,0358
inputatu	11	0,0358
edan	11	0,0358
etzan	11	0,0358

(jarraipena hurrengo orrialdean)

Aditza	Agerpenak	Portzentajea %
ondu	11	0,0358
apurtu	11	0,0358
eskertu	10	0,0325
kostatu	10	0,0325
merezi_izan	10	0,0325
urrundu	10	0,0325
aplikatu	10	0,0325
ezan	10	0,0325
izenpetu	10	0,0325
zapuztu	10	0,0325
igarri	10	0,0325
negoziatu	10	0,0325
hondatu	10	0,0325
ito	10	0,0325
agurtu	10	0,0325
ezabatu	10	0,0325
gidatu	10	0,0325
sufritu	10	0,0325
identifikatu	10	0,0325
jazo	10	0,0325
murgildu	10	0,0325
mindu	10	0,0325
auzipetu	10	0,0325
itxaron	10	0,0325
gurutzatu	9	0,0293
arautu 9	9	0,0293
defendatu	9	0,0293
balio_izan	9	0,0293
okertu	9	0,0293
banandu	9	0,0293
erakarri	9	0,0293
txikitu	9	0,0293
menderatu	9	0,0293
governatu	9	0,0293
suntsitu	9	0,0293
tirokatu	9	0,0293

(jarraipena hurrengo orrialdean)

Aditza	Agerpenak	Portzentajea %
zenbatu	9	0,0293
borrokatu	9	0,0293
ugaldu	9	0,0293
baloratu	9	0,0293
hondoratu	9	0,0293
adostu	9	0,0293
hornitu	9	0,0293
elkarretaratu	9	0,0293
islatu	9	0,0293
galazaraz	8	0,0260
laburbildu	8	0,0260
jabetu	8	0,0260
salbatu	8	0,0260
ukitu	8	0,0260
ailegatu	8	0,0260
kobratu	8	0,0260
lesionatu	8	0,0260
ondorioztatu	8	0,0260
esnatu	8	0,0260
ikaratu	8	0,0260
erdietsi	8	0,0260
goraipatu	8	0,0260
irakatsi	8	0,0260
igorri	8	0,0260
ezin_izan	8	0,0260
aldendu	8	0,0260
suertatu	8	0,0260
moderatu	8	0,0260
kaltetu	8	0,0260
baretu	8	0,0260
irentsi	8	0,0260
zulatu	8	0,0260
kantatu	8	0,0260
bisitatu	8	0,0260
ahantzi	8	0,0260
errekuperatu	8	0,0260

(jarraipena hurrengo orrialdean)

Aditza	Agerpenak	Portzentajea %
hitz_egin	8	0,0260
isuri	7	0,0228
ureztatu	7	0,0228
behatu	7	0,0228
erail	7	0,0228
haserretu	7	0,0228
pilatu	7	0,0228
zatitu	7	0,0228
konbentzitu	7	0,0228
alde_egin	7	0,0228
afizionatu	7	0,0228
planteatu	7	0,0228
espetxeratu	7	0,0228
desegin	7	0,0228
izutu	7	0,0228
helaraz	7	0,0228
animatu	7	0,0228
baldintzatu	7	0,0228
zapaldu	7	0,0228
zendu	6	0,0195
uxatu	6	0,0195
kikildu	6	0,0195
gogoraraz	6	0,0195
uztartu	6	0,0195
ekidin	6	0,0195
diseinatu	6	0,0195
laga	6	0,0195
deitoratu	6	0,0195
bazkaldu	6	0,0195
faltatu	6	0,0195
asaldatu	6	0,0195
elkarrizketatu	6	0,0195
estutu	6	0,0195
esleitu	6	0,0195
urtu	6	0,0195
astindu	6	0,0195

(jarraipena hurrengo orrialdean)

