

EPEC-RolSem: Ingelesezko PropBank-VerbNet eredura etiketatutako euskarazko corpora. Erabakiak, egokitzapenak eta bereizgarriak

Ainara Estarrona^a Izaskun Aldezabal^b Arantza Díaz de Ilarraza^a Maria Jesús Aranzabe^b

^aHizkuntza eta Sistema Informatikoak/ UPV/EHU

^bEuskal Hizkuntza eta Komunikazioa

ainara.estarrona@ehu.es izaskun.aldezabal@ehu.es

a.diazdeillaraza@ehu.es maxux.aranzabe@ehu.es

1. Sarrera¹

Artikulu hau hizkuntza naturalaren prozesamenduaren alorrean kokatzen da eta zehazki IXA Taldearen² jardunaren barruan. Azkenaldian hain modan dagoen *corpus based* edo *corpus driven linguistic* delakoen barruan kokatzen gara eta zehazki, corpusetan oinarritutako lexikografian. Kontu modernoegiak, beharbada, Henrikeren lan-estilorako, baina ezagutzen genuenok ondo dakigu bere lan-jardunaren ardatz nagusiak jakin-mina, lana eta ikerkuntza izan zirela beti, beraz, bide horretan aurrera egindako pauso oro onetsiko lukeela ez dugu zalantzarik.

IXA Taldean euskararen prozesamendurako erreferentziatzko corpus bat (EPEC-*Euskararen Prozesamendurako Erreferentzia Corpora*) eraikitzen ari gara (Aduriz *et al.* 2006). Corpus osoa, eskuz, hiru mailatan etiketatu dugu oraindaino: i) segmentazio mailan (EPEC-SEG (Aldezabal *et al.* 2007)); ii) sintaxi mailan Dependentsia Gramatikaren eredua (Tesnière 1959) jarraituz (EPEC-DEP (BDT) (Aranzabe 2008; Aldezabal *et al.* 2009)); eta iii) semantika mailan (Eusemcor), Euskal WordNet-eko adierak erabiliz eta soilik izenei gagozkiela (Pociello *et al.* 2010). Gure helburua orain, corpusari beste etiketatze-geruza bat gehitzea da. Argumentu/adjuntu hautagai diren elementuen gainean, predikatu mailako etiketatzea egingo dugu, hau da, aditzen adierak, balentziak, rol semantikoak eta, zenbaitetan, hautapen-murriztapenak ere, etiketatuko ditugu; horri deitu diogu EPEC-RolSem (Aldezabal *et al.* 2010a, 2010b, 2010c). Mota honetako baliabide semantikoak ezinbestekoak dira hizkuntzaren prozesamendu automatikoan ikertzen diren hainbat atazatan, hala nola, sintaxi mailako desanbiguazioa edo hizkuntzaren ulermena, eta era berean, aplikazio aurreratuen

¹ Lan hau Euskal Herriko Unibertsitateak (GIU09/19), Eusko Jaurlaritzak (IT344-10, S-PE11UN098, IE09-262) eta Espainiako Gobernuako Zientzia eta Berrikuntza Ministeritzak (FFI2008-02805-E/FILO, FFI2009-06497-E) emandako diru-laguntzari esker egin da.

²<<http://ixa.si.ehu.es/Ixa>>

(itzulpen automatikoa, galderak sortzeko sistemak, etab.) emaitzak hobetzeko erabil daitezke. Hizkuntzalaritza teorikoan, aldiz, fenomeno linguistikoen azterketarako balia daitezke horrelako baliabideak eta baita hizkuntzen arteko aldeak aztertzeko ere.

EPEC-RolSem corpora osatuz joan ahala eta etiketatze-lanean lortutako informazioa oinarri hartuta, euskal aditzen lexikoa ere osatu dugu: *Basque Verb Index* (BVI) (Aldezabal *et al.* 2013). Esan beharra dago, alor honetan beste hizkuntzetan jarraitzen den joera nagusia jarraitu dugula guk ere, hau da, etiketatutako corpusetatik lexikoiak sortzea; esaterako, *Penn Treebank* (Marcus 1994) eta PropBank (Palmer *et al.* 2005a) corpusak VerbNet lexikoiarekin lotzen dira (Kingsbury & Palmer 2002) eta berdin *Prague Dependency Treebanka* (Hajic 1998) Vallex lexikoiarekin (Hajic *et al.* 2003).

Etiketatzeko lan hori aurrera eramateko eredu bat aukeratzea beharrezkoa da eta aurrerago azalduko ditugun arrazoiak direla eta, PropBank-VerbNet (aurrerantzean PB-VN) eredu izan da gure hautua, Aldezabalek (2004) bere tesian egindako Euskal Aditzen Datu Baseko (EADB) aditzen azterketa oinarri hartuta.

Artikulu honetan, euskarazko EPEC-RolSem corpora etiketatzean hartu ditugun erabakiak, aukeratutako ereduari egindako egokitzapenak eta bi hizkuntzen arteko berezitasunak azalduko ditugu; eta, batez ere, azken honetan sakonduko dugu. Hala, 2. atalean, corpora etiketatzeko aukeratutako eredu aurkeztuko dugu labur; 3. atalean, etiketatze-lana aurrera eramateko erabili ditugun baliabide nagusiak azalduko ditugu; 4. atalean, aukeratutako ereduari egindako egokitzapenei buruz hitz egingo dugu; 5. atalean, corpora etiketatzean ingelesarekiko agertu zaizkigun bereizgarriak azalduko ditugu eta egoera bakoitzaren aurrean hartutako erabakiak aurkeztuko ditugu eta bukatzeko, 6. atalean, etiketatze-lan horren guztiaren emaitzak aurkeztuko ditugu.

2. Hautatutako eredu

EPEC corpora predikatu mailan etiketatzeko aukeratu dugun eredu PB-VN izan da (Agirre *et al.* 2006a) eta honako hiru arrazoietan oinarritu gara eredu aukeratzeko:

Abiapuntutzat sintaktikoki etiketatutako corpora du.

- a) Beste hizkuntza batzuetarako ere erabili izan da eredu hori: hindia (Bhatt *et al.* 2009), txinera (Palmer *et al.* 2005b; Xue 2008; Xue & Palmer 2009), koreera (Palmer *et al.* 2006), arabiera (Palmer *et al.* 2008), gaztelania (Aparicio 2007) eta katalana (Civit *et al.* 2005).
- b) Taldean alde aurretik egindako lanen antzeko ildoak proposatzen dira aditzak tratatzeko (Aldezabal 2004).

Hasiera batean, beste hizkuntzetan egiten zuten bezalaxe, PB eredia bakarrik hartu genuen jomugan, baina gerora, ikusi dugu, Merlo & Van der Plan-ekin (2009) bat eginez, VNk informazio osagarria ematen duela. VNk aditzetan rol semantikoak orokortzeko aukera ematen du eta PBk, aldiz, hobeto jasotzen ditu rol semantiko horiek gauzatzen diren egitura sintaktikoen arteko desberdintasunak. VN-n erabiltzen den rol semantikoaren multzoa, hizkuntzalaritzan, orokorrean, erabili izan denaren antzekoa da (eta bide batez, Aldezabalek (2004) erabiltzen duenaren antzekoa): *Agent, Theme, Cause, Patient...*; PBk, ordea, aditz guztientzat balentzia berberak eta aditz bakoitzeko rol espezifikoak erabiltzen ditu. PB lotuago dago funtzio gramatikalekin eta aproposagoa da ikasketa automatikorako, baina VNk informazio semantiko zehatzagoa ematen digu; beraz, argi dago biak erabilia lortuko dugula hasieran aipatu dugun helburua, hots, aditzaren informazio sintaktiko-semantiko osoa baliabide bakar batean biltzea. Jarraian, PB corpus etiketatuaren eta VN lexikoia-aren ezaugarri nagusiak azalduko ditugu.

2.1. PropBank eredia

Paul Kingsbury eta Martha Palmer jo ohi dira PB proiektuaren³ aintzindari (Kingsbury & Palmer 2002; Palmer *et al.* 2005a) gisa. Ingeleseko corpus bat semantikoki etiketatzeko proiektua da eta oinarrian *Penn Treebank*eko (Marcus *et al.* 1993) zuhaitz sintaktikoak ditu. Labur-labur esanda, etiketatzean elkarren artean independente diren 2 maila bereizten dituzte:

- a) Argumentu zenbakitua edo balentzia (0tik 4ra: Arg0, Arg1, ..., Arg4) eta adjuntuak (adjuntuek ez dute zenbakirik; besterik gabe modifikatzailearen M-a (ArgM)).
- b) Rol semantikoak.

