

**EUSKARAZKO DENBORA-EGITURAK
ETIKETATZEKO GIDALERROAK V1.0**

Begoña Atuna, María Jesús Aranzabe, Arantza Díaz de Illarraza

UPV/EHU / LSI / TR 01-2014

Euskarazko denbora-egiturak etiketatzeko gidalerroak v1.0
UPV / EHU LSI / TR 01-2014

Begoña Altuna, María Jesús Aranzabe, Arantza Díaz de Ilarraza

Laburpena

Testuetan agertzen den denborazko informazioa interpretatu ahal izateko, informazio hori kodetuko duen markaketa-lengoaia behar da, gerora informazio hori automatikoki baliatu ahal izateko. TimeML (Pustejovsky et al., 2003) etiketatze-lengoaia da zabalduena eta euskararako ere erabili dena. Lan honetan ISO-TimeMLren (ISO-TimeML working group, 2008) euskararako moldaketa aurkezten da; ingeleserako sortutako etiketa, atributu eta horien balioak aztertu ostean, euskarazko denbora-egituren informazioa hobekien islatzen dituztenak deskribatzen dira, hain zuzen ere.

Abstract

To interpret the temporal information on texts, a mark-up language that will code that information is needed, in order to make that information automatically reachable. The most used mark-up language is TimeML (Pustejovsky et al., 2003), which has also been chosen for Basque. In this guidelines we present the Basque version of ISO-TimeML (ISO-TimeML working group, 2008). After having analysed the tags, attributes and values created for English, we describe the most appropriate ones to represent Basque time structures' information.

Gaien Aurkibidea

1	SARRERA	4
2	GERTAERAK	5
2.1	Gertaerak adierazteko era sinpleak	6
2.1.1	Aditzek adierazitako gertaerak	6
2.1.2	Izenek adierazitako gertaerak	7
2.1.3	Adjektiboek adierazitako gertaerak	9
2.1.4	Adberbioek adierazitako gertaerak	10
2.2	Gertaerak adierazteko egitura konplexuak	10
2.3	Gertaeren atributuak	12
2.3.1	Gertaeren IDa (eid) eta gertaera-instantziaren IDa (eiid)	13
2.3.2	Klasea (class)	13
2.3.3	Kategoria gramatikala	14
2.3.4	Denbora eta aspektua	14
2.3.5	Polaritatea	15
2.3.6	Modalitatea	16
3	DENBORA-ADIERAZPENAK	16
3.1	<TIMEX3>	16
3.1.1	Hedadura	16
3.1.2	Atributuak	18
3.1.3	Kulturak zehaztutako adierazpenak	27
3.1.4	Dokumentuaren sorrera-unearen etiketatzea	27
3.1.5	<TIMEX3> etiketa hutsak	27
4	SEINALEAK	28
4.1	<SIGNAL>	28
4.2	Aditzak eta seinaleak	29
5	ERLAZIOAK	29
5.1	TLINK	30
5.1.1	TLINKen BNFa	30
5.1.2	relType atributua	30
5.2	Beste atributuak	32
5.3	TLINKak egiteko protokoloa	32
6	CAT Tresna	34
7	OHARRAK	36

