

Euskarazko denbora-egiturak. Azterketa eta etiketatze-esperimentua

Basque time structures. Analysis and annotating experiment

Begoña Altuna
Ixa Taldea. UPV/EHU
begona.altuna@ehu.es

María Jesús Aranzabe
Ixa Taldea. UPV/EHU
maxux.aranzabe@ehu.es

Arantza Díaz de Ilarraza
Ixa Taldea. UPV/EHU
adiazdeilarraza@ehu.es

Laburpena

Denbora-informazioa eraztea oso erabilgarria da hizkuntzaren prozesamenduan (HP), besteak beste, testuen sinplifikazioan, informazio-erazketako eta itzulpen automatikoko sistemetan balia baitaiteke. Lan honetan, euskaraz informazio hori baliagarri bihurtzeko egin diren lehen urratsak azaltzen dira: batetik, euskaraz denbora adierazteko erabiltzen diren egiturak zein eratakoak diren aztertu da gramatiketan oinarrituta, eta bestetik, egitura horiek testuetan etiketatzeko lehen erabakiak hartu dira. Halaber, ekonomiari buruzko corpus bat osatuta egin den etiketatze-lanaren esperimentua azaltzen da.

Gako-hitzak

Denbora-informazioa, denbora-adierazpenak, denborazko erlazio-hitzak, abiarazle lexikoak

Abstract

Time information extraction is very useful in natural language processing (NLP), as it can be used in text simplification, information extraction and machine translation systems. In this paper we present the first steps of making that information accessible for Basque language: on one hand, Basque structures that convey time have been analysed based on grammars and, on the other hand, first decisions on tagging those on real texts have been taken. Also, we give account of an annotating experiment we have carried out on a financial news corpus.

Keywords

Time information, time expressions, time function words, lexical triggers

1 Sarrera

Gizakiok denbora nozioa hitzez adierazteko gaitasuna dugu. Ez dugu mekanismo bakarraren bidez egiten, ordea. Ikus dezagun adibide bat:

[*Duela 13.800 milioi urte*] [*Big-Banga gertatu zenean*] [*berehala*] inflazio kosmikoa izan zela uste da. Hau da, [*ikaragarritzko denbora tarte txikian*] sekulako hedapena izan zuela unibertsoak. [*Une horretan*] gertatutakoa ez dago [*gaur egungo*] ezerekin alderatzerik. (Berria, 2014-03-18)¹

Denborako uneak (*Big-Banga gertatu zenean, berehala, gaur egungo*) edo iraupenak (*ikaragarritzko denbora tarte txikian*) adierazten dituzten denbora-egiturak identifikatu ditugu goiko testuan. *Duela 13.800 milioi urte* irakurritz gero ere, *Big-Bangaren* eta *gaur egunen* artean 13.800 milioi urteko tarte igaro dela inferi dezakegu eta bigarren perpausoko *Une horretan* egiturak zein uneri egiten dion erreferentzia (Big-Bangari) ere erraz atzeman dezakegu. Gertatzen diren ekin-tza edo egoerek ere denbora-egiturak dira: denek kronologiako une edo tarte batean jazotako gertaera bat adierazten dute eta aditz jokatuak denbora gramatikala adierazten dute (*izan zela, izan zuela, dago, etab.*).

Hizkuntzaren prozesamenduan (HP), testuaren ulermen automatiko sendoa helburu bada, hizkuntzaren azterketa sakona egin behar da eta azterketa horretan testuko gertaerak noiz jazotzen diren ere aztertu behar da. Denbora-informazioa automatikoki prozesatzea konplexua da, makinek ez baitute gizakiok dugun hizkuntza prozesatzeko berezko gaitasunik. HPrako denbora-egiturak identifikatu eta egitura bakoitzetik ahalik eta informazio gehien esplizitu egin behar da informazio hori makinarentzat ulergarria izateko. Informazio hori HPko hainbat sistematan erabili ahal izango da: informazio-erazketan, galdera-erantzunetan, testuen sinplifikazioan eta itzulpen automatikoan.

Prozesu horretan, lehenik eta behin, hizkuntzaren azterketa egin behar da, hizkuntza bakoitzean denbora nola adierazten den eta erabiltzen diren egiturak zein diren jakiteko. Infor-

¹<http://paperekoa.berria.info/plaza/2014-03-18/036/002/Uhin.egiaztatzaileak.htm>

mazio hori guztia aztertu eta modu sistemati-koan adierazi behar da. Horretarako, markaketa-lengoaia baten bidez testuko hitzei edo egiturei euren informazioa azaleratuko duen etiketa bana emango zaie. Hala, denbora-egitura bakoitzak etiketa bat hartuko du eta atributuen bidez denbora-informazioa azaleratuko da. Euskarazko denbora-informazioa makina bidez eskuragarri izateko, TimeML (Pustejovsky et al., 2003a) markaketa-lengoaia erabili dugu, hainbat hizkuntzatan (frantsesa (Bittar, 2010), koreera (Im et al., 2009), italiera (Caselli et al., 2011), etab.) ere erabili eta denbora-informazioa adierazteko estandar bihurtu baita.

Hizkuntza bakoitzak denbora adierazteko berezko erak ditu eta, horregatik, hizkuntza guztietara molda daitekeen etiketatze-lengoaia estandarra sortu bada ere, hizkuntza bakoitzarentzako denbora-egituren analisia beharrezkoa da. Euskarazko denbora-adierazpenei buruzko informazioa hainbat gramatikatan (*Euskal Gramatika Lehen Urratsak (EGLU) I eta II* (Altuna et al., 1985; Altuna et al., 1987), *Euskal Gramatika Osoa* (Zubiri & Zubiri, 1995)) aurki daiteke, baina ez da HPra bideratutako denbora-egituren azterketarik egin. Hutsune hori betetzera dator lan hau.

Lan honetan, euskarazko gramatika horietan oinarrituta, denbora-adierazpenak, denborako une edo tarte bat adierazten dutenak, zerrendatuko ditugu 2. atalean. Jarraian, 3. atalean denboraren tratamendu konputazionala azalduko dugu, denbora-informazioaren prozesamendua azalduz eta TimeML markaketa-lengoaia deskribatuz. 4. atalean, euskarazko denbora-egitura batzuk etiketatze hartu ditugun erabakiak eta etiketatze horretan laguntzeko hartu ditugun irizpideak deskribatuko ditugu. 5. atalean, hiru anotatzailek aurrera eramandako anotazio esperimentua azalduko dugu. Amaitzeko, 6. atalean, denboraren prozesamendu automatikoan aurrera egiteko jarraituko ditugun urratsak azalduko ditugu.