Hots:

- a) Argumentu batek rol semantiko desberdinak har ditzake (adib., aditzaren edo aditz-adieraren arabera Arg1 *Theme* edo *Topic* izan daiteke; eta orobat Arg0 *Agent* edo *Theme*).
- b) Era berean, rol semantiko bera argumentu desberdinetan ager daiteke (*Theme* batzuetan Arg1 da eta beste batzuetan Arg0).

Adibide batekin hobeto ulertuko dugu PB eredia. Har dezagun PB datu-basetik *to tell* aditza eta horren *tell.01* aditz-adiera (1a) —*tell.02* eta *tell.03* aditz-adierak ere baditu *to tell* aditzak—. Ondoren ikus dezagun *tell.01* aditz-adierari dagokion adibide etiketatu bat (1b). Ikusten dugunez, *tell.01* aditz-adierak hiru argumentu ditu, baina horretaz gain, PBn aditz-

³ Proiektuari buruzko xehetasun gehiagotarako: <<http://verbs.colorado.edu/~mpalmer/projects/ace.html>>.

adiera bakoitzak bere gauzaren sintaktikoak ditu (*frameak* deitzen dituzte; azken batean, alternantziak (ikus 3.1 ere). *tell.01*-i lotutako *frameak* edo alternantzia sintaktikoak (1c)-koak dira.

(1) a. rolset⁴ *tell.01*:

→Arg0 *speaker*

→Arg1 *utterance*

→Arg2 *hearer*

b. 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

c. *tell.01*

i. Ditransitive (-)

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

ii. Odd ditransitive

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

Arg0: this

REL: tells

Arg2: us

Arg1: *trace* is that US trade law is working

iii. Prepositional arg2

ohn told a story to Mary

Arg0: John

REL: told

Arg1: a story

⁴ *rolset* esaten diote aditz-adiera bakoitzaren egitura sintaktiko-semantikoari.

Arg2: to Mary

iv. Fronted

[*"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*

Goazen orain VNi buruz zerbait esatera.

2.2. VerbNet

VerbNet (VN),⁵ lexikoi bat da eta Levin-en (1993) aditz-sailkapenean dago oinarrituta; horrela izanda, aditz bakoitzari Levinen klase semantikoetako bat esleitzen zaio: 9.1, 9.2, 10.1... Hala ere, esan behar da aditzak berraztertu ahala, sailkapen xehakatuago bat egiten dutela zenbait kasutan. Adib., Levinen 37.1 klasea (*Verbs of Transfer of a Message: ask, cite, demonstrate...*), bi azpiklasetan banatu dute VN-n: i) 37.1.1⁶ (*transfer_mesg: communicate, demonstrate, explain, narrate, recite...*) eta ii) 37.1.2 (*inquire: ask, consult, enquire, inquire, pry*). Beraz, oinarrian Levin (1993) dute, baina beren moldaketak ere egin dituzte. Horretaz gain, hautapen-murriztapenak ere ageri dira zenbaitetan: \pm animate, \pm organization, \pm communication, \pm concrete, \pm location, \pm region, \pm animal... Gauzak horrela, bertsio desberdinak ateratzen doaz, etengabe klase berriak gehituz (Kipper *et al.* 2006, 2008). Memento honetan 3.2 bertsioan daude.

Esan dugun bezala, erabiltzen dituzten rolak orokorrak dira (adib.: *Agent, Theme, Beneficiary, Topic, Patient...*) eta hasiera batean behintzat klaseak errepresentatzeko pentsatuak dira (Levinen klaseak oinarrian bederen). Ikus dezagun *to spend* aditzaren adibidea. Aditz honek bi adiera ditu PBn (2a,b) bakoitza Levin klase bati lotua:

(2) a. *spend.01*: dirua erabiltzea (*use of money: pay-68* klasea)

b. *spend.02*: denbora pasatzea edo itxoitea (*bide time: consume-66-1* klasea).

1. irudian, bi klase hauetan VN-n dagoen rolei buruzko informazioa ikus dezakegu.

⁵ Xehetasunetarako jo <http://verbs.colorado.edu/~mpalmer/projects/verbnet.html> helbidera.

⁶ Era berean 37.1.1 klasea beste bi azpisailetan banatzen dute: i) 37.1.1-1 (*dictate, quote, read*) eta ii) 37.1.1-1-1 (*show, teach, tell, write*).

pay-68
Agent [+animate & +organization]
Asset [+currency]
Theme

consume-66-1
Agent [+animate & +organization]
Asset
Goal

1. Irudia. 68 eta 66-1 klaseek VN-n duten rolei buruzko informazioa

Guk gure lanean, hain zuzen, VN 3.2 lexikoiko rolak erabiliko ditugu, eta rol horiek PBko aditz-adierei lotuta ulertuko ditugu (aditz-adiera deitzen diogu PBko sarreretan agertzen diren 01, 02, 03... horiei).

Gaur egun, PB eta VN lotuak daude SemLink (Palmer 2009) proiektuari esker eta PBko datu-basea begiratzen dugunean, zuzenean ikusten dugu aditz-adierari dagokion VNko informazioa ere. SemLink proiektuan hainbat baliabide lotu dira: PropBank, VerbNet, FrameNet⁷ (FN), WordNet⁸ eta OntoNotes, eta sarean ikusgarri dago Unified Verb Index-i⁹ esker.

3. Aldez aurretik erabilitako baliabide nagusiak

Atal honetan, corpora etiketatzen hasi aurretik esku artean izan ditugun baliabide nagusiak azalduko ditugu, artikuluko hurrengo atalak ulertzeko ezinbestekoak iruditu zaizkigunak.

3.1. Euskal Aditzen Datu Basea (EADB)

EADBren diseinua eta edukia Aldezabalen (2004) tesian azaltzen da sakonki. Aldezabalek bere lanean euskarazko aditzak lantzeko proposamena egiten du EPEC corpora baliatuta eta Levin-en (1993) eta Vázquez *et al.*-en (2000) teoria eta irizpideak kontuan hartuta. Jarraian, proposamen horren oinarriak azalduko ditugu labur.

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

⁷ <<https://framenet.icsi.berkeley.edu/fndrupal/home>>

⁸ <<http://wordnet.princeton.edu/>>

⁹ <<http://verbs.colorado.edu/verb-index/>>

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 oinarritzkoak, predikatu-mota orokorrak, sailak osatzeko baliagarri izan daitezkeenak. (Aldezabal, 2008: 38)

Ikusten den bezala, egitura usu bat edo batzuk ager ditzake aditzak balio bera adierazteko. Hain zuzen, balio bera adierazten duten egitura sintaktiko desberdin guztiak, elkarren artean alternantzia dira. Aldiz, balio desberdina adierazten duten egiturek elkarren artean ez dituzte alternantziak osatzen, besterik gabe, delako aditzari dagozkion egitura sintaktiko desberdinak dira. Alternantzia osatzen duten egituretako bakoitza aldaera izango da; hain zuzen, alternantzia baten aurrean gaudenean, nahitaez bi aldaera jarri beharko ditugu harremanetan. Adibidez, *esan* aditzak bi balio semantiko ditu (i eta ii), baina hiru balio sintaktiko-semantiko (bss) (3a, 3b, 4), eta lehen balio semantikoan markatuta dauden bi bss-ak (3a, 3b) elkarren artean alternantzia dira:

i. Adierazpenezko jarduera

(3) a. Ez diogu ezer esango aitari¹⁰

DU-1: esperimentatzailea [+giz]_ERG; ¹¹ gaia [-konkr]_ABS

b. Beste bizitzaz hitz egiterakoan, han ere bere amorantea izango dela esango dio

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

ii. Ezaugarri-esleitzea

(4) Honi ere amapontekoa esaten zion gure amak

DIO-3: abiapuntua [+giz]_ERG; helburua_DAT; ezaugarria_ABS

Balio kognitibo orokor bakoitza definitzeko beharrezkoak diren rolak¹² erabiltzen ditu Aldezabalek eta ondoren, rol horiek modu koherente batean konbinatuta predikatu-mota orokorrak¹³ planteatzen ditu. Adibidez:

i. Entitate baten elkarraldatzea

¹⁰ Adibideak EPEC corpusetik ateratakoak dira, bestelakorik esaten ez bada.

¹¹ Argumentuen gauzpen sintaktikoa postposizio-atzizkietan oinarritzen da. ERG: ergatiboa; ABS: absolutiboa; DAT: datiboa; ELA_KONP: konpletiboa.

¹² Rolak EADBn 21 dira: gaia, gai ukitua, gai sortua, egoera, kokapena, denbora, helburuko kokapena, helburuko egoera, abiapuntuko kokapena, bidea, abiapuntua, helburua, esperimentatzailea, kausa, iturria, edukitzailea, edukia, ezaugarria, jarduera, neurria eta modua.