Aditza	Agerpenak	Portzentajea %
aliatu	6	0,0195
abestu	6	0,0195
integratu	6	0,0195
sakabanatu	5	0,0163
marraztu	5	0,0163
espiatu	5	0,0163
apaindu	5	0,0163
azaleratu	5	0,0163
bururatu	5	0,0163
erraztu	5	0,0163
polarizatu	5	0,0163
lehortu	5	0,0163
hustu	5	0,0163
bortxatu	5	0,0163
ugaritu	5	0,0163
busti	5	0,0163
konparatu	5	0,0163
erreparatu	5	0,0163
aspertu	5	0,0163
muntatu	5	0,0163
zimurtu	5	0,0163
afaldu	5	0,0163
sendatu	5	0,0163
dastatu	5	0,0163
landatu	5	0,0163
kondenatu	5	0,0163
kontrajarri	5	0,0163
josi	5	0,0163
itsutu	5	0,0163
kexatu	5	0,0163
orraztu	5	0,0163
jipoitu	5	0,0163
dominatu	5	0,0163
inbertitu	5	0,0163
zutitu	5	0,0163
inkomunikatu	5	0,0163

(jarraipena hurrengo orrialdean)

Aditza	Agerpenak	Portzentajea %
beR+etsi	5	0,0163
saskiratu	5	0,0163
hezi	5	0,0163
funtzionatu	5	0,0163
txalotu	5	0,0163
bainatu	5	0,0163
ahuldu	5	0,0163
gerturatu	5	0,0163
iradoki	5	0,0163
beldurtu	5	0,0163
zimeldu	5	0,0163
mehatxatu	5	0,0163
seinalatu	5	0,0163
miatu	5	0,0163
helmugaratu	5	0,0163
blokeatu	5	0,0163
jorratu	5	0,0163
kudeatu	5	0,0163
poztu	5	0,0163
barneratu	5	0,0163
kontratatu	5	0,0163
ase	5	0,0163
trebatu	4	0,0130
parekatu	4	0,0130
harrotu	4	0,0130
larritu	4	0,0130
ahitu	4	0,0130
ezindu	4	0,0130
aitzinatu	4	0,0130
zaildu	4	0,0130
atsekabetu	4	0,0130
gutxietsi	4	0,0130
sakondu	4	0,0130
azkartu	4	0,0130
ikuskatu	4	0,0130
existitu	4	0,0130

(jarraipena hurrengo orrialdean)

Aditza	Agerpenak	Portzentajea %
sostengatu	4	0,0130
gogortu	4	0,0130
kriminalizatu	4	0,0130
aditu	4	0,0130
galdekatu	4	0,0130
erne	4	0,0130
ohart+araz	4	0,0130
taxutu	4	0,0130
lurperatu	4	0,0130
puskatu	4	0,0130
etxeratu	4	0,0130
gizarteratu	4	0,0130
barkatu	4	0,0130
birziklatu	4	0,0130
oroitu	4	0,0130
imajinatu	4	0,0130
erotu	4	0,0130
geldiaraz	4	0,0130
bidaiatu	4	0,0130
desbideratu	4	0,0130
eskegi	4	0,0130
motibatu	4	0,0130
estreinatu	4	0,0130
gogora+araz	4	0,0130
kotizatu	4	0,0130
urratu	4	0,0130
kateatu	4	0,0130
bihurritu	4	0,0130
okupatu	4	0,0130
deskubritu	4	0,0130
eraitsi	4	0,0130
igurtzi	4	0,0130
irudikatu	4	0,0130
zoratu	4	0,0130
erlazionatu	4	0,0130
xedatu	4	0,0130

(jarraipena hurrengo orrialdean)