2 Denbora-adierazpenen sailkapena

Sarreran aipatu bezala, denbora-informazioa duten formak dira denbora-egitura. Egitura horietako batzuk gertaerak denboran kokatzeko baliatzen direnak dira: denbora-adierazpenak. Atal honetan, euskaraz denborako une bat edo iraupen bat adierazteko erabiltzen diren denbora-adierazpenak deskribatuko ditugu. Denbora-adierazpenek perpausaren denborakokapena adierazten dute; hau da, eurek seinalatzen dute perpausako gertaerari dagokion

denbora-unea edo luzapena. Adierazpide horiek guztiek, formaz ugariak badira ere, *noiz?*, *noiztik?*, *noiz arte?*, *noizko?* eta *zenbat denbora?* galderei erantzuten diete eta denbora une edo tarte mugatu bat (iraupen bat) irudikatzen dute.

Egitura horien azterketa beharrezkoa da horiek etiketatze jarraituko dugun etiketatze-eskema aukeratzeko eta IXA ikerketa-taldean² garatu ditugun tresnen berrerabilera aztertze-ko. Hori kontuan hartuta eta gramatiketan oinarrituta (*Euskal Gramatika, Lehen Urratsak (EGLU) I eta II* (Altuna et al., 1985; Altuna et al., 1987), *Euskal Gramatika Osoa* (Zubiri & Zubiri, 1995) eta Sareko Euskal Gramatika³), denbora-adierazpenen sailkapena egingo dugu jarraian.

2.1 azpiatalean denbora adierazten duten aditzondo eta adizlagunak ikusiko ditugu. 2.2 eta 2.3 azpiataletan egitura zabalagoak, postposizio-lokuzioak eta hitz anitzeko unitate lexikalak (HAUL), izango ditugu hizpide eta 2.4 azpiatalean ordui eta datei arreta berezia eskainiko diegu.

2.1 Aditzondoak eta adizlagunak

Denborazko aditzondoek zirkunstantzia bat adierazten dute. EGLUn jasotzen den moduan, adberbioen artean bi sail nagusi bereizten dira irizpide morfologikoak erabiliz: i) aditzondoak, berez eta postposizio-atzizkirik gabe aditzari laguntzen dioten aditz-sintagmako elementuak, eta ii) adizlagunak, izen-sintagmak adizlagun bilakatze-ko atzizki edo postposizio jakin batzuen beharra dutenak.

Aditzondoak formaz anitzak dira; Euskaltzaindiak (Altuna et al., 1987) hainbat motatakoak batzen ditu: aditzondo bakunak (*atzo* (1), *maiz* (2) edo *luze* (3)), zehaztugabe konposatuak ((4) adibideko *noizbait* eta (5) adibideko *inoiz*) eta bestelakoak ((6) adibideko *dagoeneko*).

- (1) *Atzo* programa berria estreinatu zuten telebistan.
- (2) *Maiz* joaten gara Frantziara ostrak jatera.
- (3) Ebakuntzak *luze* joko du erizainaren arabera.
- (4) *Noizbait* entzuna nuen Mikelek gaur kontatu duen istorioa.

²<http://ixa.si.ehu.es>

³<http://www.ehu.es/seg/aurkezpena>

(5) *Inoiz* horrelako soineko politik dendan ikusten baduzu, eros iezadazu.

(6) *Dagoeneko* euritakoa eta neguko arropa atera behar izan dugu.

Gainera, 2.2. atalean ikusiko dugun moduan, aditzondoak egitura handiagoetan, postposizio-lokuzioetan, ager daitezke eta *gaurtik aurrera* (7) edo *noizean behin* moduko esamoldeak (8) sortu.

(7) *Gaurtik aurrera* ez duela gehiago erreko zin egin du.

(8) *Noizean behin* jatetxe garestietan bazkaltea gustatzen zaio Maiteri.

Horiez gain, adizlagun itxura duten *askotan* (9), *sasoiz* (10) eta gisakoak aipatu behar ditugu, adizlagun itxura izanagatik ere, aditzondoak baitira.

(9) *Askotan* ahazten ditut eguzkitarako betaurrekoak hondartzara noanean.

(10) Lanera *sasoiz* heltzeko ordubete lehenago esnatu ohi naiz.

Adizlagunak, aditzondoak ez bezala, postposizio-atzizkidun sintagmak dira. Orokorrean leku-denborazko atzizkiak (inesiboa nagusiki (11), adlatiboa (12), ablatiboa (13), leku genitiboa (14)) hartzen dituzte sintagmok, baina batzuetan sozietiboa (15) eta instrumentala (16) ere hartzen dute. Postposizio askeak ere har ditzakete sintagma horiek 2.2. azpiatalean aztertuko dugun bezala.

(11) *Goizean* irten zen etxetik.

(12) Iluntzetik *egunsentira* eten behar izan dituzte erreskate operazioak.

(13) *Gabonetatik* egon zen Mikelen eskutitzaren zain.

(14) Eskatutakoa *sanferminetarako* prest izango duzu.

(15) Andra Mari eguna *ostegunarekin* jausten da aurten.⁴

(16) *Bi orduz* egon naiz zure zain.

Aditzondoek bezala, denborako une (11–15) edo iraupen bat (16) adieraziko dute. Bereizketa hori oso erabilgarria izango da prozesamendu automatikoan ikusiko dugunez, une edo iraupen izan ezberdin tratatuko direlako.

2.2 Postposizio-lokuzioak

Postposizio-lokuzioak “adposizio-sintagma baten burua izateko gauza diren unitate fraseologikoak” (Lorente, 2001) direla esan izan da tradizioan. Postposizio atzizkidun forma osagarri eta postposizio aske batez osatzen dira (17–22).

(17) Kontzertua *hamaikak aldera* hasi zen, iragarritakoa baino ia ordubete beranduago.

(18) *Bi minutu barru* ez bada agertzen, joan egingo gara.

(19) *Bostak irian* gertatu zen istripua.

(20) Eskabidea *epez kanpo* aurkeztu zuen eta ez zioten diru-laguntza eman.

(21) *Berandura arte* esna egoteak osasunari kalte egiten diola adierazi du neurozientzialariak.

(22) *Gaur eta bihar bitartean* amaituko dut.

Ez dira, ordea, egitura guztiz zurrunik. Postposizio-lokuzioetan kontuan izan behar da forma bat baino gehiago izan ditzaketela, nahiz eta esanahiari eusten dioten. Horren adierazle ditugu, esaterako, (23–25) adibideetako *Ostegun arte*, *osteguna arte* eta *ostegunera arte* postposizio-lokuzioak:

(23) *Ostegun arte* egongo gara Bartzelonan.

⁴ *Andra Mari eguna* ere denbora-adierazpena da, baina adibide honekin sozietiboaren denbora balioa nabarmendu nahi dugu.