¹³ Predikatu-mota orokorrak 13 dira: egoera-aldaketa, kokapen-aldaketa, aldaketa, sortzea, jarduera, elkarraldatzea, edukitzea, ezaugarria esleitzea, izatea (existitzea), kokapena, egoera, deskribapena eta ustezko adierazpena.

abiapuntua+gaia+helburua (eman): Foru Erkidegoko Parlamentuak “terrorismoaren biktimei” Nafarroako domina ematea salatu dute

ii. Entitate bati ezaugarria esleitzea

abiapuntua+ezaugarria+helburua (esan): Honi ere amapontekoa esaten zion gure amak

iii. Entitate bat edukitzea

edukitzailea+edukia (izan): Eibarrek ere bere aukerak izan zituen

edukitzailea+edukia+kokapena (ekarri): Idazlearen lan guzien izenen zerrenda, bukaeran dakar [liburuak?]

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 (adib. *jan*: jarduerazko aditza). Ordea, aditz batzuek predikatu orokor bat baino gehiago adieraz dezakete, eta horiei loturiko alternantziak ere izan ditzakete. Horrelakoetan sortzen dira guztiak elkarren artean bereizteko zailtasunak (eta beharrak, aldi berean); eta aditz batzuk mota semantiko batekotzat jotzen badira ere (adib. *eman*: elkarraldatzeazko aditza; *ekarri*: kokapen-aldaketazkoa), bestelako predikatu motetan ere parte hartzen dute (adib. *eman*: jarduerazko predikatu motan; *ekarri*: sortzeazko eta eduteazko predikatu motan).

3.2. EPEC-DEP (BDT)

EPEC corpora predikatu mailan etiketatzeke, dependentzietekin etiketatutako corpusean oinarritu gara (EPEC-DEP (BDT));¹⁴ Aldezabal *et al.* 2009), aditzaren mendean agertzen den osagai orori *arg_info* etiketa (argumentu-informazioa) semantikoa esleitzeko. EPEC-DEP zuhaitz-bankuan, 300.000 hitz etiketatu dira eskuz Dependentsia Gramatikaren Teoria (Tesnière 1959) jarraituz. Teoria horretan, esaldiko hitzak binaka lotuz esaldiaren zuhaitz sintaktikoa (dependentsia-zuhaitza ere deitua) lortzen da. 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 (Aranzabe 2008) bidez. Dependentsietekin etiketatutako corpus horretan 1.457 aditz eta 31.639 agerpen daude guztira. Gure lan honetan 30 agerpenetik gora duten aditzak landu ditugu.

¹⁴ <<http://ixa.si.ehu.es/Ixa/Produktuak/1306407157>>

3.3. Etiketatzeko tresna eta etiketa: AbarHitz eta *arg_info*

Rol sematikoak etiketatzeko erabili dugun tresna etiketatze sintaktikorako erabilitako bera da: AbarHitz. Ikus Díaz de Ilarraza (*et al.* 2004) eta Aldezabal *et al.* (2010b), tresnari eta interfazeari buruzko xehetasunetarako, hurrenez hurren. Hemen xehe azalduko duguna etiketatzeko erabili dugun etiketa izango da. Eredua hautatuta, etiketatzerakoan informazioa nola formalizatu da hurrengo pausoa. Formalizazio horretan *arg_info* etiketa sortu dugu eta honako zazpi eremuak ditu: *arg_info*: (*VNrs*, *aditza*,¹⁵ *hitz aztergaia*, *BAL*, *VNrol*, *EADBrol*, *HM*). Adibide hipotetiko batez (5), jarraian, eremu bakoitzaren azalpena emango dugu.

(5) *arg_info*: (tell_01/say_01, esan, nik, arg0, agent, esperimantatzailea, [+giz])

- *VNrs* (PB aditza eta *rolset* zenbakia): PBko aditza eta horri dagozkion *rolset* zenbakientzako eremua. Askotan bat baino gehiago egongo da, adibidez, *tell_01/say_01*, *02*, *03*, *04*, *05*, *06*, *07*... Ingeleseko ordain desberdin bakoitzeko (edo ordain multzo desberdin bakoitzeko) euskarazko adiera bat izango dugula suposatuko dugu. Adibidez: *esan_01 = tell_01/say_01*; *esan_02 = call_01*.
- *Aditza*: lantzen ari garen aditza (burua, alegia), dependentzia-etiketarik kopiatua.
- *Hitz aztergaia*: argumentu/adjuntu hautagaia, dependentzia-etiketarik kopiatua.
- *BAL*(entzia): argumentu edo adjuntuentzako eremua. Balioak ondokoak dira: Arg0, Arg1, Arg2, Arg3, Arg4, ArgM. Aztergai den unitateari horietako zein dagokion zehazten da.
- *VNrol* (VerbNet rola): VNri dagokion rola jartzen da eremu honetan.
- *EADBrol* (Euskal Aditzen Datu Baseko rola): EADBko rola ezartzeko eremua.
- *HM* (Hautapen-murriztapenak): hautapen-murriztapenak jartzeko eremua. Balioak, printzipioz, ondorengoak dira: [+biz], [-biz], [+konkr], [-konkr], [+giz], [-giz].

Etiketa hau aditzaren mende dagoen dependentzia-etiketa bakoitzaren ondoren erantsiko da, 2. irudian ikus daitekeen bezala:

(6) Hau esanik, gartzelak, bi modutan sailkatuko genituzke:

¹⁵ Bigarren eta hirugarren elementuek *arg_info* etiketa modu independentean erabiltzeko aukera ematen dute.

ncobj (abs, esanik[w305], Hau[w304], Hau[w304], obj)
arg_info (tell_01/say_01, esanik[w305], Hau[w304], arg1, topic, gaia, -konkr)

xmod (mod, sailkatuko[w311], esanik[w305], esanik[w305])
arg_info (classify_01, sailkatuko[w311], esanik[w305], argM, MNR, modua, -)

ncobj (abs, sailkatuko[w311], gartzelak[w307], gartzelak[w307], obj)
arg_info (classify_01, sailkatuko[w311], gartzelak[w307], arg1, Theme, gai, -biz)

detmod (- , modutan[w310], bi[w309])

ncmod (ine, sailkatuko[w311], modutan[w310], modutan[w310])
arg_info (classify_01, sailkatuko[w311], modutan[w310], argM, MNR, modua, -)

auxmod (- , sailkatuko[w311], genituzke[w312])

2. irudia. (6) adibidearen dependentzia-etiketak dagozkien *arg_info* etiketekin

Etiketatzeko-prozesuan laguntzeko bestelako baliabide batzuk ere erabili ditugu, hala nola, Levinen aditz guztien zerrenda euskaratua (Aldezabal 2010), Levinen aditzen itzulpena PBrekin gurutzatuta eta corpusarekin lotuta, eta EADBko 100 aditzen informazioaren esleipen automatikoa postposizio-atzizkietan oinarrituta. Beroietan sakontzeko, jo bedi Aldezabal *et al.* 2013ra.

4. PB-VN eredia erabiltzean hartutako erabaki orokorrak eta ereduari egindako egokitzapenak

PB-VN eredia euskarara aplikatzean, hainbat egokitzapen eta moldaketa egin behar izan ditugu. Atal honetan bi motatako moldaketak azalduko ditugu; alde batetik, eredia erabiltzean hartu ditugun erabaki nagusiak (4.1 eta 4.2 atalak) eta bestetik, euskarara egokitzean egin behar izan ditugun moldaketak (4.3 eta 4.4 atalak).

4.1. Arg0/Arg1 aukeratzen

PBk bi maila independente bereizten ditu: argumentuak (edo adjuntuak) eta rolak, baina gero ikusten dugu etiketatzean Arg1 beti dela *Theme* eta Arg0 *Agent*. Ez dirudi arrazoi linguistiko garbirik dagoenik hori horrela izateko. Ikus jarraian Kingsbury & Palmer-ek (2003:3) ematen duten argumentua:

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:

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

Figure 1.