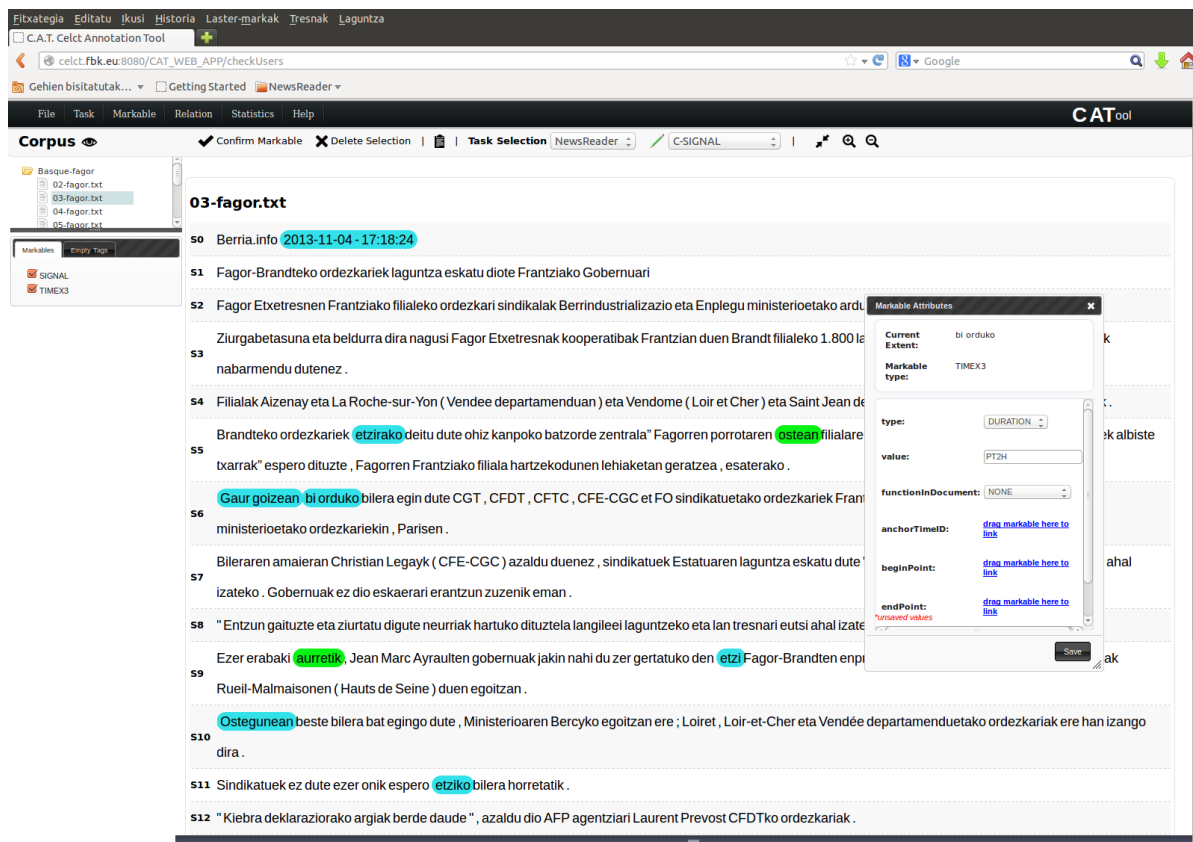
1 SARRERA

Dokumentu honetan euskarazko denbora-egiturak etiketatzeko gidalerroak aurkitzen dira. Hainbat dira hizkuntzan denbora adierazten duten formak: denbora-adierazpenak (*gaur, 2014ko martxoaren 12an, Arratsaldeko bostetatik*); seinaleak¹, bestela esan, erlazio-hitz edo esamoldeak, (*bostak [arte], ondoren, baino beranduago*) eta gertaerak (*etorri, lasterketa*). Denbora-adierazpenen eta gertaeren artean, gainera, denbora-erlazioak sortzen dira (*A baino lehenago B, A eta B batera* etab.). Erlazio horiek TLINK loturen bidez adieraziko dira. Gidalerro hauetan horiek guztiak nola etiketatu eta zein atributu adierazi eta nola azalduko da.

ISO-TimeML (ISO-TimeML working group, 2008) etiketatze-lengoaian oinarritutako <TIME3> etiketak erabiliko ditugu denbora-adierazpenak markatzeko eta <SIGNAL> etiketak “denbora-lokailu edo seinaleentzat. Gertaerek <EVENT> eta denbora-erlazioek <TLINK> hartuko dute. Etiketatuko ditugun elementuen luzera kurtsibaz eta, behar izanez gero, [] kortxete karratu bidez adieraziko da gidalerroetan. Gidalerroak idazteko NewsReader proiektuko gidalerroak (Tonelli et al., 2014) hartu dira eredutzat, euskarazko denbora-egituren analisi eta prozesaketa proiektu horren segidatzat hartu nahi baita.

Eskuzko etiketatzea egitean ez dira etiketen atributu guztiak kontuan hartuko batzuk era automatikoan esleitzen baitira. Anotatzaileak ez du <TIME3> etiketetan `tid`, `valueFromfunction` eta `temporalFunction` atributuez arduratu behar, adibidez. Automatikoki esleitzen diren etiketak ez dira interfazean agertuko, baina bai *output*ean. Etiketatzea Fondazione Bruno Kessler-en (FBK) CAT etiketatzeko tresnarekin (Bartalesi et al., 2012) egingo da. 1. irudian ikus daiteke etiketatze tresnaren interfazea; etiketatu beharreko testua, etiketa motak, denbora-erlazioak eta etiketa hutsak sortzeko baliabideak daude ikusgai besteak beste:

¹TimeML markaketa-lengoaian “signal” erabiltzen da denborazko erlazio bat adierazten duten formak markatzeko.