(24) *Osteguna arte egongo gara Bartzelonan.*

(25) *Ostegunera arte egongo gara Bartzelonan.*

Denboraren adierazpenak berezkoak ditu zenbait postposizio-lokuzio: oraintsu aipatutako *alde*, *arte* eta *barru*, adibidez, baina horiek ez dira bakarrik. Usuenak eta euren formak (Aduriz et al., 2008) 1. taulan aurki ditzakegu. Taulan, postposizio askearekin edo beregainarekin batera, forma osagarriaren kasua eta elementu beregainarena berarena ere ageri dira, hala nola, denborazko erabilera adibideak.

Taulako denborazko postposizio-lokuzio arruntenen zerrenda honakoek osatzen dute: *-en aitzin*, *-tik aitzina*, *-en barren*, *-tik goiti*, *-0 irian*, *-z kanpo*, *-0 ondoren*, *-en oste*, *-0 parte* eta *-en pe*. Horien erabilera urriagoa da.

Postposizio-lokuzio horiei aparteko sail bat eskaini nahi izan diegu, elementu osagarriaren burua denbora-adierazpen izango delako eta elementu beregaina “seinale”. Seinaleok erlazio-hitzak izango dira eta denbora-adierazpena gertaera edo beste denbora-adierazpen batekin lotzeko balio izango dute:

(26) *Liburutegia seiak arte egongo da zabalik.*

(26) adibidean ikus dezakegunez, *arte* elementu beregainak denbora-muga adierazten du, kasu honetan liburutegia zabalik egoteko tartea. *Seiakek*, bere aldetik, kronologiako une bat adierazten du, liburutegia itxiko den ordua, hain zuzen ere.

2.3 Hitz anitzeko unitate lexikalak (HAUL)

Hitz anitzeko unitate lexikalek arreta berezia hartzen dute, aurretik azaldutako egituren forma badiute ere, esanahi konposizionala duten egitura ihartuak baitira. Gure eguneroko jardunean ere oso arruntak dira eta prozesamendu automatikoari begira oso garrantzitsua izango da horiek identifikatu eta unitatetzat hartzea analisi zuzena lortu nahi izanez gero. (27) eta (28) adibideetako denborazko HAULak eta beste hainbat (Urizar, 2012)-an aurki daitezke.

(27) *Patata-arrautzopila egiteko lehenik eta behin patatak zuritu behar dira.*

(28) *Gaurko pilota partida luze gabe hasiko da, pilotariak berotze-arietak amaitu baitutuzte.*

2.4 Orduak eta datak

Esan bezala, orduei eta datei atal berezia eskaintzea erabaki dugu. Egitura horiek erraz identifikatzen diren forma zurrunik hartzen dituzte euskaraz. Prozesamendu automatikoari begira, beste denbora-adierazpenetatik aparte aztertzea komenigarria da erregelen bidez aise erazagut baitaitezke. Normalean adizlagun funtzioa betetzen dute ondoko adibideetan ikus daitekeenez:

(29) *Bostetan etorri zen menditik.*

(30) *Autobusa ordu bata eta hamarrean pasatzen da nire etxe azpitik.*

(31) *15:00etan hasiko da emanaldia.*

(32) *Gernikako bonbardaketa 1937ko apirilaren 26an gertatu zen.*

(33) *Bilbon, 2013ko ekainaren 19an.*

(34) *Gaur zortzi izango da liburuaren aurkezpena.*

Baina ezaugarri sintaktikoei dagokienez, perpauseko beste funtzio batzuk ere har ditzakete, subjektua (35) eta objektuarena (36), baita izenlagunarena ere (37)an ageri denez:

(35) *Ordu biak jota ziren heldu zirenerako.*

(36) *Gaur hamalau ditu hilak.*

(37) *Martxoaren 25eko greba eguna oso jendetsua izatea espero da.*

Denbora-egiturak aurkeztu ondoren, 3. eta 4. ataletan azalduko dugu egitura horien analisia nola egin dugun.

3 Ikuspegi konputazionala

Lan honen sarreran aipatu bezala, euskarazko denbora-adierazpenen azterketa hizkuntzaren prozesamenduari begira egin dugu, denbora-informazioaren prozesamenduari begira bereziki. Ataza horretan zer urrats eta baliabide erabili edo erabiliko diren azalduko dugu jarraian.

Postposizio beregaina	Forma osagarria	Elementu beregaina	Adibidea
alde (IZE)	-ABS	-ra	<i>Hirurak aldera</i> bazkalduko dugu
arte (IZE)	-ABS -0	-0/-ko -0	Ikastaroa <i>zortziak artekoa</i> da <i>Bihar arte</i> ez dago autobusik
aurre (IZE)	-tik	-ra	<i>Gaurtik aurrera</i> ez du gehiago erreko
barru (IZE)	-0	-0	<i>Bi egun barru</i> entregatuko dut lana
bitarte (IZE)	-ra -ABS	-0/-an/-ko -an	<i>Etæera bitartean</i> kontatuko dizut hori <i>Bostak bitartean</i> hemen egongo zarete
buru (IZE)	-en	-an	<i>Bi egunen buruan</i> jakin zuen emaitza
gero/geroztik (ADB)	-z	-0/-ko	<i>Istripuaz gero</i> ez du ezer gogoratzen
inguru (IZE)	-ABS	-0/-an	<i>Zortziak inguruan</i> esnatu gara

1. taula: Denborazko postposizio-lokuzioak.

3.1 Denbora-informazioaren prozesamendua

Denbora-informazioa esplizitu egiteko, lehenik eta behin denbora adierazten duten egiturak identifikatu behar dira. Lan hau denbora-adierazpenei eta zenbait erlazio-hitzi baino ez diegu eskaini, baina gertaerak (gertatzen diren egoera edo ekintzak) ere denbora-egiturak dira. Azken horiek denbora-informazioaren ikuspegitik lantzeke ditugu oraindik.

Denbora-adierazpenak azpimultzotan sailkatuko ditugu: data, ordua, iraupena edo errepikapena, eta definituko ditugun atributu batzuen bidez, euren informazioa azalerratu dugu. Denbora-erlazioak adierazten dituzten erlazio-hitzak ere identifikatuko ditugu. Denbora-adierazpen eta gertaeren artean denbora-erlazioak sortzen dira; gertaera bat zein unetan gertatu den edo zenbat iraun duen, adibidez, eta erlazio horietako batzuk testuan esplizituki adierazten dira erlazio-hitzen bidez. Denbora-adierazpenak eta erlazio-hitzak etiketatzeke jarraian azalduko dugun TimeML markaketa-lengoaia (Pustejovsky et al., 2003a) erabili dugu.