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 (2006:76): “In *John and Mary come* the NP *John and Mary* is a constituent in Treebank and it is also marked as ‘Arg0’ in PropBank”. Baina PBn zuzenean begiratuta, 3. irudian agertzen den informazioa topatzen dugu:

<i>come.01</i>
roles:
Arg1: entity in motion (<i>theme</i>)
Arg2: extent
Arg3: start point
Arg4: end point

3. irudia. *come.01* aditza PBn

Inkonsistentzia horiek ikusita, PBko gidalerroetan esaten dutenari jarraiki, gure erabakia izan da bi mailen arteko independentzia mantentzea; beraz, ez ditugu automatikoki Arg0 *Agent* gisa eta Arg1 *Theme* gisa etiketatu. Gure ustez, *kokapen-aldaketa* adierazten duten aditz iragangaitzeko subjektua aldi berean da ekintza burutzen duen entitatea eta ekintza hori jasotzen duena (Vázquez *et al.* 2000). Beraz, aditz horien subjektuak Arg0 balentziarekin

etiketatu ditugu. Erabaki hau PBko 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* bezalako aditz inakusatibo baten kasuan, zeinetan aldaera intrantsitiboa baino ez den posible, ekintza gauzatzen duena eta jasotzen duena entitate bera direla ulertzen dugu eta beraz, Arg0_Theme etiketak jarriko dizkiogu. PBn, aldiz, mota horretako aditzen subjektuak Theme gisa etiketatzen dituzte, baina Arg1 balentzia jartzen diete. Gure ustez, 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). Bestalde, *apurtu* bezalako aditz kausatibo/inkoatiboetan, Theme rolari beti jarriko diogu Arg1 balentzia, iruditzen zaigulako esaldian agertu ez arren, ekintzaren “kausa” (Arg0) beti existitzen dela.

Esan beharra dago PB eredia erabili duten beste hizkuntzetan Arg0_Agent, Arg1_Theme irizpidea jarraitu dutela.

4.2. VNk bi klase edo gehiago esleitzen dituenan

VNk zenbait kasutan aditz-adiera bati Levinen klase bat baino gehiago esleitzen dizkio. Jar dezagun *separate.01* aditz-adieraren adibidea, arazoa hobeto ulertzeko:

separate.01, separate, vncls: 23.1-2, 10.1, 29.10:

roles:

Arg0: entity causing separation (vnrole: 23.1-2-Agent, 10.1-Agent, 29.10-Agent)

Arg1: thing being divided (vnrole: 23.1-2-Patient1, 10.1-Theme, 29.10-Theme)

Arg2: source, separated from (vnrole: 23.1-2-Patient2, 10.1-Source)

4. irudia. *separate.01* aditza PBn

4. irudian ikusten denez, *separate.01* aditz-adierari 3 Levin klase esleitzen zaizkio¹⁶ VN-n (23.1-2, 10.1, 29.10) eta klase bakoitzak rol desberdinak hartzen ditu; orduan, zein rolekin etiketatu behar dugu guk 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*, *gai ukitua* eta *gai ukitua* rolak agertzen ditu:

DA-1: *gai ukitua* (ABS), *gai ukitua* (ABL¹⁷)

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

Beraz, PB-VNko *Agent*, *Patient1* eta *Patient2* aukeratuko ditugu, hau da, 23.1-2 klaseari dagozkion rolak. Jarraian, eredia euskarara egokitzeke egin ditugun bi moldaketa garrantzitsuenak azalduko ditugu.

4.3. PB-VN-n euskaraz posible ez den argumentu bat agertzen da

Euskaraz posible ez diren argumentuak ez ditugu kontuan hartu eta beraz, ez ditugu etiketatu (eta jakina, ezta BVIn definitu ere). Horien artean, zenbait mugimenduzko aditzetan agertzen den *Extent* argumentua daukagu. Ikus dezagun *etorri* aditzaren adibidea. 1. taulan ikus dezakegu *come.01* aditz-adierak PB-VN eredian duen informazioa:

1. taula. *come.01* aditz-adiera PB-VN-n

<i>PB-VN: come.01</i>
Arg1: entity in motion/"comer" (theme)
Arg2: <i>extent -- rare</i>
Arg3: start point (source ¹⁸)
Arg4: end point (destination)

PBn bertan esaten dute, 1. taulan ikusten den bezala, Arg2_Extent hori ez dela ohikoa (*rare* marka du ingelesez) eta gainera, ez dute horren adibide bat bera ere ematen. VN-n, aldiz, ez

¹⁶ Hiru klaseak OntoNotes (Hovy *et al.* 2006) baliabidearen adiera berdinarekin lotzen dira, *Unified Verb Index*-en ikus dezakegun bezala: *Sense number 1: (cause to) divide or segregate into parts, sections or smaller groups.*

¹⁷ ABL: ablatiboa.

¹⁸ VN-n etengabe ari dira lexikoa eguneratzen eta azkenaldian, euren rolak LIRICS (Linguistic InfRrastructure for Interoperable resourCes eta Systems (<<http://lirics.loria.fr>>) proiektuko rol estandarrekin berdintzeko bidea hartu dute. Gure lana hasi genuen garaian, Arg3 argumentuari *Source* rola esleitzen zion VNk, baina gaur egungo datuak begiratzen baditugu, *Initial Location* jartzen duela ikusiko dugu. Guk hasierako VNko datuak jarri ditugu gure BVI lexikoian.

dute rolik ematen argumentu horrentzat, hau da, ez dute *Extent* argumentua sartu *come* aditzaren sarreran.

EADBN ere ez dugu argumentu hori *etorri* aditzaren sarreran: *etorri_1*: gai ukitua_ABS; abiapuntuko kokapena/bidea_ABL; helburuko kokapena_ALA.¹⁹ Izan ere, euskaraz ez baita posible zenbateko distantzia *etorri* den esatea:

(7) * Mikel 200 metro etorri da handik hona²⁰

Beraz, *Arg2_Extent* hori ez dugu kontuan hartu BVIko *etorri* aditzaren lehen adiera definitzerakoan, eta ondoren datozen argumentuak aurreratu egin ditugu, 2. taulan ikus daitekeen bezala:

2. taula. *etorri_1* aditz-adiera BVI lexikoian

<i>BVI: etorri_1</i>
Arg0: ²¹ <i>theme</i> , gai ukitua (ABS)
Arg1: <i>source</i> , abiapuntuko kokapena (ABL)
Arg2: <i>destination</i> , helburuko kokapena (ALA)

4.4. *Path* rola gehitu

Corpusa etiketatzen joan ahala, beharrezkoa iruditu zaigu *Path* (bidea) rol semantikoa gehitzea. Rol semantiko hori ez da agertzen VN-n, baina bai EADBN. Hualde & Ortiz de Urbinak (2003: 373) ere *pasatu* aditzaren egitura honetaz hitz egiten dute eta *ibaia pasatu dut* adibidea ematen dute.²² Hona hemen *pasatu* aditzaren adibide bat:

(8) Zure etxetik pasatu naiz gaur goizean

Kasu honetan ABLa ez da “abiapuntuko_kokapena”, baizik eta aditzak adierazten duen mugimendua egiterakoan ibiltzen den “bidea”.

5. Aditzak etiketatzerakoan aurkitu ditugun ingelesarekiko bereizgarriak eta hartutako erabakiak

Atal honetan, EADBN agertzen diren 100 aditzak etiketatzerakoan topatu ditugun arazoak aztertuko ditugu, agerian jarritz euskara eta ingelesaren artean azaldu zaizkigun aldeak.

¹⁹ ALA: adlatiboa

²⁰ Asmatutako adibidea.

²¹ Ikus 4.1 atala Arg0/Arg1 balentzia aukeratzeko irizpidearen azalpenerako.

²² “Only the path-structure *pasatu* takes a transitive auxiliary”, Hualde & Ortiz de Urbina (2003: 373).

Agertutako kasuistika eta hartutako erabakiak hobeto ulertzeko, multzo bakoitza adibideak erabiliz azalduko dugu.

5.1. EADBko eta PB-VNko informazioa bat etorri dira: *aldatu_2*

Arazorik eman ez diguten kasuetan, bat datoz ingeleseko eta euskarazko aditz-adieren argumentu-egiturak eta argumentuei esleitutako rol semantikoak. Beraz, zuzenean aplikatu ahal izan dugu euskaraz PB-VN ereduak. Horrelako kasua dugu *aldatu* aditza. Aditz honek bi adiera edo predikatu orokor ditu euskaraz, EADBn aztertutakoaren arabera beti: i) kokapenahiz egoera-aldaketa eta ii) ordezkatzeko-jarduera. Bigarren adiera itzultzeko *replace.01* aditza aukeratu dugu eta 3. taulan ikus dezakegu PB-VN-n eta EADBn duen informazioa:

3. taula. PB-VNk eta EADBk *replace.01* eta *aldatu_2* aditz-adierentzat ematen duten informazioa

<i>PB-VN: replace.01</i>	<i>EADB: aldatu_2</i>
Arg0: replacer (agent ²³)	esperimentatzailea (ERG)
Arg1: old thing (theme1)	gaia (ABS)
Arg2: new thing (theme2)	gaia (INS ²⁴)

Bietan ikusten ditugu 3 argumentu, eta rolak 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* da.

5.2. EADBko eta PB-VNko informazioa ez dira bat etorri

Arazoetako bat PB-VN-n eta EADBren arteko adostasunik eza izan da. 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, kasu onargarrien artean daudenean. Adibide batekin hobeto ulertuko dugu: *lortu* aditza (4. taula):

²³ Parentesi artekoa VNko rola da.