Aditza	Agerpenak	Portzentajea %
akabatu	4	0,0130
kontzentratu	4	0,0130
margotu	4	0,0130
zintzilikatu	4	0,0130
artxibatu	4	0,0130
hunkitu	4	0,0130
desjabetu	4	0,0130
moteldu	4	0,0130
jin	4	0,0130
arindu	4	0,0130
baliogabetu	4	0,0130
oihukatu	4	0,0130
interesatu	4	0,0130
maitatu	3	0,0098
txunditu	3	0,0098
zizelatu	3	0,0098
biribildu	3	0,0098
bizi_izan	3	0,0098
zartatu	3	0,0098
makurtu	3	0,0098
jiratu	3	0,0098
dimititu	3	0,0098
berragertu	3	0,0098
baimendu	3	0,0098
ahoskatu	3	0,0098
tartekatu	3	0,0098
errotu	3	0,0098
jausi	3	0,0098
manifestatu	3	0,0098
aurpegiratu	3	0,0098
paseatu	3	0,0098
pitzatu	3	0,0098
bakartu	3	0,0098
higitu	3	0,0098
ekoitzi	3	0,0098
eraberritu	3	0,0098

(jarraipena hurrengo orrialdean)

Aditza	Agerpenak	Portzentajea %
gosaldu	3	0,0098
eraldatu	3	0,0098
ebatzi	3	0,0098
legitimatu	3	0,0098
probokatu	3	0,0098
zanpatu	3	0,0098
zirikatu	3	0,0098
estimatu	3	0,0098
hauteman	3	0,0098
kantxaratu	3	0,0098
usteldu	3	0,0098
errekonozitu	3	0,0098
ezberdindu	3	0,0098
libratu	3	0,0098
konektatu	3	0,0098
txukundu	3	0,0098
sorta+araz	3	0,0098
errekurritu	3	0,0098
susmatu	3	0,0098
zahartu	3	0,0098
erreatratu	3	0,0098
miretsi	3	0,0098
fidatu	3	0,0098
tolestu	3	0,0098
esplikatu	3	0,0098
gal+araz	3	0,0098
berpiztu	3	0,0098
garraiatu	3	0,0098
arbuia	3	0,0098
indargetu	3	0,0098
berriztatu	3	0,0098
aholkatu	3	0,0098
presionatu	3	0,0098
gorritu	3	0,0098
laburtu	3	0,0098
inposatu	3	0,0098

(jarraipena hurrengo orrialdean)

Aditza	Agerpenak	Portzentajea %
errenditu	3	0,0098
omendu	3	0,0098
birlanda	3	0,0098
engainatu	3	0,0098
ihartu	3	0,0098
estaldu	3	0,0098
itsatsi	3	0,0098
ahalbidetu	3	0,0098
baketu	3	0,0098
gomendatu	3	0,0098
probatu	3	0,0098
lurrindu	3	0,0098
afiliatu	3	0,0098
ilundu	3	0,0098
normalizatu	3	0,0098
geldi+araz	3	0,0098
deliberatu	3	0,0098
leundu	3	0,0098
tipifikatu	3	0,0098
bonbardatu	3	0,0098
abisatu	3	0,0098
estraditatu	3	0,0098
berrets	3	0,0098
izoztu	3	0,0098
suposatu	3	0,0098
jalgi	3	0,0098
ferekatu	3	0,0098
tokatu	3	0,0098
gaixotu	2	0,0065
dopatu	2	0,0065
garestitu	2	0,0065
itxuratu	2	0,0065
derrigortu	2	0,0065
errematatu	2	0,0065
presiditu	2	0,0065
suspertu	2	0,0065

(jarraipena hurrengo orrialdean)

Aditza	Agerpenak	Portzentajea %
mitifikatu	2	0,0065
finantzatu	2	0,0065
deskribatu	2	0,0065
funditu	2	0,0065
apaldu	2	0,0065
itaundu	2	0,0065
zorabiatu	2	0,0065
erreakzionatu	2	0,0065
inplikatu	2	0,0065
oldartu	2	0,0065
zakartu	2	0,0065
trabatu	2	0,0065
esku_hartu	2	0,0065
otu	2	0,0065
haizatu	2	0,0065
matxuratu	2	0,0065
zorpetu	2	0,0065
kalkulatu	2	0,0065
arrestatu	2	0,0065
portatu	2	0,0065
inprobisatu	2	0,0065
kitatu	2	0,0065
pribatizatu	2	0,0065
oroit+araz	2	0,0065
idealizatu	2	0,0065
tragatu	2	0,0065
aberastu	2	0,0065
handiagotu	2	0,0065
ihesi	2	0,0065
isil+araz	2	0,0065
disolbatu	2	0,0065
zozketatu	2	0,0065
planifikatu	2	0,0065
solidotu	2	0,0065
bete+araz	2	0,0065
artatu	2	0,0065