TimeML baliatuz etiketatu dugu azterketarako lagina. Eskuz etiketatutako lagina abiapuntu hartuta, denbora-informazioa etiketatzeke eske ma definitzea eta ikerketa taldeko tresnen bererabilgarritasuna aztertzea lortuko dugu. Ondoren, erdi-automatikoki etiketatuko litzateke corpusa eta, behin eskuzko gainbegiratzea eginda, denbora-informazioa *gold standarda* izango litzatekeen corpusa sortuko genuke.

3.2 TimeML

TimeML lengoaia naturaleko testuetako denbora-informazioa etiketatzeke markaketa-lengoaia da. Setzerren lanari (Setzer, 2001) eta TIDES proiektuaren barruko TIMEX2 etiketa multzoari (Ferro et al., 2003) jarraipena emanez sortu zen TimeML (Pustejovsky et al., 2003a); horien proposamenak hobetuz eta osatuz denbora-egitura guztiei etiketa bat esleitzeko etiketatze-eskema sendoa proposatu zen. TimeML XML (eXtensible Markup Language) markaketa-lengoiaren gainean garatu zen eta hizkuntza guztietan aplikagarri egin zen. Egun ISO-TimeML markaketa-lengoaia (ISO-TimeML working group, 2008), International Organization for Standardization-ek (ISO) estandarizat hartutakoa, hainbat hizkuntzatan erabili izan da. TimeMLn oinarrituta zenbait hizkuntzatan denboraren arabera etiketatutako corpusak sortu dira, TimeBank 1.1 (Pustejovsky et al., 2003b) eta TimeBank 1.2 (Pustejovsky et al., 2006) besteak beste.

TimeMLk bai uneak, bai iraupenak etiketatzeke balio du eta etiketatutako denbora-adierazpenek <TIMEX3> etiketa hartzen dute, baita euren informazioa esplizitu egingo duen hainbat atributu ere, hala nola, mota (**type**), denborako unea edo denbora-tartearen luzera adierazten duen zenbakizko balioa (**value**) eta hasiera (**beginPoint**) eta amaiera puntuak (**endPoint**) iraupenen kasuan.

Lengoaia horretan, erlazio-hitzak etiketatzeke <SIGNAL> etiketa erabiltzen da. Gertaerak, “gertatzen diren egoerak” (Pustejovsky et al., 2003a), <EVENT> etiketa hartuko dute eta denbora-erlazioek, denbora-adierazpen eta gertaeren artean sortzen diren erlazioek, <TLINK>.

Denbora adierazpen mota	Adibideak
DATE	sanferminak, 2014ko martxoaren 20an, etzi
TIME	arratsaldeko bostetan, 17:26, gaur goizean
DURATION	50 urte, bost minutuz, bi asteko
SET	egunero, astean birritan

2. taula: Denbora-adierazpenen motaren araberrako sailkapena.

Euskararako ere estandar hori baliatzea erabaki dugu, egun denbora markaketa-lengoaia osoen eta aurreratuen baita.

TimeML euskarara moldatzeko lehen urratsa deskribatzen dugu; alegia, markaketa-lengoaia hori erabilia euskarazko denbora-egiturak identifikatzeko lehen urratsak. Hala, 2. atalean, gramatiketan oinarrituta sailkatu ditugun denbora-adierazpenak TimeMLk proposatzen duen sailkapenaren arabera antolatzen saiatuko gara orain. TimeMLk denbora-adierazpenak lau motatan sailkatzen ditu: DATE (eguna bezain luze edo luzeagoak diren datak), TIME (eguna baino laburragoak diren denborak), DURATION (iraupena) eta SET (errepikapena) (2. taula). Seinaleak 4.1.1 azpiatalean azalduko ditugu.

4 Euskarazko denbora-adierazpenen etiketatzea

Lan honen sarreran esan bezala, testuko denbora-adierazpenak etiketatzeke, TimeML markaketa-lengoaian oinarritutako etiketak baliatuko ditugu. Jarraian, etiketa horiek erabilia sarrerako testuko denbora-egituren informazioa agerian utzi dela ikus dezakegu:

```
<TIMEX3 tid="t1" type="DURATION"
  value="P13800000000Y"
  beginPoint="-13799997986"
  endPoint="2014">
Duela 13.800 milioi urte </TIMEX3>.
Big-Banga gertatu zenean berehala inflazio
kosmikoa izan zela uste da.
Hau da, <TIMEX3 tid="t2" type="DURATION"
  value="PTXS">
ikaragarriko denbora tarte txikian </TIMEX3>
sekulako hedapena izan zuela unibertsoak.
<TIMEX3 tid="t3" type="TIME" value="PAST_REF">.
Une horretan </TIMEX3> gertatutakoa ez dago
<TIMEX3 tid="t4" type="DATE" value="PRESENT_REF">
gaur egungo </TIMEX3> ezerekin alderatzerik.
(Berria, 2014-03-18)
```

Denbora-egiturek etiketak hartu dituzte eta etiketa horien barruko atributuek egituren ezauzgarriak (mota, balioa, hasiera-puntua, etab.) adierazten dituzte. Identifikatzailea (tid) esleitu diegu lehenik eta behin, eta gero deskribape-

nezko atributu dei ditzakegunak: `type`, `value`, `beginPoint` eta `endPoint`. Informazio hori guztia <TIMEX3> XML etiketa baten barruan batu dugu, ondoren automatikoki prozesagarria izateko.

Ikus daitekeenez, ez diegu denbora-adierazpen guztiei etiketa eman (*Big-Banga gertatu zenean* eta *berehala*). Egitura horiek denborazkoak badira ere, lehen urrats honetan ez ditugu etiketatu, denborazko menderagailuak nola tratatu eta zein balio esleitu erabakitzeke azterketa-lan zehatzagoa egin behar delako. *Berehala* moduko aditzondoei balioa esleitzea zaila da eta horiek ere azterketa sakonagoaren beharra izango dute.

Jarraian, aurreko testutik erauzitako denbora-adierazpen baten etiketatze adibide bati helduko diogu. (38) adibidean denbora-adierazpen baten denboraren araberrako etiketatzea ikus daiteke; zein atributu hartzen duen eta atributuok zein balio hartzen duten.

```
(38) <TIMEX3 tid="t1" type="DURATION"
  value="P13800000000Y"
  beginPoint=-13799997986"
  endPoint="2014"> Duela 13.800 milioi urte
</TIMEX3>
```

(38) adibideko egitura denbora-adierazpena dela adierazteke, <TIMEX3> etiketa esleitu zaio, baita “tid” identifikatzailea ere. Iraupen bat adierazten du egitura horrek eta hala adierazi da “DURATION” mota esleitzean eta, iraupena izaki, hasiera puntua eta amaiera puntua ere esleitu zaizkio. Denbora-adierazpenaren balio kronologikoa balueren bidez islatzen da; kasu honetan, iraupen bat denez, balio horrek denbora tarte adieraziko du. Halaber, `beginPoint` eta `endPoint` atributuek iraupen horren hasiera eta amaiera puntua adieraziko dituzte.