²⁴ INS: instrumentala.

4. taula. PB-VN-n eta EADBn *obtain.01* eta *lortu_1* aditz-adierentzat ematen den informazioa

<i>PB-VN: obtain.01</i>	<i>EADB: lortu_1</i>
Arg0: receiver (agent)	esperimentatzailea (ERG)
Arg1: thing gotten (theme)	gaia (ABS)
Arg2: received from (source)	-

4. taulan ikusten dugunez, PB-VNk hiru argumentu esleitzen dizkio eta EADBk bi baino ez; baina *Source* (ABL) hori kasu onargarrien artean dago euskarazko adieran.²⁵ Gainera, aditz horrek adiera edo predikatu orokor bakarra du, hau da, ez dago beste adiera batetik bereiztu beharrik, eta, beraz, ez dago arazorik argumentu gisa etiketatzeko.

Asko izan dira horren antzeko kasuak. Gure irizpide orokorra PB-VN eredua jarraitzea denez, EADBn argumentu ez diren horiek, argumentu gisa etiketatu ditugu.

5.3. Aditz korapilatsuak

Atal honetan, etiketatzean arazoak eman dizkiguten aditzak aipatuko ditugu eta horien sailkapena edo multzokatzea egiten saiatuko gara.

5.3.1. Mugimenduzko aditzak: *igo/jaitsi, atera/sartu*

Motion verbs delako aditzak bereziki aztertuak izan dira hainbat hizkuntzatan. Talmy-ren (1985) tipologia jarraituta, euskara *verb-framed* hizkuntza izango litzateke, gaztelania edo turkiera bezala, eta ingelesa, aldiz, *satellite-framed* hizkuntza, alemana bezala. *Satellite-framed* hizkuntzek aditzetik kanpo uzten dute mugimenduaren norabidea adierazten duen elementua (*out, in, up, etab.*) eta *verb-framed* hizkuntzek, aldiz, aditzean bertan adierazten dute norabide hori. Horren adibide dira: *igo (go up)/jaitsi (go down); atera (take out/come out/go out)/sartu (come in/go in)*.

Mugimenduzko aditz horiek arazoak eman dizkigute euskarazko aditzari ingelesezko PB-VN-n ordaina topatzerakoan, ingelesezko *phrasal verb* horiek ez daudelako sistematikoki landuta PB-VN-n. Adibidez, *atera* aditzaren *entitate baten kokapen-aldaketa* adierentzat ingelesez *take out, come out* edo *go out* erabiltzen dute. PB-VN-n badago *take out*, baina ez adiera horretan, baizik eta *lortu* adieran (*obtain*). Gauza bera gertatzen da *come out* aditzarekin, *appear* esanahia adierazteko erabiltzen dute. PBn *go out* ere agertzen da, baina 2 argumentu baino ez ditu eta gure *kokapen-aldaketak* 3 edo 4 argumentu eskatzen ditu:

²⁵EADBn aditz-adiera bakoitzean ezinbestekoak diren kasuez gain (argumentuak), kasu onartuak eta ez-onartuak zeintzuk diren zehazten da.

atera-DA: gai ukitua_ABS; abiapuntuako kokapena/bidea_ABL; helburuko kokapena_ALA

atera-DU: kausa_ERG; gai ukitua_ABS; abiapuntuko kokapena/bidea_ABL; helburuko kokapena_ALA

Azkenean, *come out*, *take out* ordainak aukeratu ditugu gure *atera* aditzaren *kokapen-aldaketa* adierentzat, baina ez PBn agertzen diren adiera horiekin baizik eta gure *atera*-ren pareko den *kokapen-aldaketa* adierarekin; hau da, ingelesez esanahi hori nola adierazten den kontuan hartu dugu eta gure *atera* aditzaren *kokapen-aldaketa* adieraren argumentu eta rolak jarri dizkiogu. Ikusi 5. taulan *atera_1* aditz-adierak BVIn duen informazioa:

5. taula. *atera_1* sarrera BVIn

<i>BVI: atera_1 (come_out_01/extract_01/take_out_01)</i>
Arg0: agent, kausa (ERG)
Arg1: theme, gaia (ABS)
Arg2: source, abiapuntuko kokapena (ABL/DAT)
Arg3: destination, helburuko kokapena (ALA)

Bestalde, *go up*, *go down*, *come out* eta antzekoak *go* eta *come* aditzen sarreretan daude sartuta. Adibidez *go.01* aditz-adieraren azpian honako adibide hau aurkitzen dugu PBn:

(9) 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 ordez, *rise*, *raise*, *ascend* eta antzeko ordainak erabili behar izan ditugu. Gauza bera gertatu zaigu antzeko beste aditzekin.

5.3.2. Hitz Anitzeko Unitate Lexikaletarako joera handia duten aditzak

Badira Hitz Anitzeko Unitate Lexikaletarako (HAUL) joera berezia duten aditzak (*eman*, *jarri*, *hartu*, *pasatu*, *egin*, *etorri*...) 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, honek are ageriago uzten du aditz-esapideen aurrean gaudela.

Horrelako adibideak etiketatzean, HAULa zurrera bada (Urizar 2012), besterik gabe etiketatu gabe utzi eta oharren eremuan “*aditz-esapidea*” dela markatu dugu (*parte hartu*

esaterako) (Aldezabal *et al.* 2010b), baina zalantzazko kasuetan, hau da, HAULetara hurbiltze bidean dauden horien kasuan, agerpena etiketatu egin dugu, baina oharretan horiei ere “aditz-esapidea” oharra gehitu 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 urundu egiten da. Horregatik iruditzen zaigu horrelakoetan HAULetatik hurbilago gaudela: *amore eman, urratsak eman...*
- *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, 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* gisa etiketatzea INEa onartuta alternantzia gisa (alegia, ABS/INE arteko alternantzia planteatzea), baina “aditz-esapidea” oharra ere 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: i) entitate baten kokapen-aldaketa, ablatiboa (ABL) azpikategorizatzen duena (10) eta ii) entitate baten edukitzea, INEa azpikategorizatzen duena (esanahi berezi samarra azken hau (11)).

(10) Lasterketa buruan hartu zuen lekukoa Monique Hennagan kidearen eskutik, baina Hemmings jamaikarra gain-gainean zuela

(11) Ibarretzek Lehendakaritzako aretoan hartu zuen Díez

Jakin badakigu aditz horrek HAULetarako 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. Horrek guztiak *leku-aldatze* esanahitik urrundu egiten ditu. Beste batzuk, aldiz, bigarren adieratik hurbilago daude, baina INEa ez da izaten kokapena edo lekua eta hauetan 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 beharbada horrek erakusten digu, hain zuzen, HAULEtatik hurbilago daudela.

- Azkenik, jakin badakigu *egin* aditzarekin osaturiko HAULEn multzoa oso handia dela. Aditz horri adiera bakarra definitzen zaio EADBN, *sortze-prosezua*, hain zuzen ere; baina badira balio horretatik aldentzen diren hainbat eta hainbat adibide. Gehienetan iruditzen zaigu euskaraz joera dagoela predikatuari esanahia ematen dion elementua aditzetik kanpo geratzeko;²⁶ hau da, ingelesa bezalako hizkuntza batean esaterako *to bike* edo *to walk* bezalako aditzak ditugu eta horiek euskaraz *bizikletaz ibili* eta *oinez ibili* izango lirateke. Orobat gaztelaniazko *traicionar*, *presionar*, *morder* bezalako 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. Hona EPEC corpusean agertu zaizkigun adibide batzuk:

(12) urrats eta pauso handiak egin zituzten

(13) traizio egin diola

(14) barrutik boikota egiteko

(15) zuzeneko emankizuna egingo dute

(16) errieta egingo digute

(17) koska egin

(18) presioa egin

Bukatzeko esan, HAULEtarako joera handia duten aditzetan, gure zalantza nagusia adiera berriaren eta HAULaren arteko mugak bereiztean sortu zaigula. Gehienetan HAULak direla erabaki dugu, baina hala ere, zalantzazko adibideak markatuta utzi ditugu rolei eta HAULEi buruzko azterketa sakonagoa egitea erabakitzen dugunerako.

²⁶ Ikus 5.3.1 eta 5.4 ataletan kontrako fenomenoari buruzkoak.