(jarraipena hurrengo orrialdean)

Aditza	Agerpenak	Portzentajea %
pozoitu	2	0,0065
akuratu	2	0,0065
hoztu	2	0,0065
tindatu	2	0,0065
iraindu	2	0,0065
desitxuratu	2	0,0065
eskalatu	2	0,0065
goxatu	2	0,0065
kontzientziatu	2	0,0065
samurtu	2	0,0065
isladatu	2	0,0065
euskaldundu	2	0,0065
nozitu	2	0,0065
sorraz	2	0,0065
barreiatu	2	0,0065
biraraz	2	0,0065
alboratu	2	0,0065
profitatu	2	0,0065
deskargatu	2	0,0065
apailatu	2	0,0065
erein	2	0,0065
baietsi	2	0,0065
urdindu	2	0,0065
kargugabetu	2	0,0065
lardaskatu	2	0,0065
erraustu	2	0,0065
obeditu	2	0,0065
hetsi	2	0,0065
deportatu	2	0,0065
ehizatu	2	0,0065
hobetsi	2	0,0065
jalki	2	0,0065
kolonizatu	2	0,0065
zartaraz	2	0,0065
maitemindu	2	0,0065
osotu	2	0,0065

(jarraipena hurrengo orrialdean)

Aditza	Agerpenak	Portzentajea %
sakatu	2	0,0065
trenkatu	2	0,0065
kontserbatu	2	0,0065
partekatu	2	0,0065
merkatu	2	0,0065
polikromatu	2	0,0065
lotsatu	2	0,0065
torturatu	2	0,0065
irauli	2	0,0065
bizkarreratu	2	0,0065
konkistatu	2	0,0065
amestu	2	0,0065
ostu	2	0,0065
kreditatu	2	0,0065
esportatu	2	0,0065
damutu	2	0,0065
besarkatu	2	0,0065
kausitu	2	0,0065
egituratu	2	0,0065
arrazoitu	2	0,0065
bikoiztu	2	0,0065
lekutu	2	0,0065
zaletu	2	0,0065
desertatu	2	0,0065
kommutatu	2	0,0065
jauzi	2	0,0065
desplazatu	2	0,0065
manipulatu	2	0,0065
munduratu	2	0,0065
txertatu	2	0,0065
nahastekatu	2	0,0065
hausnartu	2	0,0065
berregitura	2	0,0065
dardaratu	2	0,0065
ebatu	2	0,0065
argaldu	2	0,0065

(jarraipena hurrengo orrialdean)

Aditza	Agerpenak	Portzentajea %
arriskatu	2	0,0065
enmarkatu	2	0,0065
lerrokatu	2	0,0065
jarraiki	2	0,0065
oroitaraz	2	0,0065
puztu	2	0,0065
destakatu	2	0,0065
ausartu	2	0,0065
samartu	2	0,0065
ihinztatu	2	0,0065
isildu	2	0,0065
aztoratu	2	0,0065
berreraiki	2	0,0065
egosi	2	0,0065
gizendu	2	0,0065
kondentsatu	2	0,0065
marratu	2	0,0065
saritu	2	0,0065
berezitu	2	0,0065
inbaditu	2	0,0065
debutatu	2	0,0065
zurrupatu	2	0,0065
kargatu	2	0,0065
null	2	0,0065
herriratu	2	0,0065
uste_izan	2	0,0065
eguneratu	2	0,0065
ezetsi	2	0,0065
gogaitu	2	0,0065
prefabrikatu	2	0,0065
zaharkitu	2	0,0065
mozkortu	1	0,0033
freskatu	1	0,0033
amaiaraz	1	0,0033
beR+ets	1	0,0033
irristatu	1	0,0033