4.1 Seinaleak eta abiarazle lexikoak

Denbora-informazioaren azterketan, beste elementu batzuk ere izan behar dira kontuan. Jarraian denborazko erlazio-hitzak (seinaleak) eta denbora-adierazpenak identifikatzeko baliagarri

izango diren denborazko abiarazle-lexikoak azalduko ditugu. Lehenek denborazko erlazio baten berri ematen dute eta bigarrenek denbora-adierazpenak antzematen laguntzen dute, berez denbora-adierazten duen elementu bat edo gehiagoz osatuta baitaude.

4.1.1 Seinaleak

Seinale deituko diegu denbora erlazio-hitz edo egiturei, ingelesez TimeML lengoian (Pustejovsky et al., 2003a) “signal” erabiltzen baita horiek izendatzeko eta hurbileko kalkoa erabiltzea egokia dela uste dugulako. Horiek perpausoko gertaeren artean edo denbora-adierazpenen artean sortzen diren erlazioak adieraziko dituzte. Hainbat kategoriatakoak izan daitezke seinaleak:

- denborazko postposizio-lokuzioak:⁵ *-ABS/-0 arte, -ra/-ABS bitarte, -z gero, -tik aurre.*
 - denborazko lokailuak: *bitartean.*
 - denborazko adberbioak eta adberbio lokuzioak: *ondoren, eta gero, baino lehenago.*
 - karaktere bereziak: *- eta /*
- (39) Sanferminak astebete *barru* amaituko dira, baina ordura *arte* ez da aspertzeko betarik izango.
- (40) Patatak frijituko ditugu. *Bitartean* arrautzak irabiatuko ditugu.

(39) adibidean *-0 barru* postposizio-lokuzioa dugu *astebete* denbora-adierazpena eta *amaituko dira* gertaeraren arteko erlazioa azalduz. Kasu honetan amaiera unea inferi dezakegu: iterazio mementoa baino astebete beranduago. *Artek*, ordea, *ordura* denbora-adierazpena eta *ez da izango* egoera lotzen ditu baldintza berria noiz hasiko den adieraziz. (40) adibidean, *frijitu* eta *irabiatu* gertaeren arteko denbora-erlazioa nolakoa den adierazten du *bitarteanek*.

Funtzio bera betetzen dute denborazko mendeko perpausuetako erlazio-atzizkiek, *-(e)n*, *-(e)la*, baina TimeMLn denbora-informazioaren etiketatzea hitzaren mailan egiten denez (ez morfema mailan), horiek nabarmentzea oso zaila da eta alde batera uztea erabaki dugu. Ondorioz, erraztasunaren izenean, denborazko erlazio-atzizkiak ez markatzea erabaki dugu.

⁵Etiketatzean postposizioen elementu beregaina baino ez da etiketatzen.

4.1.2 Abiarazle lexikoak

Abiarazle lexikoak (Ferro et al., 2003)-en arabera, “*a word or numeric expression whose meaning conveys a temporal unit or concept*” (denborazko unitate edo kontzeptu bat adierazten duen hitzezko edo zenbakizko adierazpenak) dira. Euskaraz ere, berez denbora adierazten duten esapideak aurki ditzakegu: *ordu, gaur, ekain*, eta abar. 2.1 ataleko adibideetan ikusi ahal izan denez, denbora-adierazpenek berez denbora adierazten duen unitatea izango dute buru askotan ((11) adibideko *goizean* eta *gauera*, (14) adibideko *sanferminetarako*, (15) adibideko *ostegunarekin* eta (16) adibideko *orduz*). Abiarazle lexiko horiek identifikatzea garrantzitsua izango da hizkuntzaren tratamendu automatikoa egitean, egiturek berez denbora adierazten baitute.

Forma horiek orokorrean denbora adierazten duten esapideetan agertzen dira. Ondorioz, abiarazle lexiko horiek aurretik deskribatutako egituraren batean agertuz gero, denbora adierazten dutela, izan denbora-une edo iraupen bat, suposa dezakegu. Hartara, euskarazko hainbat abiarazle lexiko zerrendatu ditugu (ikus eranskina). Denbora adierazten duten hitzoren zerrenda zabaldu eta osatzeko, euskarazko WordNet (Pociello, 2008) kontsultatu dugu, ingelesezko *time* eta euskarazko *denbora* hitzen euskarazko hiponimoak bilatuz. Besteak (Altuna et al., 1987) eta (Ferro et al., 2003)-ko adibideetatik itzulita edo moldatuta lortu ditugu. 3. taulan abiarazleotako lagin bat kategoria gramatikalaren arabera sailkatu dugu.

Hala ere, denbora-adierazpen guztiek ez dute abiarazle lexiko bat hartzen (*askotan* (9), *bostetan* (29)) eta hori ere kontuan hartzekoa da egiturek automatikoki nola erauzi proposatzean. Egiturek denborazkoak direla erabakitzeke testuinguruaren beharra izango dugu. (41) adibidean *bostetan* denborazko adizlaguna da eta ordua adierazten du; (42) adibideko *bostetan* sin-tagmak, ordea, lekua:

(41) *Bostetan* geratu ginen kafea hartzeko.

(42) Bost etxe zituen eta *bostetan* jarri zuen jaccuzia.

Bestalde, 3. taulan adierazitako abiarazle lexikoak denbora adierazle moduan hartu izan diren arren, gerta daiteke testuinguru batzuetan hitzok denborarik ez adieraztea:

(43) Arratsalde on!

Kategoria	Abiarazle lexikoak
Izen arruntak	aro, arratsalde, asteburu, astelehen, eguerdi, etorkizun, mende, minutu, ordu, otsail, sanfermin, sasoi, seiurteko, solstizio, udazken
Izen bereziak	Gabon, San Joan
Adjektiboak	eguneroko, hilabetekari, bienal
Denbora-patroiak	13:03, 2014/02/12, 1992ko
Adberbioak	gaur, berandu, orain, lehen, berehala, egundo
Zenbakiak	5etan, 6an

3. taula: Kategoriaren araberrako abiarazle lexikoen sailkapena.

(44) Lan hau lau pertsonaren artean egitekoa da, *orduan* (= ondorioz) Mikeli deitu behar izango diogu lana amaitu nahi badugu.

(43) adibideko *arratsalde* denborazkotzat ez hartzea erraza izan daiteke unitate fraseologikoen azterketa eginez gero, esaldiko bi tokenak egitura fosilizatu batean agertzen baitira. (44) adibideko *orduan*, ordea, askoz ere zailagoa izango da hitz anbigua baita.