5.3.3. *Euskaraz aditz bakar batekin esaten dena, ingelesez ordain desberdinekin itzultzen da*

Bi egoera daude horrelako kasuetan. Alde batetik, euskarazko aditz-adiera bat adierazteko ingelesezko ordain bat baino gehiago erabil daitezke eta, bestetik, euskaraz aditz bakar batekin esaten dena, ingelesez bat baino gehiagorekin esaten da testuinguruaren arabera. Bi egoera horiek ondorio desberdinak dituzte argumentuei eta rolei dagokienez.

a) Batzuetan, ingelesezko ordain desberdinak erabil daitezke aditz-adiera berdina adierazteko, baina aditz-adiera horien argumentu 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 hori adierazteko. PB-VN begiratzen badugu, *finish.01* edo *end.01* VN klase berekoak (55.4-1 klasea) direla ikusten dugu eta beraz, biek dituztela argumentu eta rol berberak: *Agent, Theme, Instrument*. Gure *amaitu* aditzaren *egoera-aldaketa* adierak bi argumentu ditu: *kausua* eta *gai ukitua*, eta ez dago inolako arazorik *amaituren* balio semantiko hori PB-VNko *finish.01* eta *end.01* aditz-adierekin etiketatzeko.²⁷

b) Beste batzuetan, euskaraz aditz bakar batekin esaten dena, ingelesez aditz desberdinekin esaten da testuinguruaren arabera. Horren adibide garbia da *egin* aditza (baita *eskatu* edo *pasatu* ere). Euskaraz hainbat gauza desberdin esateko erabiltzen da *egin* aditza, baina esanahia gorabehera, aditza 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. 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, adibidez:

(19) kanta asko egin zituen: *compose_02*

(20) Baina biek ere joko alaiagia egiten zuten ACB baterako: *practice_01*

(21) Baserrietan sagardoa etxeko sagarrekin egiten zen: *manufacture_01*

(22) ateen bidez eginiko erabilpen anitzeko zirkuitu konbinazionalak:

²⁷ Jakin badakigu kontrako fenomenoak ere gertatzen dela (ingelesez *to break* eta euskaraz *apurtu, puskatu* edota *hautsi*).

build_01/construct_01

(23) bilera bat egin: hold_04

(24) Boticak 27 puntu egin zituen: score_01

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; esaterako:

- *compose_02: agent, product, beneficiary*
- *practice_01: agent, theme, instrument*
- *manufacture_01: agent, product, beneficiary*
- *ask_02: agent, topic, recipient*
- *build_01: agent, product, material, beneficiary*

PB-VN ereduari jarraituz etiketatzen ari garenez, gure erabakia izan da ordain bakoitzari dagozkion argumentu eta rolak etiketatzea, baina jakin badakigu euskara hutsezko lexikoia egiten dugunean, *egin* aditzak ezingo dituela rol desberdinak jaso agerpen bakoitzean baldin eta ez baditugu adiera desberdinak definitzen. Hau da, *galdera bat egin dela, maiteari poesia bat egin dela, 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, *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. EADBn ez dira azken bi hauek argumentutzat jo eta iruditzen zaigu gehiago direla adjuntuak, baina, bestalde, *egin* aditzak adiera bakarra izanda, ez dago arazorik argumentu gisa etiketatzeko.

5.4. Ingelesez existitzen ez diren aditzak: *bazkaldu (have_lunch)* eta *galdu (get_lost)*

Euskaraz, 5.3.2 atalean aipatu dugun bezala, askotan gertatzen da predikatuari esanahia ematen dion elementua aditzetik kanpo geratzea; ingelesez, aldiz, gauza bera esateko hitz bakar bat izatea. Fenomeno horren erakusgarri *egin* aditzaren hainbat adibide jarri ditugu, 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), *to get lost* (galdu) eta baita *go up/go down* eta antzekoak ere.²⁸ 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 argumentu eta rolak esleitu dizkiegu asmatutako "aditz-esapide" horiei.

Ikusten dugunez, fenomeno honek bi norabideetan sortzen dizkigu arazoak, eta bietan erabaki desberdinak hartu ditugu euskarazko eta ingelesezko ereduak elkarrengana hurbiltzeko.

5.5. EADBn landuta ez dauden balio semantiko berriak aurkitu ditugu corpusean

Zenbait kasutan lexikoietan agertzen ez diren adierak agertu zaizkigu corpusean, hau da, EADBk eta ingelesezko ordain egoki gisa aukeratu dugun PB-VNko aditzak definituta ez dituzten adierak aurkitu ditugu corpusean. Etiketatzeko-lan horrek, beraz, EADBn landutako aditzen sarrerak aberasteko aukera eman digu. Jarraian, horrelako kasuen adibide aipagarri batzuk emango ditugu:

- *egin* aditzean EADBn landuta ez zeuden hainbat adiera aurkitu ditugu corpusean, nahiz eta askotan HAULen aurrean gauden zalantza sortzen diguten. Esana dugun bezala, EADBn *egin*ek ez du DATA azpikategorizatua, onartua baizik, eta ekintza zuzentzen zaion entitatea adierazten du, hots, *helburua*. Horregatik, DATA onartzen ez duten adibideak aurkitzen ditugunean, predikatu orokorretik aldentzen den adiera baten aurrean gaudela esan dezakegu. Ikusi (26) adibidea:

(25) Ordu batzuk egin dituzte hantxe

Adibide horretan DATA agertzea ezinezkoa da eta ez dirudi *sortze-prozesuko* predikatu orokorrean sar dezakegunik. Horregatik, gure BVIin adiera berri bat proposatu dugu, PB-VNko *spend_02* aditzaren parekoa litzatekeena, hain zuzen ere. Ikus 6. taula:

6. taula. *spend.02* eta *egin_2* aditz-adierak PB-VNn eta BVIin duten informazioa

<i>PB-VN: spend.02</i>	<i>BVI: egin_2</i>
Arg0: bider, waiter (agent)	Arg0: <i>agent</i> , esperimantatzailea (ERG)
Arg1: unit of time (asset)	Arg1: <i>asset</i> , neurria (ABS)
Arg2: activity (goal)	Arg2: <i>goal</i> , jarduera (INE)

²⁸ Baita 5.3.1 atalean ikusi ditugun *go up/go down* eta antzekoak ere.

- *eman* aditzean ere, EADBn definituta ez dagoen adiera berri bat aurkitu dugu corpusean. EADBn 3 adiera ematen dira aditz honentzat: i) elkarraldatzea, ii) denbora-epe baten barruan egiten den jarduera eta iii) gertakizun baten inguruko ustea adierazten duen jarduera.

Corpusean (26) adibidea aurkitu dugu:

(26) Amak esan zidanean zure leihoak eraikuntzaren gibelean zegoen arkadira ematen zuela [...]

Adiera honetan ALA agertzea ezinbestekoa da; EADBn definitutako hiru adieretan, aldiz, ALA atzizkia onartu gabekoen²⁹ 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-VN eta BVIn adiera horren informazioa 7. taulan ikus daiteke:

7. taula. *overlook.02* eta *eman_3* aditz-adierak PB-VN-n eta BVIn

<i>PB-VN: overlook.02</i>	<i>BVI: eman_3</i>
Arg0: place ³⁰	Arg0: <i>agent</i> , esperimentatzailea (ERG)
Arg1: direction	Arg1: <i>destination</i> , helburua (ALA)

Badira zalantzazko kasuak ere. Batzuetan, ez dago oso argi adiera berri baten aurrean gauden edo testu mota jakin bateko erabilera berezi baten aurrean gauden. Ikusi adibidez, *atera* aditzaren hurrengo adibidea:

(27) Epaileak txartela atera dio

Kasu honetan zeren aurrean gaude? *Kokapen-aldaketa* bat ala *sortze-prozesu* bat? Badirudi gehiago dela lehenengoa, hau da, “txartela epailearen poltsikotik kanpora atera dela”, “jokalariaren aurrera atera dela”, baina *abiapuntuko* eta *helburuko kokapen* horiek ez dira inoiz esplizitu agertzen, testuinguruagatik ulertzen ditugu. Kirolaren inguruko corpus batean horrelako zerbait ikusiz gero, denok ulertzen dugu “txartela poltsikotik atera dela”, baina inoiz ez dugu esplizitu idatzita ikusiko. Honenbestez, adiera berri baten aurrean gaude ala kirolari buruzko corpus berezitan *atera* aditzaren *kokapen-aldaketa* adierak hartzen duen

²⁹ Ikus Aldezabal 2004-n C eranskineko 79., 81. eta 82. orrialdeak.

³⁰ VN-n ez da *overlook* aditza agertzen, beraz, ez ditugu VNko rolak esleitu. VNko sistema jarraituz, guk gehitu ditugu ingelesezko rolak BVIn.

balio sintaktiko-semanticoko berezi baten aurrean? Iruditzen zaigu gai hau sakonago aztertzeke arazoa dela eta euskarazko aditzentzat proposatzen ditugun adiera orokorrak corpus berezietan nola gauzatzen diren ikustea izan litekeela etorkizuneko lan interesgarri bat.