(jarraipena hurrengo orrialdean)

Aditza	Agerpenak	Portzentajea %
borobildu	1	0,0033
jira+araz	1	0,0033
dibertitu	1	0,0033
kontseilatu	1	0,0033
erasan	1	0,0033
absolbitu	1	0,0033
juzkatu	1	0,0033
labaindu	1	0,0033
ergeldu	1	0,0033
estuaraz	1	0,0033
agitu	1	0,0033
nazkatu	1	0,0033
haztatu	1	0,0033
separatu	1	0,0033
kartzelatu	1	0,0033
exekutatu	1	0,0033
obratu	1	0,0033
hezetu	1	0,0033
inkulpatu	1	0,0033
bigundu	1	0,0033
izu+araz	1	0,0033
ehortzi	1	0,0033
asentatu	1	0,0033
bataiatu	1	0,0033
desegokitu	1	0,0033
emigratu	1	0,0033
errezatu	1	0,0033
goretsi	1	0,0033
proporzionatu	1	0,0033
iratzarri	1	0,0033
eskulpitu	1	0,0033
amatu	1	0,0033
igurikatu	1	0,0033
beR+gizartera	1	0,0033
trukatu	1	0,0033
irudi+araz	1	0,0033

(jarraipena hurrengo orrialdean)

Aditza	Agerpenak	Portzentajea %
konsideratu	1	0,0033
kraskatu	1	0,0033
eratorri	1	0,0033
mastekatu	1	0,0033
esplotatu	1	0,0033
nagusiaraz	1	0,0033
partitu	1	0,0033
sedimentatu	1	0,0033
kalera+araz	1	0,0033
ebakuatu	1	0,0033
erlatibizatu	1	0,0033
merkaturatu	1	0,0033
aguantatu	1	0,0033
gastatu	1	0,0033
hel+araz	1	0,0033
sinbolizatu	1	0,0033
indarhartu	1	0,0033
bertaratu	1	0,0033
urrikaldu	1	0,0033
itzarri	1	0,0033
katxondeatu	1	0,0033
laudatu	1	0,0033
odoldu	1	0,0033
arnastu	1	0,0033
poltsikoratu	1	0,0033
higatu	1	0,0033
desagerraraz	1	0,0033
kiribildu	1	0,0033
lekualdatu	1	0,0033
erremolkatu	1	0,0033
ajatu	1	0,0033
preskribitu	1	0,0033
hondoraraz	1	0,0033
irabiatu	1	0,0033
birkoka	1	0,0033
deserritu	1	0,0033

(jarraipena hurrengo orrialdean)

Aditza	Agerpenak	Portzentajea %
komentatu	1	0,0033
endekatu	1	0,0033
modelatu	1	0,0033
finitu	1	0,0033
alderanztu	1	0,0033
gorrotatu	1	0,0033
atxiktu	1	0,0033
protestatu	1	0,0033
belaunikatu	1	0,0033
bizkortu	1	0,0033
desordenatu	1	0,0033
kontsultatu	1	0,0033
futitu	1	0,0033
amildu	1	0,0033
ihardetsi	1	0,0033
txartu	1	0,0033
zimelaraz	1	0,0033
erauzi	1	0,0033
esprimatu	1	0,0033
apartatu	1	0,0033
hastiatu	1	0,0033
segurtatu	1	0,0033
ikusta+araz	1	0,0033
berrargitara	1	0,0033
itsasoratu	1	0,0033
dantzatu	1	0,0033
ebatsi	1	0,0033
metatu	1	0,0033
etsaitu	1	0,0033
ahalbideratu	1	0,0033
gaztigatu	1	0,0033
sinplifikatu	1	0,0033
indultatu	1	0,0033
betatu	1	0,0033
itzulika+araz	1	0,0033
kautotu	1	0,0033