1. irudia: CAT tresnaren interfazea

6. atalean interfazearen deskribapen sakonagoa eta horren erabilera azaltzen dira. Gidalerro hauen 2. atalean gertaerak nola etiketatu azalduko da. 3. atalean denbora-adierazpenak eta 4.ean denbora-seinaleak ere nola etiketatu azalduko da. 6. atalean CAT etiketatze-tresna nola erabili azalduko da. Amaieran, ohar batzuk gehitu dira gidalerrook eraketa fasean dauden artean informazio gehigarria eman ahal izateko. Gidalerroetan zehar denbora-egiturak eta denbora-erlazioak argitzeko hainbat adibide eskaintzen dira. Adibideotako batzuk euskarazko idatzizko komunikabideetatik hartu dira, baina beste asko beren-beregi sortuak dira ahalik eta adibide argiena erakusteko asmoz.

2 GERTAERAK

Gertaerak edo ebentoak gertatzen diren egoerak edo ekintzak dira. Puntukariak izan daitezke edo denbora-tarte bat iraun. Euskaraz aditz jokatuak (1) zein ez jokatuak (2), izenek (3), adjektiboak (4) eta adberbioak (5) adieraz ditzakete gertaerak.

- (1) Bihar *batuko dira* euskararen inguruko eragileak, egoerari irtenbidea aurkitzeko.
- (2) *Bazkaldu* ondoren siesta egitea osasuntsua dela adierazi dute ikerlariek.

- (3) Azken hiruhilabetean auto *salmentek* gora egin dute Hego Euskal Herrian.
- (4) Enpresa *bideragarria* dela diote langileek.
- (5) Helegite epea *zabalik* egongo da hilaren amaierara arte.

Izenordain anaforikoek ere adieraz dezakete gertaera bat:

- (6) Partidu handiak hauteskundeetan *porrota* pairatzeko beldur dira. *Hori* ekiditea izango da zerrendaburuen eginkizuna.

Adibideotan (1–6) ikusi ahal izan dira euskaraz gertaerak adierazteko formak. Esaldi bakoitzean, gainera, nabarmendutakoaz aparte, beste gertaera batzuk ere topa daitezke, baina argitasunaren izenean adibide bakoitzean mota bakarrari eman zaio garrantzia.

2.1 Gertaerak adierazteko era sinpleak

Orokorrean gertaerei token bateko hedapena esleituko zaie, berez unitatea bi token baino gehiagok osa badezaketeen ere. Har ditzagun berriro aurreko adibideak. Oraingoan gertaera guztiak kortxete artean agertuko dira eta gertaeren burua (etiketatu behar dena) kurtsibaz adieraziko da (7–12):

- (7) Bihar [*batuko* dira] euskararen inguruko eragileak, [*egoerari*] [*irtenbidea*] [*aurkitzeko*].
- (8) [*Bazkaldu*] ondoren [*siesta*] [*egitea*] [*osasuntsua*] [*dela*] [*adierazi* dute] ikerlariek.
- (9) Azken hiruhilabetean auto [*salmentek*] [*gora*] [*egin* dute] Hego Euskal Herrian.
- (10) Enpresa [*bideragarria*] [*dela*] [*diote*] langileek.
- (11) [*Helegite*] epea [*zabalik*] [*egongo* da] hilaren amaierara arte.
- (12) Partidu handiak [*hauteskundeetan*] [*porrota*] [*pairatzeko*] [*beldur* dira]. [*Hori*] [*ekiditea*] [*izango* da] zerrendaburuen [*eginkizuna*].

Goiko (7–12) adibideetan bi gauza ikus daitezke: hasieran esan bezala, alde batetik hainbat gramatika-kategoriatako hitzek adieraz ditzaketela gertaerak (izenak: (7)-ko *egoera*; aditz jokatuak: (11)-ko *egongo da*; adjektiboak: (8)-ko *osasuntsua* etab.) eta bestetik, gertaera adierazten duen sintagma hitz bat baino gehiagoz osatua bada ere, token bakarrak hartuko duela etiketa (kortxetez sintagma, kurtsibaz buruak).