Adiera berriak aurkitzearena etenik ez duen fenomeno da, batetik, corpus berriak ustiatu ahala gertatuko baita eta, bestetik, beharrezkoa baita hizkuntzak bizirik iraungo badu.

6. Etiketatzeko lanaren emaitzak

Azken atal honetan, egindako lanaren ondorio gisa orain arteko gure lanaren emaitza aurkeztuko dugu. Esan bezala, alde batetik, EPEC-RolSem corpus etiketatua dugu eta, bestetik, aditzen lexikoia (BVI) osatu dugu.

EPEC corpusean guztira 1.457 aditz desberdin daude, baina horietatik 167 aditzek baino ez dute 30 agerpen edo gehiago. Guk EPEC-RolSem corpusean, oraingoz 244 aditz landu ditugu (364 aditz-adiera eta 22.343 agerpen guztira); alegia, corpusaren % 71 etiketatu dugu. Landutako aditz horien artean 140 dira 30 agerpen edo gehiagokoak. Gure helburua 30 agerpen edo gehiago duten aditz guztiak etiketatzea da, beraz, 27 aditz geratzen zaizkigu etiketatzeko eta horretan ari gara.

Aditzen lexikoari (BVI) dagokionez, aditz baten sarreran honela antolatzen da informazioa:

- 1) Euskarazko aditzaren adiera eta dagokion PB-VNko ordaina.
- 2) Aditz-adieraren egitura sintaktiko-semanticoa: i) argumentu kopurua edo balentzia, ii) VNko rola, iii) EADBko rola, iv) postposizio-atzizkia eta (zenbaitetan) v) hautapen-murriztapena (+/- biziduna, +/- gizakia, +/- konkretua).

Halaber, esan beharra daukagu corpuseko eta lexikoiko informazioa bilatu eta ustiatzeko tresna bat ere sortu dugula, e-ROLda izenekoa eta <<http://ixa2.si.ehu.es/e-rolda/index.php>> helbidean erabilgarri dagoena. Tresna horrek hiru informazio mota ematen dizkigu euskarazko aditz jakin bati buruz galdetuz gero: i) BVIko informazioa; ii) aditz-adiera bakoitzaren PropBank (PB), FrameNet (FN) eta Basque WordNet (BWN) loturak eta iii) EPEC-RolSem corpuseko adibideak.

e-ROLda-k informazioa ikusteko eta bere gainean bilaketa zehatzak egiteko aukera ere ematen digu; hau da, galde dezakegu aditz baten adiera jakin bati buruz, rol semanticoko jakin bati buruz, postposizio atzizki bati buruz edo horien denen arteko konbinazio jakin bati buruz. Hori guztia kontutan hartuta, e-ROLda ezinbesteko baliabidea iruditzen zaigu corpusean eta lexikoian gordetzen den informazio guztia bistaratzeko, arakatzeko eta ustiatzeko.

Bibliografia

- Aduriz, Itziar; Aranzabe, Maria Jesús; Arriola, Jose Mari; Atutxa, Aitziber; Díaz de Ilarraza, Arantza; Ezeiza, Nerea; Gojenola, Koldo; Oronoz, Maite; Soroa, Aitor & Urizar, Ruben, 2006, "Methodology and steps towards the construction of EPEC, a corpus of written Basque tagged at morphological and syntactic levels for automatic processing". In A. Wilson, P. Rayson & D. Archer (arg.), *Corpus Linguistics Around the World*. Netherlands: Rodopi. Book series: Language and Computers. Vol. 56, 1-15.
- Agirre, Eneko; Aldezabal, Izaskun; Etxeberria, Jone & Pociello, Elixabete, 2006, "A Preliminary Study for Building the Basque PropBank". In *Proceedings of the 5th International Conference on Language Resources and Evaluations (LREC)*. Genoa, Italy: European Language Resources Association (ELRA), 981-986.
- Aldezabal, Izaskun, 2004, *Aditz-azpikategorizazioaren azterketa. 100 aditzen azterketa zehatza, Levin (1993) oinarri harturik eta metodo automatikoak baliatuz*. Leioa: UPV/EHU. Doktore tesia.
- Aldezabal, Izaskun; Ceberio, Klara; Esparza, Itsaso; Estarrona, Ainara; Etxeberria, Jone; Iruskieta, Mikel; Izagirre, Elixabeta & Uria, Larraitz, 2007, *EPEC (Euskararen Prozesamendurako Erreferentzia Corpusa) segmentazio-mailan etiketatzeko eskuliburua*. Barne Txostena, Informatika Fakultatea. Donostia: UPV/EHU / LSI / TR 11-2007.
- Aldezabal, Izaskun, 2008, "Aditz-azpikategorizazioa". In X. Artiagoitia & J. A. Lakarra (arg), *Gramatika Jaietan. Patxi Goenagaren omenez. ASJUren gehigarria*. Donostia: Gipuzkoako Foru Aldundia-EHU, 29-46.
- Aldezabal, Izaskun; Aranzabe, Maria Jesús; Arriola, Jose Mari & Díaz de Ilarraza, Arantza, 2009, "Syntactic annotation in the Reference Corpus for the Processing of Basque (EPEC): Theoretical and practical issues". *Corpus Linguistics and Linguistic Theory*. Berlin-New York: Mouton de Gruyter, 5:2, 241-269.
- Aldezabal, Izaskun; Aranzabe, Maria Jesús; Díaz de Ilarraza, Arantza & Estarrona, Ainara, 2010a, "Building the Basque PropBank". In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odjik, S. Piperidis, M. Rosner eta D. Tapias (arg.), *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC 2010)*. Valletta (Malta): European Language Resources Association (ELRA), 1414-1417.
- Aldezabal, Izaskun; Aranzabe, Maria Jesús; Díaz de Ilarraza, Arantza; Estarrona, Ainara; Fernández, Kike & Uria, Larraitz, 2010b, *EPEC-RS: EPEC (Euskararen Prozesamendurako Erreferentzia Corpusa) rol semantikoekin etiketatzeko eskuliburua. Barne-txostena*. Informatika Fakultatea. Donostia: UPV/EHU / LSI / TR 02-2010.
- Aldezabal, Izaskun; Aranzabe, Maria Jesús; Díaz de Ilarraza, Arantza; Estarrona, Ainara & Uria, Larraitz, 2010c, "EusPropBank: Integrating Semantic Information in the Basque Dependency Treebank". In A. Gelbukh (arg.), *Lecture Notes in Computer Science (LNCS) n° 6008, Computational Linguistics and Intelligent Text Processing*. Berlin-Heidelberg-New York: Springer, 60-73.
- Aldezabal, Izaskun; Aranzabe, Maria Jesús; Díaz de Ilarraza, Arantza; Estarrona, Ainara; 2013, *A methodology for the Semiautomatic Annotation of EPEC-Rolsem, a Basque Corpus Labeled at Predicate Level following the PropBank-Verb-Net Model*. Barne-txostena. Informatika Fakultatea. Donostia: UPV/EHU/LSITR01-2013.
- Aldezabal, Izaskun, 2010, "Basis for the annotation of EPEC-RolSem". *Interdisciplinary Workshop on Verbs. The identification and Representation of Verb Features*. Scuola Normale Superiore - Laboratori di Linguistica. Pisa (Italia): Università di Pisa, Dipartimento di Linguistica, 92-97.
- Aparicio, Juan José, 2007, *Clasificación semántica de los predicados en español*. Barcelona: Universitat de Barcelona.
- Aranzabe, Maria Jesús, 2008, *Dependentzia-ereduan oinarritutako baliabide sintaktikoak: zuhaitz-bankua eta gramatika konputazionala*. Leioa: UPV/EHUko doktore tesia, Euskal Filologia Saila.
- Babko-Malaya, Olga, 2005, *PropBank Annotation Guidelines*. <<http://verbs.colorado.edu/~mpalmer/projects/ace/PBguidelines.pdf>>
- Babko-Malaya, Olga; Bies, Ann; Taylor, Ann; Yi, Szuting; Palmer, Martha; Marcus, Mitch; Kulick, Seth & Shen, Libin, 2006, "Issues in Synchronizing the English Treebank and PropBank". *Proceedings of the Workshop on Frontiers in Linguistically Annotated Corpora*. Sidney, Australia: Association for Computational Linguistics, 70-77.
- Bhatt, Rajes; Narasimhan, Bhuvana; Palmer, Martha; Rambow, Owen; Sharma, Dipti Misra & Xia, Fei, 2009, "A Multi-Representational and Multi-Layered Treebank for Hindi/Urdu". In *Proceedings of the Third Linguistic Annotation Workshop*. Suntec, Singapore: ACL-IJCNLP 2009, 186-189.
- Civit, Montse; Aldezabal, Izaskun; Pociello, Elixabete; Taulé, Mariona; Aparicio, Juan José; Márquez, Luis; Navarro, Borja & Martí, María Antonia, 2005, "3LBLEX: léxico verbal con frames sintáctico-semánticos". *Procesamiento del Lenguaje Natural*, 367-374.
- Díaz de Ilarraza, Arantza; Garmendia, Aitzpea & Oronoz, Maite, 2004, "Abar-Hitz An Annotation Tool for the Basque Dependency Treebank". In M.T. Lino, M.F. Xavier, F. Ferreira, R. Costa, R. Silva, C. Pereira, F. Carvalho, M. Lopes, M. Catarino & S. Barros (arg.), *Proceedings of Fourth International Conference on Language Resources and Evaluation (LREC 2004)*. Lisboa: European Language Resources Association (ELRA), 251-254.