(jarraipena hurrengo orrialdean)

Aditza	Agerpenak	Portzentajea %
laxatu	1	0,0033
erreformatu	1	0,0033
ezagut+araz	1	0,0033
ailagatu	1	0,0033
barautu	1	0,0033
sufri+araz	1	0,0033
bilakaratu	1	0,0033
xigortu	1	0,0033
enkargatu	1	0,0033
akordatu	1	0,0033
goititu	1	0,0033
hormadiratu	1	0,0033
birmolda	1	0,0033
zangopilatu	1	0,0033
jariatu	1	0,0033
desesperatu	1	0,0033
modulatu	1	0,0033
belztu	1	0,0033
blindatu	1	0,0033
zedarritu	1	0,0033
jazarri	1	0,0033
desprestigiatu	1	0,0033
esmaltatu	1	0,0033
gainbegiratu	1	0,0033
amortizatu	1	0,0033
txigortu	1	0,0033
kualifikatu	1	0,0033
erdi-isolatu	1	0,0033
mehetu	1	0,0033
galdezkatu	1	0,0033
pattaltu	1	0,0033
haustu	1	0,0033
iltzatu	1	0,0033
berrekin	1	0,0033
itsusitu	1	0,0033
militatu	1	0,0033

(jarraipena hurrengo orrialdean)

Aditza	Agerpenak	Portzentajea %
normaldu	1	0,0033
herdoildu	1	0,0033
balankatu	1	0,0033
ingresatu	1	0,0033
usatu	1	0,0033
egonkortu	1	0,0033
erregutu	1	0,0033
hirukoiztu	1	0,0033
inspiratu	1	0,0033
desarroilatu	1	0,0033
elikatu	1	0,0033
enroilatu	1	0,0033
gomitatu	1	0,0033
bedeinkatu	1	0,0033
momifikatu	1	0,0033
formalizatu	1	0,0033
alokatu	1	0,0033
idortu	1	0,0033
beR+egiN	1	0,0033
transkribatu	1	0,0033
boikotatu	1	0,0033
jendeztatu	1	0,0033
detektatu	1	0,0033
harilkatu	1	0,0033
auziperatu	1	0,0033
berenganatu	1	0,0033
bukaraz	1	0,0033
doblatu	1	0,0033
kunbiatu	1	0,0033
garaitu	1	0,0033
penatu	1	0,0033
sentiaraz	1	0,0033
zuritu	1	0,0033
eden	1	0,0033
lapurteratu	1	0,0033
minorizatu	1	0,0033

(jarraipena hurrengo orrialdean)

Aditza	Agerpenak	Portzentajea %
ahantzaraz	1	0,0033
herrimindu	1	0,0033
inhibitu	1	0,0033
demokratizatu	1	0,0033
leherraraz	1	0,0033
ohoratu	1	0,0033
praktikatu	1	0,0033
hitzeman	1	0,0033
interbenitu	1	0,0033
zafratu	1	0,0033
entregatu	1	0,0033
federatu	1	0,0033
ondulatu	1	0,0033
atertu	1	0,0033
begiztatu	1	0,0033
bistaratu	1	0,0033
mailakatu	1	0,0033
frakasatu	1	0,0033
atzitu	1	0,0033
irradiatu	1	0,0033
bordatu	1	0,0033
diagnostikatu	1	0,0033
koskortu	1	0,0033
marrazkitu	1	0,0033
abokatu	1	0,0033
mutildu	1	0,0033
gaitz_esan	1	0,0033
paratu	1	0,0033
sarraskitu	1	0,0033
isolatu	1	0,0033
mentalizatu	1	0,0033
nazionalizatu	1	0,0033
pentsaraz	1	0,0033
dedikatu	1	0,0033
geraraz	1	0,0033
pobretu	1	0,0033

(jarraipena hurrengo orrialdean)