2.1.1 Aditzek adierazitako gertaerak

Aditz guztiek gertaera bat adierazten dutela kontuan hartuta, guztiak etiketatuko dira, baina aditz-sintagmaren burua baino ez da etiketatuko. Aditz trinkoen (13) eta forma ez-jokatuen (14) kasuan buru horrek izango du informazio guztia, baina beste aditzei (15) dagokienez, informazioaren zati bat aditz laguntzaileak izango du. Informaziorik ez galtzeko, ordea, atributu batzuk definituko dira informazio hori gertaerarekin batera gorde ahal izateko. Hurrengo adibideetan, (13–15), argitasunaren izenean azpimarratu nahi diren formak baino ez dira nabarmendu, besterik ere badagoen arren.

- (13) Zientzialariek urteak *daramatzate* fusiozko energia nuklear merkearen bila.
- (14) Akordiorik *lortu* ezean, grebari eutsiko diote osasun saileko langileek.
- (15) Goizeko 8etan *atera* dira mendizaleak mendi tontorrerantza.

2.1.2 Izenek adierazitako gertaerak

Izen guztiak ez dira gertaera izango. Izena gertaera izango bada, honako testuinguruetak bitan agertu behar da ((TimeML Working Group, 2010)):

- IZENAK minutuak/egunak/urteak... iraun zuen.
- IZENA oso azkar/bat batean gertatu zen.
- IZENA denbora-adierazpenean gertatuko da.
- IZENA/IZENAK denbora adierazpenean hasi/jarraitu/bukatu zen/zuen.

Adibidez “galleta” izena ez da gertaera izango, baina bai, adibidez, “uholde”:

- (16) a. Galletak bi ordu iraun zuen.
b. Uholdeak bi ordu iraun zuen.
- (17) a. * Galleta oso azkar gertatu zen.
b. Uholdea oso azkar gertatu zen.
- (18) a. * Galleta etzi gertatuko da.
b. Uholdea etzi gertatuko da.
- (19) a. * Galleta atzo hasi zen.
b. Uholdea atzo hasi zen.

Goiko (16–19) adibideetan ikus daiteke “galleta” hitza testuinguru baten baino ez dela egokia, eta, ondorioz, ez da gertaeratzat automatikoki hartuko. “Uholde” hitza, ordea, egokia da lau testuinguruetan eta, bi baino gehiagotan (lauretan) egokia denez, gertaera izango da.

Hala, honako irizpideak jarraituko dira izenen bidez adierazitako gertaerak etiketatzeko:

Aurreko testuinguruetan egokiak diren izenak BETI etiketatuko dira. Horrelakoak izango dira, adibidez, (20) eta (21) esaldietako *leherketa*, *akordio*, *greba* eta horien antzekoak: *krisi*, *batzar*, *erreskate*, *eraso*, *lurrikara*... Aditz-izenak ere multzo honetakoak izango dira, adibidez, (22) eta (23) esaldietako *lekualdatze* eta *euste*, eta horien antzekoak: *sakabanatze*, *galtze*...

- (20) *Leherketak* hainbat zauritu utzi ditu eta inguruak husteru behartu du polizia.
- (21) *Akordiorik* lortu ezean, *grebari* eutsiko diote osasun saileko langileek.
- (22) Lantegiaren *lekualdatzea* galarazteko asmoz atera dira langileak kalera.
- (23) Sindikatuek *grebari euste*a erabaki dute patronalaren entzungortasunaren aurrean.

Aurreko izenez gain beste batzuk ere badira gertaerak adierazten dituztenak, baina testuinguru zehatzetan baino ez dira etiketagarri izango:

Egoera bereizgarriak (*sortal states*) horrelakoak izan daitezke:

- izen agentiboak (*agentive nominals*): *irabazle, gidari, bezero, bazkide...*
- ogibideak, eginkizunak eta postuak adierazten dituzten izenak: *irakasle, doktore, zuzendari...*
- adierazle zurrunak, hau da, beti pertsona berari erreferentzia egingo diotenak: *Estatu Batuetako 27. presidentea, 17. aita santua...*

Horiek BAKARRIK izango dira gertaera bezala etiketagarri predikazio-osagarriaren buru badira, bestela esan, honako aditzen osagarri badira:

- emendiozko predikatuak: *izan, iruditu...*

(24) Atlético Madrid Ligako *irabazle* izan da aurten.

- predikatu inkoatiboak (egoera aldaketa bat aditzera ematen dutenak (Altuna et al., 1987)): *bihurtu, sartu...*

(25) Chávezen heriotzaren ondoren Maduro *presidente* bihurtu zen.