- Hajic Jan, 1998, "Building a Syntactically Annotated Corpus: The Prague Dependency Treebank". In E. Hajicova (arg.), *Issues of Valency and Meaning Studies in Honor of Jarmila Panevová*. Prague (Czech Republic): Charles University Press, 106–132.
- Hajic, Jan; Panevova, Jarmila; Urešova, Zdenka; Bemova, Alevtina; Kolarova, Veronika & Pajas, Petr, 2003, "PDT-VALLEX: Creating a Largecoverage Valency Lexicon for Treebank Annotation". In *Proceedings of the Second Workshop on Treebanks and Linguistic Theories*. Sweden: University of Vaxjo, 57–68.
- Hovy, Eduard; Marcus, Mitch; Palmer, Martha; Pradhan, Sameer; Ranshaw, Lance & Weischedel, Ralph, 2006, "OntoNotes: The 90% solution". *Proceedings of the Human Language Technology / North American Association of Computational Linguistics conference*. New York, 57-60.
- Hualde, Jose Ignacio & Ortiz de Urbina, Jon, 2003, *A Grammar of Basque*. Berlin/New York: Mouton de Gruyter.
- Kingsbury, Paul & Palmer, Martha, 2002, "From Treebank to PropBank". In *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC-2002)*. Las Palmas.
- Kingsbury, Paul & Palmer, Martha, 2003, "PropBank: The Next Level of Treebank". In *Proceedings of Treebanks and Lexical Theories*. Växjö, Sweden.
- Kipper, Karin; Korhonen, Anna; Ryant, Neville & Palmer, Martha, 2006, "Extending VerbNet with Novel Classes". In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC)*. Genoa, Italy: European Language Resources Association (ELRA), 1027-1032.
- Kipper, Karin; Korhonen, Anna; Ryant, Neville & Palmer, Martha, 2008, "A large-scale classification of English verbs". *Language Resources and Evaluation Journal*, 42:1, 21-40.
- Levin, Beth, 1993, *English Verb Classes and Alternations. A preliminary Investigation*. Chicago and London: The University of Chicago Press.
- Marcus Mitchell; Santorini, Beatrice & Marcinkiewicz, Mary Ann, 1993, "Building a Large Annotated Corpus of English: The Penn Treebank". *Computational Linguistics* 19: 2, 313–320.
- Marcus, Mitchell, 1994, "The Penn TreeBank: A revised corpus design for extracting predicate argument structure". In *Proceedings of the ARPA Human Language Technology Workshop*. Princeton, New Jersey.
- Merlo Paola & Van der Plas, Lonneke, 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, Singapore: 288–296.
- Palmer, Martha, 2009, "Semlink: Linking PropBank, VerbNet, and FrameNet". In *Proceedings of the Generative Lexicon Conference (GenLex-09)*. Pisa, Italy, 9-15.
- Palmer, Martha; Gildea, Daniel & Kingsbury, Paul, 2005a, "The Proposition Bank: A Corpus Annotated with Semantic Roles". *Computational Linguistics* 31:1, 71-106.
- Palmer, Martha; Nianwen, Xue; Babko-Malaya, Olga; Chen, Jinying & Snyder, Benjamin, 2005b, "A Parallel Proposition Bank II for Chinese and English". *Proceedings of the Workshop on Frontiers in Corpus Annotation II: Pie in the Sky*. Stroudsburg, USA: Association for Computational Linguistics, 61-67.
- Palmer, Martha; Ryu, Shijong; Choi, Jinyoung; Yoon, Sinwon & Jeon, Yeongmi, 2006, *Korean PropBank*. Philadelphia: Linguistic Data Consortium, LDC2006T03.
- Palmer, Martha; Babko-Malaya, Olga; Bies, Ann; Diab, Mona; Maamouri, Mohamed; Mansouri, Aous & Zaghouni, Wajdi, 2008, "A Pilot Arabic PropBank". In N. Calzolari, Kh. Choukri; B. Maegaard; J. Mariani; J. Odijk; St. Piperidis; D. Tapias (arg.), *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*. Marrakech, Morocco: European Language Resources Association (ELRA), 3467-3472.
- Pociello, Elixabete; Agirre, Eneko & Aldezabal, Izaskun, 2010, "Methodology and Construction of the Basque WordNet". *Language Resources and Evaluation (LRE)*, Netherlands: Springer. 45: 2, 121-142.
- Talmy, Leonakd, 1985, "Lexicalization patterns: Semantic structure in lexical forms". In T. Shopen (arg.), *Language typology and syntactic description, Vol. 3. Grammatical categories and the lexicon*. Cambridge: Cambridge University Press. 36-149.
- Taulé, Mariona; Martí, Maria Antonia & Borrega, Oriol, 2011, "AnCora 2.0: Argument Structure Guidelines for Catalan and Spanish". *Working paper 4: TEXT-MESS 2.0 (Text-Knowledge 2.0)*. Centre de Llenguatge i Computació (CLiC). Universitat de Barcelona.
- Tesnière, Lucien, 1959, *Eléments de Syntaxe Structurale*. Paris: Klincksieck.
- Urizar, Ruben, 2012, *Euskal lokuzioen tratamendu konputazionala*. Leioa: UPV/EHUko Doktore tesia.
- Vázquez, Gloria; Fernández, Ana; Martí, Maria Antonia, 2000, *Clasificación verbal. Alternancias de diátesis*. Quaderns de Sintagma 3. Edicions de la Universitat de Lleida.
- Xue, Nianwen, 2008, "Labeling Chinese predicates with semantic roles". *Computational Linguistics* 34: 2, 225-255.
- Xue, Nianwen & Palmer, Martha, 2009, "Adding semantic roles to the Chinese Treebank". *Natural Language Engineering* 15: 1, 143-172.

eXeLearning euskaraz erabiltzen baduzu eta hobekuntzaren bat proposatu nahi baduzu, mezu honi erantzun zure ekarpenarekin. Saiatu ahalik eta zehatzena izaten (zuzendu edo hobetu behar den testua, non agertzen den, zure hobekuntza proposamena, etc.). Denon artean tresna hobetuko dugu! TM. Author. Posts. You must be logged in to reply to this topic. Comments are closed.

eXeLearning loves its Community. @inproceedings{Ibarloza2015EPECRoISemI, title={EPEC-RoISem: Ingelesezko "PropBank-VerbNet" eredura etiketatutako euskarazko corpusa : erabakiak, egokitzapenak eta bereizgarriak}, author={Ainara Estarrona Ibarloza and Izaskun Aldezabal Roteta and A. S{a}nchez and Mar{ia} Jes{u}s Aranzabe Urruzola}, year={2015} }.¹ Sarrera Artikulu hau hizkuntza naturalaren prozesamenduaren alorrean kokatzen da eta zehazki IXA Taldearen jardunaren barruan. Azkenaldian hain modan dagoen corpus based edo corpus driven linguistic delakoen barruan kokatzen gara eta zehazki, corpusetan oinarritutako lexikografian. EPEC-RS: EPEC (Euskararen Prozesamendurako Erreferentzia Corpusa) rol semantikoekin etiketatzeko eskuliburua [Guidelines to tag semantic roles in the EPEC corpus (the Reference etc.; I Aldezabal, MJ Aranzabe, A D{a}z de Ilaraza, A Estarrona, K Fern{a}ndez, Internal Report, UPV/EHU/LSI/TR 02-2010, 2010.² EPEC corpusa predikatu-mailan etiketatzeko oinarriak: EPEC-RoISem, BVI eta e-ROLda. AE Ibarloza. Euskal hizkuntzalaritzaren egungo zenbait ikerlerro: hizkuntzalari etc., 2014. A methodology for the semiautomatic annotation of EPEC-RoISem, a Basque corpus labeled at predicate level following the PropBank-VerbNet model. Digital Scholarship in the Humanities, Vol. 31, Issue. 3, p. 470. CrossRef. Google Scholar. Li, Tianshi Li, Qi and Chang, BaoBao 2016. Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data.