Aditza	Agerpenak	Portzentajea %
baliotu	1	0,0033
izpiritualtu	1	0,0033
kidetu	1	0,0033
ezeztatu	1	0,0033
aitzinaraz	1	0,0033
asebete	1	0,0033
presentatu	1	0,0033
bat-egin	1	0,0033
interpretatu	1	0,0033
birgizartera	1	0,0033
entzerratu	1	0,0033
lokartu	1	0,0033
ongarritu	1	0,0033
atseden	1	0,0033
prometatu	1	0,0033
hurreratu	1	0,0033
jastatu	1	0,0033
deskontsolatu	1	0,0033
konfirmatu	1	0,0033
eskiatu	1	0,0033
mozorrotu	1	0,0033
aurkatu	1	0,0033
zibilizatu	1	0,0033
jiraraz	1	0,0033
diktatu	1	0,0033
marroitu	1	0,0033
abusatu	1	0,0033
apaiztu	1	0,0033
azkandu	1	0,0033
ikusitu	1	0,0033
italiartu	1	0,0033
zokoratu	1	0,0033
kakanahastu	1	0,0033
laboratu	1	0,0033
erkatu	1	0,0033
perfilatu	1	0,0033

(jarraipena hurrengo orrialdean)

Aditza	Agerpenak	Portzentajea %
baieztu	1	0,0033
obsesionatu	1	0,0033
gerizatu	1	0,0033
sosegatu	1	0,0033
inplementatu	1	0,0033
uzkaili	1	0,0033
desageR+araz	1	0,0033
kimatu	1	0,0033
leitu	1	0,0033
erremediatu	1	0,0033
aitzineratu	1	0,0033
asimilatu	1	0,0033
biribilkatu	1	0,0033
desegonkorktu	1	0,0033
komeni_izan	1	0,0033
emititu	1	0,0033
entzuN+araz	1	0,0033
erromaratu	1	0,0033
mobilizatu	1	0,0033
findu	1	0,0033
ordenatu	1	0,0033
huts_egin	1	0,0033
behetitu	1	0,0033
zaticatu	1	0,0033
jaurtiki	1	0,0033
maneiatu	1	0,0033
berdetu	1	0,0033
bortxaz_hartu	1	0,0033
zikindu	1	0,0033
dirdiratu	1	0,0033
kontsolatu	1	0,0033
matxinatu	1	0,0033
adiskidetu	1	0,0033
aparkatu	1	0,0033
azpimarkatu	1	0,0033
ikusmiratu	1	0,0033

(jarraipena hurrengo orrialdean)

Aditza	Agerpenak	Portzentajea %
berrantolatu	1	0,0033
uhinkaratu	1	0,0033
ebalatu	1	0,0033
mesfidatu	1	0,0033
nerbiostu	1	0,0033
areagoaktu	1	0,0033
pintatu	1	0,0033
urrinzatu	1	0,0033
deformatu	1	0,0033
lautu	1	0,0033
errebelatu	1	0,0033
odoleztatu	1	0,0033
populatu	1	0,0033
higi+araz	1	0,0033
desagertaraz	1	0,0033
kiskali	1	0,0033
goibeldu	1	0,0033
batzartu	1	0,0033
taularatu	1	0,0033
jaregin	1	0,0033
luberritu	1	0,0033
gortu	1	0,0033
pujatu	1	0,0033
trafikatu	1	0,0033
irazi	1	0,0033
blaitu	1	0,0033
konplikatu	1	0,0033
kontsumitu	1	0,0033
amorratu	1	0,0033
ozpindu	1	0,0033
handi-handitu	1	0,0033
salbuetsi	1	0,0033
berebaitaratu	1	0,0033
irudiztatu	1	0,0033
disimulatu	1	0,0033
erbesteratu	1	0,0033

(jarraipena hurrengo orrialdean)