- aspektuzko predikatuak: *hasi, jarraitu, amaitu, ibili, ari izan...*

(26) Maialen Mungian *liburuzain* dabil.

- egoera aldaketazko predikatuak: *erretiratu, aukeratu, izendatu...*

(27) Ikasleen *ordezkari* aukeratu zuten Garazi ostegunean.

- ebaluazio eta deskribapen predikatuak: *kontsideratu, -tzat hartu, deskribatu.*

(28) Zuzendariak postaria bere *idazkaritzat* hartu zuen.

Izenezko predikatuetan ere sintagma burua baino ez da etiketatuko. (29) adibidean sintagma *Venezuelako presidente* bada ere, “presidente” hitza baino ez da etiketatuko gertaera bezala.

(29) Chvezen heriotzaren ondoren Maduro [Venezuelako *presidente*] bihurtu zen.

Hitz anitzeko izen-predikatuak Sintagmaren burua baino ez da etiketatuko: *krisi* ekonomikoa, Aita Santuaren *bisita* (30). Berdin jokatu da “Bigarren Mundu *Gerra*” edo “Industri *Iraultza*” modukoekin.

(30) Dena prest dago astearteko Aita Santuaren *bisitarako*.

Izenaren modifikatzaile diren izenak EZ dira etiketatuko hitz elkartuetako osagai badira; hau da, mendekotasunezko izen konposatuetan ez dira mugatzaileak etiketatuko, esaterako, *hauteskunde-porrota* eta *itxaron-gela* hitz elkartuetako *hauteskunde* eta *itxaron* osagaiak. Etiketatuko dira, ordea, mugakizunak: *hauteskunde-porrota* hitz elkartuko *porrota* osagaia.

(31) Martxoko *hauteskunde-porrotaren* ostean, partiduko idazkari nagusiak ez-ohiko kongresua deitu du.

Magnitude izenak dituzten egiturak magnitude izenak (*tenperatura, tamaina, biztanleria, intentsitatea, defizita...*) ere gertaera izango dira hainbat testuingurutan (32–34). Horiek entitate bat (pertsona bat, pertsona multzoa, leku bat...) hartuko dute argumentu moduan eta eskala baten balio bat izango dute, zenbakizkoa izan edo ez (*altu, baju, bero, hotz, 5°, 20 m...*).

- (32) Lurrazaleko *tenperaturak* gora egitea da berotegi efektuaren arriskurik handiena.
- (33) Donostiak 180.000 pertsonako *biztanleria* du.
- (34) Estatuaren *defizita* % 0,3 murriztu da aurten.

Aurreko baldintzak betetzen ez dituzten izenak baina gertaera izan daitezkeenak ere baditugu azkenik: *esperantza* (35), *nahia, desira, zintzotasun, oniritzi* (36)... Horiek ere gertaera izango dira jarraian datozen adibideen modukoetan:

- (35) Liskarrak laster amaituko direlako *esperantza* du gobernuak.
- (36) Enpresak epailearen *oniritziarekin* egin du gaur goizeko ekintza.

2.1.3 Adjektiboek adierazitako gertaerak

Gertaerak adierazten dituzten adjektiboek orokorrean egoera estatikoak adierazten dituzte. Izenen osagarri bezala (min *jasanezina, bidezko* epaiketa, lorategi *ederra*) edo predikatu-osagarri moduan (*haurdun* dago) ager daitezke adjektiboak. Izenen osagarri agertzen diren adjektiboak EZ dira etiketatuko, ezta (37) moduko perpausetan agertzen direnak ere.

- (37) Txinako bandera *gorria* da. EZ DA ETIKETATUKO.

Predikatu-osagarri funtzioa hartzen duten adjektiboak, ordea, honako aditzekin agertuz gero izango dira etiketagarri:

- emendiozko predikatuak: *izan, iruditu...*
 - (38) Sartu den emakumeak *gaztea* dirudi.
- aspektuzko predikatuak: *hasi, jarraitu, amaitu...*
 - (39) Egoerak *lasai* jarraitzen du errebeldeek kontrolatzen duten lurraldean.
- pertzepzio predikatuak: *iruditu, eman...*
 - (40) *Iskanbilatsuak* diruditen zaleei estadiorako sarrera debekatzeko baimena aldarrikatu du klubak.
- ebaluazio eta deskribapen predikatuak: *kontsideratu, -tzat hartu, deskribatu...*
 - (41) Anbotoko igoera *nekagarritzat* hartzen dute askok.