5 Esperimentazioa

Arestian aipatu bezala, denbora-adierazpen eta seinaleen deskribapenarekin batera, etiketatze saiakera bat egin dugu. NewsReader proiektuaren⁶ metodologia jarraituz, hiru anotatzaileek euskarazko 4 kazetaritza-testu etiketatu dituzte denbora-adierazpenak eta denbora erlazio-hitzak (seinaleak) markatuz. Testuak euskarazko egunkari bateko albisteak dira eta enpresa baten itxiera eta horren erosketari buruzkoak dira. Erabili dugun tresna CAT (CELCT Annotation Tool) (Bartalesi Lenzi, Moretti, & Sprugnoli, 2012) izan da, denbora-informazioa etiketatzekeo aproposa baita.

Etiketatzailerik guztira 56 esaldi aztertu dituzte euskararako denbora-egiturak etiketatzekeo osatu diren etiketatze-gidalerroak jarraituz (Altuna, Aranzabe, & Díaz de Ilarraza, 2014). Gidalerro horiek TimeML markaketa-lengoaian eta NewsReader proiektuko eskuzko etiketatze-irizpideetan (Tonelli, Sprugnoli, & Speranza, 2014) oinarritutakoak izan dira. Etiketatzeko tresna, halaber, proiektu horretan erabiltzen ari direna da.

5.1 Anotatzaileen arteko adostasuna

Anotatzaileen arteko adostasuna neurtu dugu ondoren, testuan ezarritako etiketa kopurua eta etiketa horien luzera (token berak hartzen dituzten) kontuan hartuta. Horretarako, Diceren koefizientea (Dice, 1945) erabili dugu denbora-adierazpen eta seinaleen gaineko adostasuna neurtzeko (ikusi 4. eta 5. taulak). Diceren koefizienteak etiketatzekeo token berak aukeratu diren neurtzen du adostasun osoa islatuz.

4. eta 5. tauletan etiketaren gaineko adostasuna ikus daiteke. *Markablek*⁷ etiketaren luzera zehatzaren gaineko adostasuna adierazten du eta *tokenek* token batzuetan adostasuna izan dela. Denbora-adierazpenetan oso hurbileko emaitzak lortu izanak egiturak identifikatzean anotatzaileek adostasun handia izan dutela adierazten du. Seinaleen kasuan, token bakarrek izanik orokorrean, *markable* eta *tokenentzat* emaitza berak lortu dira.

Testuan ezarritako <TIMEX3> etiketekin batera, bost <TIMEX3> etiketa huts ere etiketatu dituzte. Etiketa hutsak informazio implizitua azaleratzeko erabiltzen diren testuaz kanpoko etiketak dira. Horien gaineko adostasuna oso eskasa izan da A anotatzaileak 5, Bk 1 eta Ck 0 etiketatu baititu. Hiru anotatzaileek jarri dituzten etiketak, testukoak zein hutsak, guztira 33 izan dira eta adostasuna % 66,7koa izan da. Seinaleen aldetik, 16 izan dira markatu direnak eta % 31,2ko adostasuna lortu da hiru anotatzaileen artean.

Denbora-adierazpenen balio eta motaren gaineko adostasuna ere neurtu da. 6. taulan *type* eta *value* atributuen gaineko adostasuna ikus dezakegu, bai bikotekakoa bai orokorra.

Emaitza orokorrak emateaz aparte, binakako ebaluazioak egin ditugu, adostasuna anotatzaile guztien artean maila berekoa den edo ez aztertzeko.

⁶<http://www.newsreader-project.eu/>

⁷Markatu nahi den edozein entitate.

Anotatzaile bikoteak	Micro-average (Markable)	Micro-average (Token)	Macro-average (Markable)	Macro-average (Token)
A – B	0.96	0.976	0.969	0.977
A – C	0.943	0.965	0.923	0.965
B – C	0.902	0.94	0.892	0.942
Guztira	0.935	0.96	0.928	0.961

4. taula: Denbora-adierazpenen (TIMEX3) anotatzaileen arteko adostasuna.

Anotatzaile bikoteak	Micro-average (Markable)	Micro-average (Token)	Macro-average (Markable)	Macro-average (Token)
A – B	0.75	0.75	0.79	0.79
A – C	0.556	0.556	0.479	0.479
B – C	0.444	0.444	0.393	0.393
Guztira	0.583	0.583	0.554	0.554

5. taula: Denbora erlazio hitzen (SIGNAL) anotatzaileen arteko adostasuna.

5.2 Emaitzen ebaluazioa

Etiketatu denbora-adierazpen eta seinaleak alderatuz gero, adostasuna non lortu den edo ez ikusi ahal izan dugu. Alde batetik, *markable* bati etiketa bera jarri zaion neurtu dugu eta beste aldetik, etiketa horiek hedapen bera (token berak) hartu duten. Denbora-adierazpenen kasuan emaitza onak lortu dira, adostasun osoa beti izan da 0,89tik gorakoa testuan jarritako etiketatzen, baina etiketa hutsak kontuan hartuz gero, emaitzek behera egin dute. Seinaleen kasuan emaitzak nabarmen baxuagoak dira. Anotatzaileak bat ez etortzeko arrazoi nagusiak ondokoak izan dira:

- Denbora-adierazpen edo seinale bat etiketatu ez izana edo etiketa huts osagarriak sortu ez izana.
- Etiketatu behar ez zen tokena etiketatu izana. Hau da, denbora-egituratzat hartu da denbora-egitura ez dena.
- Tokenei etiketa ezberdina eman izana. Denbora-egitura zuzen identifikatu, baina etiketa okerra eman izana.
- Denbora-adierazpenen atributuei balio ezberdinak eman izana.

Emaitzen ebaluaziorako, anotatzaileak batu eta adostasunaren edo ez-adostasunaren arrazoiak aztertu ditugu. Gidalerroen interpretazio okerra izan da arazorik handiena; eta, ondorioz, anbiguotasuna edo argitasun falta agertzen zuten atalak berrikusi eta zuzendu ditugu. Halaber, gidalerroak idaztean kontuan hartu ez ziren denbora-egiturak, lagintzat hartutako corpusean

agertu ez zirelako, gehitu ditugu, egituren zerrenda osatzeko. Horrekin batera, denbora-egitura batzuk, “jadanik” eta parekoak, esaterako, alde batera uztea erabaki dugu, ez baitute denbora-informazio berririk gehitzen eta beste hizkuntzetan ere albo uzten direlako. Aldaketa horien bidez espero dugu seinaleetan nagusiki hobekuntza nabaritzea, horien etiketatzea mugatu baita gehien aldaketak egitean.