Aditza	Agerpenak	Portzentajea %
apetatu	1	0,0033
patruilatu	1	0,0033
azpitudatu	1	0,0033
ikustaraz	1	0,0033
zorroztu	1	0,0033
eroraraz	1	0,0033
neutralizatu	1	0,0033
gehiagotu	1	0,0033
solastatu	1	0,0033
infiltratu	1	0,0033
itzuri	1	0,0033
ken+araz	1	0,0033
laztandu	1	0,0033
errefusatu	1	0,0033
ailega+araz	1	0,0033
postulatu	1	0,0033
xurgatu	1	0,0033
desarautu	1	0,0033
kobinatu	1	0,0033
elbarritu	1	0,0033
errepasatu	1	0,0033
eztandaraz	1	0,0033
aktibatu	1	0,0033
onbideratu	1	0,0033
tentatu	1	0,0033
birsor	1	0,0033
zangozabatu	1	0,0033
desgizarterattu	1	0,0033
konbokatu	1	0,0033
erabilaraz	1	0,0033
flotatu	1	0,0033
gosetu	1	0,0033
atzeraraz	1	0,0033
beR+aukera	1	0,0033
transferitu	1	0,0033
esna+araz	1	0,0033

(jarraipena hurrengo orrialdean)

Aditza	Agerpenak	Portzentajea %
pagatu	1	0,0033
txikiagotu	1	0,0033
diziplinatu	1	0,0033
erdibitu	1	0,0033
mekanizatu	1	0,0033
esterilizatu	1	0,0033
nahigabetu	1	0,0033
apoiatu	1	0,0033
pausatu	1	0,0033
uleR+!taraz	1	0,0033
itsutatu	1	0,0033
datatu	1	0,0033
ahalmendu	1	0,0033
plantatu	1	0,0033
herrestatu	1	0,0033
sor+araz	1	0,0033
betikotu	1	0,0033
usnatu	1	0,0033
ketu	1	0,0033
lehenetsi	1	0,0033
erreklamatu	1	0,0033
sumindu	1	0,0033
zabartu	1	0,0033
desberdindu	1	0,0033
kodetu	1	0,0033
entenditu	1	0,0033
errepresentatu	1	0,0033
falta_izan	1	0,0033
ateraraz	1	0,0033
produzitu	1	0,0033
iraitzi	1	0,0033
eraendu	1	0,0033
luzitu	1	0,0033
formulatu	1	0,0033
alteratu	1	0,0033
orokortu	1	0,0033

(jarraipena hurrengo orrialdean)

Aditza	Agerpenak	Portzentajea %
atzerriratu	1	0,0033
beR+eraiki	1	0,0033
transmititu	1	0,0033
deuseztu	1	0,0033
erantzi	1	0,0033
espainiaratu	1	0,0033
musukatu	1	0,0033
paralizatu	1	0,0033
txoratu	1	0,0033
zirimolatu	1	0,0033
judikatu	1	0,0033
doitu	1	0,0033
kurritu	1	0,0033
erdiratu	1	0,0033
menperatu	1	0,0033
nardatu	1	0,0033
pentsa+araz	1	0,0033
sentiberatu	1	0,0033
uniformatu	1	0,0033
euskaratu	1	0,0033
sorrarazi	1	0,0033
inketatu	1	0,0033
demonizatu	1	0,0033
basitu	1	0,0033
birbizi	1	0,0033
desblokeatu	1	0,0033
kolaboratu	1	0,0033
likidotu	1	0,0033
aldararaz	1	0,0033
onetsi	1	0,0033
goratu	1	0,0033
atundu	1	0,0033
promenatu	1	0,0033
hurrendu	1	0,0033
irakin	1	0,0033
biziraun	1	0,0033

(jarraipena hurrengo orrialdean)

Aditza	Agerpenak	Portzentajea %
zarramazkatu	1	0,0033
deskonposatu	1	0,0033
konfesatu	1	0,0033
1.211	30.740	% 100

B.1 taula: EPEC-DEP corpuseko aditzen maiztasunak.