Adjektibo predikatuak etiketatzeko, beraz, ondoko gidalerroak jarraituko dira:

Adjektibo predikatiboak SOILIK izenak adierazitako entitatearen ezaugarri ez-egonkor bat adierazten badute etiketatuko dira. Ezaugarri ez-egonkorak diren erabakitzeo, ondoko hiru baldintzetatik bat bete behar izango dute:

- Ezaugarria argi eta garbi ez-egonkorra da.

(42) Mikel *gaixo* dago.

(43) Egoera *larria* da.

- Ezaugarria denborako une edo tarte bati lotuta dago.

(44) XIX. mendean Euskal Herria ekonomikoki *atzeratua* zen.

- Ezaugarria eztabaidagai edo iritzi edo jakintza moduan agertzen da.

(45) Itsasontzian zihozen gehienak *euskaldunak* ziren.

(46) Elkarretaratuek *euskaldunak* direla aldarrikatu dute.

Adjektibo bat etiketatzeko zalantzarik izanez gero, EZ etiketatu.

2.1.4 Adberbioek adierazitako gertaerak

Adberbioak predikatu-osagarri funtzioa betetzen dutenean etiketatuko dira. [*Etenda*] *dago* (47) edo [*zail*] *izan* (48) modukoak izango dira horiek. [*Korrika*] *joan*, [*Kolokan*] *egon* eta [*ezbaian*] *jarri* modukoak eta [*kontuan*] *hartu*, [*aurrera*] *egin* edo [*martxan*] *jarri* (49) modukoak ere multzo honetan sartuko ditugu.

(47) Produktzioa *etenda* dago enpresaren hiru lantegitan.

(48) Errealak *zail* du datorren denboraldian Europan jolastea.

(49) Ertzaintzak jarraian jarri zuen *martxan* suteentzako protokoloa.

Multzo honetan sartuko ditugu ere [*neke-neke*] *eginda egon*, [*ziplo*] *erori* eta gisakoak.

Dena dela, kontuan izan behar da etiketatze-lanean aurrera egin ahala, gertaerak adierazteko beste forma batzuk identifikatu eta deskribatu beharko direla.

2.2 Gertaerak adierazteko egitura konplexuak

Aurreko paragrafoetan maiz ikusi ahal izan dugunez, hainbat gertaera egitura berean ager daitezke:

- aspektuzko egiturak: gertaera bat adierazten duen elementu bat + aspektuzko aditz bat (*hasi*, *jarraitu*, *amaitu*, *ekin*, *utzi*...) edo izen bat (*hasiera*, *amaiera*)...

(50) Julia Robertsek zinea [*egiteari*] [*ekin dio*] berriz ere.

(51) Udal garbiketa zerbitzuak hondartzak [*atontzen*] [*amaitu du*].

Aspektuzkoa zein beste gertaera independenteki etiketatuko dira. Bi gertaerak ALINK baten bidez (aurrera azalduko dena) lotuko dira.

- predikatu inkoatiboak: *bihurtu* (52), *sartu* moduko aditzak + sortutako egoera adierazten duen elementuaz osatzen dira (alderantzizko ordenan). BIAK etiketatuko dira gertaera independente moduan:

(52) Umea jaiotzean, Mirari [*amona*] [*bihurtu*] zen.

Horiek ere ALINK baten bidez lotuko dira.

- kausazko predikatuak: horiek kausa adierazten duen aditz bat (*eragin*, *ahalbidetu*, *ekidin*, *eraman*, *sortu*...) eta ondorioa adierazten duen beste aditz batez osatuta daude. Bi motatakoak izan daitezke:

- jatorrizko gertaera + ondorio gertaera + kausazko aditza (hiru gertaerak euren aldetik etiketatuko dira):

(53) Atzo [*euriteak*] [*uholdeak*] [*eragin zituen*].

- ondorio gertaera + kausazko aditza:

(54) Jonek [*sutea*] [*abiarazi zuen*].

TimeMLk ((TimeML Working Group, 2010)) ez du gertaera horiek lotzeko LINK loturarik, denbora-erlazioaz (TLINK) aparte. NewsReader proiekturako (Tonelli et al., 2014), ordea, CLINK, kausazko lotura, sortu da eta euskarazko testuetan ere erabiliko dela aurreikusi da.