Etiketa hutsen sorrera eta erabilera ere arazo iturri izan da. Horien bidez, testuan esplizitu agertzen ez den, baina deduzi daitekeen denbora-informazioa adierazten da. Gidalerroetan horien erabilera argiago azaldu dugu eta adibide praktikoak ere egin ditugu anotatzaileen lana hobetzeko.

6 Ondorioak eta etorkizunerako lanak

Lan honen helburua euskarazko testuen ulermen automatikoan urratsa egitea izan da. Horretarako, denbora-informazioa adierazten duten euskarazko denbora-adierazpenak eta erlazio-hitzak aztertu ditugu. Behin azterketa hori eginda, elementu horiek etiketatzeko erabiliko den TimeML denbora markaketa-lengoaia (Pustejovsky et al., 2003a) aukeratu dugu beste hizkuntzatan erabili delako eta euskararako ere egokia delako.

Eredu horri jarraituz, eta zenbait egokitzapen egin ondoren, euskarazko denbora-egiturak kodetzeko baliatuko den etiketatze-eskema definitu dugu. Etiketatzeko eskema horren egokitasuna egiaztatzeko, hiru etiketatzailek 56 esaldiko lagin bat etiketatu dute NewsReader proiektuaren metodologia jarraituz eta CAT tresna (Bartalesi Lenzi, Moretti, & Sprugnoli, 2012) erabi-

Anotatzaile bikoteak	A–B	A–C	B–C	Orokorra
type	0.76	0.64	0.75	0.55
value	0.76	0.68	0.46	0.45

6. taula: type eta value atributuen gaineko adostasuna.

lita. Emaitzak ebaluatu ondoren, definitutako etiketatze-eskeman zenbait zuzenketa egin ditugu argi ez zeuden edota anbiguotasuna sortzen zuten atalak zehazteko. Lan horren ondorioz, gidalerroen lehen bertsioa (Altuna, Aranzabe, & Díaz de Ilarraza, 2014) osatu dugu.

Etorkizuneko lanen artean, gidalerro horiek zabaltzea aurreikusi dugu; izan ere, tamaina handiagoko corpus bat osatzea dugu helburu eta corpus horren azterketak denbora-egituren kasuistika gehituko du.

Corpus horretan, denbora-egitura guztiak (denbora-adierazpenak, erlazio-hitzak eta gertae-rak) eta horien artean sortzen diren denbora-erlazioak markatuko dira denboraren arabera finkatutako irizpideak jarraituz eta garatutako etiketatze-tresnak baliatuz. Corpus hori *gold standard* modura erabiltzea eta eskuragarri jar-tzea da gure asmoa.

Amaitzeko, etiketatze-lanetan laguntzeko, be-rez denbora adierazten duten hitzen zerrenda egin dugu, baina abiarazle-lexikoen zerrenda ho-ri osatzeko asmoa ere badugu. Lan hau idatzi bitartean TempoWordNet (Dias et al., 2014) ga-ratu da. Datu-base horretan *synset* bakoitzari denbora-informazioa esleitu zaio. Abiarazle le-xikoak bezala, TempoWordNet-eko informazioa denbora-egiturak identifikatzeko baliagarria izan-go dela uste dugu. Denboraren araberrako anali-si ona lortzeko, baliabide hori nola integratu ere erabaki behar izango dugu.

Eskerrak

Ikerketa lan hau Eusko Jaurlaritzak emandako ikertzaileak prestatzeko doktoretza-aurreko A modalitateko bekari esker (Erref. zk.: PRE_2013.1_959) egin da.

Eskerrak Itziar Aldaberi corpora osatzeko be-re laguntzagatik eta Arantxa Otegiri formatu arazoetan hain baliagarri izateagatik.

Eskerrak errebisatzaileei lan honen hobekun-tzan hartutako lanagatik.

Eranskina - ABIARAZLE LEXIKOAK

Jarraian abiarazle lexiko izendatu ditugun for-men adibideak batu eta sailkatuko dira.

• IZEN ARRUNTAK

Abendu, abuztu, aldi, apiril, aro, arrats, arratsalde, aste, astearte, asteazken, aste-lehen, asti, azaro, betikotasun, bider, con-tinuum, denbora, eguerdi, egun, ekain, etor-kizun, garai, gau, gaur-bihar, gero, geroal-di, goiz, goizalde, hamarkada, hilabete, igan-de, iluntze, infinitu, iragan, iraganaldi, irail, iraunaldi, iraupen, larunbat, lehen, lehenal-di, maiatz, martxo, memento, mende, milur-teko, minutu, momentu, negua, orain, orai-naldi, ordu, oren, ostegun, ostiral, otsail, sa-soi, segundo, uda, udaberri, udazken, une, urri, urtaro, urtarril, urte, uztail.

• IZEN BEREZIAK

Gabon, San Joan.

• ADJEKTIBOAK

Berantiar, egunkari, goiztiar, hilabetekari.

• ADBERBIOAK

Arestian, aspaldi, atzo, aurki, aurrenik, aur-temein, aurren, bart, behiala, behin, behin edo behin, behinik behin, berandu, berri-ki, berritan, beti, bihar, birritan, edonoiz, egun, etzi, etzidamu, gaur, gaurgero, gero, geurtz, goiz, goizik, harrezkero, herenegun, hirutan, honezkero, horrezkero, iaz, inoiz, lehen, lehen baino lehen, lehenbailehen, lehe-nengo eta behin, lehenik, luzaz, maiz, noiz-bait, noizbehinka, noizean behin, noiznahi, ondoan, orain, oraino, orduan, ostean, oste-ra, sarri, sasoz, sekula, usu.

• DENBORA-ESAPIDE BEREZIAK

2014ko apirilaren 15ean, 2015/09/19, 12:40.