- aditz arinak: gertaera izena + aditz arina (*eman*, *egin*, *izan*, *jazo*...).

(55) Tontorrera igotzeko hiru saiakera egin eta gero, gaur goizeko eguraldi txarra dela kausa, mendizale taldeak [*amore*] [*eman du*] eta kanpalekura itzultzea erabaki du.

(56) Egoerari [*buelta*] [*eman*] ezinik daude.

- emendiozko egiturak: *izan*, *egon* edo *iruditu* aditza + osagarria (*buru izan* (57), *beldurrak egon* (58), *gose izan* (59)). BI elementuak etiketatuko dira, bigarrena, goian (gertaerak adierazten dituzten izen eta aditzen ataletan) azalduko irizpideen arabera.

(57) Emakume bat [*izango da*] ikerketa sailaren [*buru*].

(58) Auzokoak [*beldurrak*] [*daude*] kaleratzea noiz gertatuko zain.

(59) Munduan une honetan milioika ume [*gose*] [*dira*].

Ez dago aditzaren eta osagarriaren artean lotura adierazteko biderik TimeMLn, NewsReader-en, ordea, GLINK gramatika lotura erabiltzen da. Loturen gaineko gogoeta egitean azalduko da honelako erlazioak nola adierazi.

- Magnitude izenak dituzten egiturak: magnitude izenak (tenperatura, tamaina, biztanleria, intentsitatea...) hiru egituratan ager daitezke:

- Magnitude izena x da.

(60) Euskal Herriaren *biztanleria* 3 miliokoa da.

– Norbaitek/zerbaitek x-ko magnitude izena du.

(61) Robin Foodek 500.000 pertsonako *ikusleria* du.

– Magnitude izena x-tik y-ra igo/jaitsi/aldatu da.

(62) Amazonaseko oihan tropikalaren *tamaina* milioi bat hektareatik erdira jaitsi da.

WordNet ((Pociello, 2008)) kontsultatu ostean, hauek dira “magnitude” hitzaren hiponimo bezala agertu direnak: *handitasun*, *tamaina*, *proporzio*, *kantitate*, *kopuru*, *neurri*, *ugaritasun*, *zabaltasun*, *maila*, *hedadura* eta *larritasun*. Ondoriozta daiteke, beraz, hitzok berriki aipatutako testuinguruetan gertaera izan daitezkeela, aurretik zerrendaturikoekin batera.

Kasu horietan funtzio izena eta aditza etiketatuko dira, bakoitza bere aldetik. TimeMLk ez du bi elementu horiek lotzeko erlazio zehatzik, baina NewsReader-en GLINK erabiltzen da.

- Gertaera instantziak: gertaera-adierazpen batzuek bi instantzia edo gehiagori egiten diote erreferentzia. Adibidez:

(63) Realak entrenamendua izango du astelehen eta asteartean.

(63) adibidean bi gertaera ezberdin adierazten dira: *asteleheneko entrenamendua* eta *asteartekoa*. Kasu horretan gertaera bakarra, baina bi gertaera instantzia sortuko dira (erabaki behar da nola), baita bi denbora erlazio ere.

2.3 Gertaeren atributuak

Gertaren ezaugarriak adierazteko <EVENT> etiketak atributu batzuk hartuko ditu. Honakoak dira gertaerek hartuko dituzten atributuak:

- Gertaeraren identifikatzailea (*eid*) eta gertaera-instantzien identifikatzaileak (*eiid*).
- Klasea
- Katetoria gramatikala
- Denbora
- Aspektua
- Polaritatea
- Modalitatea

Hemen ikus daiteke EVENT etiketaren BNFa (Backus-Naur formatua):

```
attributes ::= eid eiid class tense aspect pos polarity [modality]
eid ::= e<integer>
eiid ::= ei<integer>
class ::= REPORTING | PERCEPTION | ASPECTUAL | I_ACTION |
        I_STATE | STATE | OCCURRENCE
tense1 ::= PRESENT | -PRESENT | 'NONE'
```

```

tense2 ::= PAST | -PAST | 'NONE'
aspect1 ::= PERFECT | -PERFECT | 'NONE'
aspect2 ::= FUTURE | -FUTURE | 'NONE'
pos ::= ADJECTIVE | NOUN | VERB | PRONOUN | ADVERB | OTHER
polarity ::= NEG | POS {default, if absent, is POS}
modality ::= AHAL | NAHI | BEHAR

```

2.3.1 Gertaeren IDa (eid) eta gertaera-instantziaren IDa (eiid)

Atributu horiek automatikoki esleitzen dira.