Bibliografia

Aduriz, Itziar, Izaskun Aldezabal, María Jesús Aranzabe, Jose Mari Arriola, Klara Ceberio, Ainara Estarrona, Mikel Iruskieta, Mikel Lersundi, Elisabete Pociello, Larraitz Uri-a, Ruben Urizar, & Edurne Aldasoro. 2008. Euskarazko postposizio-lokuzioen tratamendu konputazionala. Txosten teknikoa, Lengoaia eta Sistema Informatikoak Saila, UPV/EHU. UPV / EHU LSI / TR 07-2008. <http://ixa.si.ehu.es/Ixa/Argitalpenak/>

- Barne_txostenak/1220881618/publikoak/Postposizioen%20barne-txostena.
- Altuna, Begoña, María Jesús Aranzabe, & Arantza Díaz de Ilarraza. 2014. Euskarazko denbora-egiturak etiketatze gidalerroak. Txosten teknikoak, Lengoaia eta Sistema Informatikoak Saila, UPV/EHU. UPV / EHU LSI / TR 01-2014. http://ixa.si.ehu.es/Ixa/Argitalpenak/Barne_txostenak/1414871293/publikoak/Denbora-egiturak%20etiketatzeko%20gidalerroak.
- Altuna, Patxi, Pello Salaburu, Patxi Goenaga, María Pilar Lasarte, Lino Akesolo, Miren Azkarate, Piarres Charriton, Andolin Eguskitza, Jean Haritschelhar, Alan King, Jose Mari Larrarte, Jose Antonio Mujika, Beñat Oyharçabal, & Karmele Rotaetxe. 1985. *Euskal Gramatika Lehen urratsak (EGLU) I*. Euskaltzaindiko Gramatika batzordea, Euskaltzaindia, Bilbo.
- Altuna, Patxi, Pello Salaburu, Patxi Goenaga, María Pilar Lasarte, Lino Akesolo, Miren Azkarate, Piarres Charriton, Andolin Eguskitza, Jean Haritschelhar, Alan King, Jose Mari Larrarte, Jose Antonio Mujika, Beñat Oyharçabal, & Karmele Rotaetxe. 1987. *Euskal Gramatika Lehen Urratsak (EGLU) II*. Euskaltzaindiko Gramatika Batzordea, Euskaltzaindia, Bilbao.
- Bartalesi Lenzi, Valentina, Giovanni Moretti, & Rachele Sprugnoli. 2012. CAT: the CELCT Annotation Tool. In N. Calzolari, K. Choukri, T. Declerck, M. Uğur Doğan, B. Maegaard, J. Mariani, J. Odiijk, & S. Piperidis, editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, pages 333–338, Istanbul, Turkey. European Language Resources Association (ELRA).
- Bittar, André. 2010. *Building a TimeBank for French: a Reference Corpus Annotated According to the ISO-TimeML Standard*. Ph.D. thesis, Université Paris Diderot, Paris. <http://www.linguist.jussieu.fr/~abittar/docs/Bittar-PhD.pdf>.
- Caselli, Tomasso, Valentina Bartalesi Lenzi, Rachele Sprugnoli, Emanuele Pianta, & Irina Prodanof. 2011. Annotating Events, Temporal Expressions and Relations in Italian: the It-TimeML Experience for the Ita-TimeBank. In *Proceedings of the 5th Linguistic Annotation Workshop*, pages 143–151, Association for Computational Linguistics, Portland, Oregon, USA.
- Dias, Gael, Mohammed Hasanuzzaman, Stéphane Ferrari, & Yann Mathet. 2014. TempoWordNet for Sentence Time Tagging. In *Proceedings of the 4th ACM Temporal Web Analytics Workshop (TEMPWEB) associated to the 23rd International World Wide Web Conference*, pages 833–838, Seoul, South Korea. International World Wide Web Conferences Steering Committee.
- Dice, Lee Raymond. 1945. Measures of the Amount of Ecologic Association between Species. *Ecology*, 26:297–302.
- Ferro, Lisa, Laurie Gerber, Inderjeet Mani, Beth Sundheim, & George Wilson. 2003. TIDES 2003 Standard for the Annotation of Temporal Expressions. Txosten teknikoak, MITRE, McLean, USA, September. http://www.mitre.org/sites/default/files/pdf/ferro_tides.pdf.
- Im, Seohyun, Hyunjo You, Hayun Jang, Seungho Nam, & Hyopil Shin. 2009. KTimeML: Specification of Temporal and Event Expressions in Korean Text. In *Proceedings of the 7th workshop on Asian Language Resources in conjunction with ACL-IJCNLP 2009*, pages 115–122, Suntec City, Singapore. Association for Computational Linguistics.
- ISO-TimeML working group. 2008. Language resource management — Semantic Annotation Framework (SemAF) — Part 1: Time and events. International Standard ISO/CD 24617-1(E), ISO. http://lirics.loria.fr/doc_pub/SemAFCD24617-1Rev12.pdf.
- Lorente, Mercé. 2001. Altres elements lèxics. In *Gramàtica del Català Contemporani*. Empúries, Barcelona, pages 831–888.
- Pociello, Elisabete. 2008. *Euskararen ezagutzabasa lexikala: Euskal WordNet*. Ph.D. thesis, Euskal Filologia Saila, Euskal Herriko Unibertsitatea, Leioa. <http://ixa.si.ehu.es/Ixa/Argitalpenak/Tesiak/1204622545/publikoak/2008Tesi-txostena-eranskinak-aurrezpena.rar>.
- Pustejovsky, James, José M Castaño, Robert Inghria, Roser Saurí, Robert J Gaizauskas, Andrea Setzer, Graham Katz, & Dragomir R Radev. 2003a. TimeML: Robust Specification of Event and Temporal Expressions in Text. *New directions in question answering*, 3:28–34.
- Pustejovsky, James, Patrick Hanks, Roser Saurí, Andrew See, Robert Gaizauskas, Andrea Setzer, Dragomir Radev, Beth Sundheim, David

- Day, Lisa Ferro, & Marcia Lazo. 2003b. The TimeBank Corpus. In D. Archer, P. Rayson, A. Wilson, & T. McEnery, editors, *Proceedings of Corpus Linguistics 2003*, pages 647–656, Lancaster, UK. UCREL, Lancaster University.
- Pustejovsky, James, Marc Verhagen, Roser Saurí, Jessica Littman, Robert Gaizauskas, Graham Katz, Inderjeet Mani, Robert Knippen, & Andrea Setzer. 2006. TimeBank 1.2. Txosten teknikoa, Linguistic Data Consortium. <https://catalog.ldc.upenn.edu/LDC2006T08>.
- Setzer, Andrea. 2001. *Temporal Information in Newswire Articles: An Annotation Scheme and Corpus Study*. Ph.D. thesis, University of Sheffield, Sheffield, UK. ftp://ftp.dcs.shef.ac.uk/home/robertg/theses/setzer_thesis.pdf.
- Tonelli, Sara, Rachele Sprugnoli, & Manuela Speranza. 2014. NewsReader Guidelines for Annotation at Document Level. version 4.1. Txosten teknikoa, Fondazione Bruno Kessler. <http://www.newsreader-project.eu/files/2013/01/NWR-2014-2.pdf>.
- Urizar, Ruben. 2012. *Euskal lokuzioen tratamendu konputazionala*. Ph.D. thesis, Euskal Filologia Saila, Euskal Herriko Unibertsitatea, Donostia. <http://ixa.si.ehu.es/Ixa/Argitalpenak/Tesiak/1342621075/publikoak/TESIA>.
- Zubiri, Ilari & Entzi Zubiri. 1995. *Euskal Gramatika Osoa*. Didaktiker, Bilbao.