2.3.2 Klasea (class)

Zazpi gertaera mota dago:

1. Diskurtsiboak (REPORTING): *esan, adierazi, aipatu, galdetu...*
 - (64) Endesak *iragarri du* Garoña berriro zabaltzeko baimena eskatuko duela.
2. Pertzepziozkoak (PERCEPTION): *ikusi, entzun...*
 - (65) Epaileak gaur *entzungo ditu* inputatuak.
3. Aspektuzkoak (ASPECTUAL): *hasi, amaitu, jarraitu...*
 - (66) Atzo *hasi zen* erreklamazioak egiteko epea.
4. Ekintza intensionalak (*intensional actions*); testuan esplizituki agertu eta gertaera adierazten duten argumentu bat aukeratzen duten gertaera dinamikoak (I_ACTION): *saiatu, aztertu, atzeratu, ekidin, eskatu, zin egin, izendatu, aldarrikatu...*
 - (67) Endesak iragarri du Garoña berriro zabaltzeko baimena *eskatuko* duela.
5. Egoera intensionalak (*intensional states*) (I_STATE): *sinestu, pentsatu, nahi izan, espero izan, gorrotatu, behar izan, prest egon, gai izan...*
 - (68) Endesak *pentsatzen* du Garoña berriro zabaltzeko *gai izango* dela.
6. Egoerak (zirkunstantziak) (STATE): egoera intensionalak ez diren beste guztiak.
 - (69) Maika *haurdun* dago.
7. Jazoerak (OCCURRENCE): gertatzen diren ekintzak orokorrean.
 - (70) Bart *sutea* piztu zen lantegi baten inguruan.

2.3.3 Kategoria gramatikala

- Aditzak
- Izenak
- Adjektiboak
- Adberbioak
- Izenordainak (anaforikoak)

2.3.4 Denbora eta aspektua

Euskaraz aspektua eta tempusa edo aditz-denbora ez da inguruko hizkuntzetan bezala gauzatzen. Hizkuntza indoeuropar askok orainaldiaren, lehenaldiaren eta geroaldiaren arteko bereizketa egiten dute eta modalizazioa aditz askeekin egiten dute. Euskaraz, ordea, geroaldia aspektuaren bidez adierazten da eta modalizazioa aditz nagusiaren baitan ager daiteke.

Hala, euskararako **aspect** (1 eta 2) eta **tense** (1 eta 2) atribuentzat balio ezberdinak proposatu dira, baita **mod** atributuarentzat balio zehatzak ere. Alde batetik, euskaraz historikoki horrela egin izan delako eta, bestetik, denbora-informazioaren analisia analisi-katean txertatu ahal izateko, honako balioak proposatu dira atributu horientzat:

Aspektua euskaraz aditz nagusiaren formak ematen du: -TU, -TUKO, -TZEN edo aditz trinkoen forma zehatzak. Horien arabera, aspektuak balio hauek har ditzake euskaraz:

- PERFECT: perfektiboa, burutua.

(71) Unai janaria *prestatu du*.

- -PERFECT: ez-perfektiboa, ez-burutua.

(72) Enpresak 3.500 langile *ditu*.

- FUTURE: geroaldia adierazten duena.

(73) Reala datorren denboraldian UEFA Ligan *ibiliko da*.

- -FUTURE: geroaldia adierazten ez duena.

(74) Gontzal egunero *joaten da* metroz lanera.

Burututasunaren eta gerokotasunaren arteko gurutzaketa gertatzen da euskarazko aditzetan eta horregatik euskaraz bi aspektu atributu sortu behar izan dira (**aspect1** (+/- PERFECT) eta **aspect2** (+/- FUTURE)). Aditzek horrelako aspektu konbinaketak hartu ahal izango dituzte (1. taula) eta ondorioz, anotatzaileak aspektu atributu bakoitzean ezaugarri bat bete behar izango du:

aspect2/aspect1	PERFECT	-PERFECT
FUTURE		sartuko dira, egingo dute
-FUTURE	sartu dira, egin dute	sartzen dira, egiten dute

1. taula: Euskarazko aspektuak

(75) adibidean, aspektua etiketatzeko adibide bat ikus daiteke:

- (75) Vladimir Putin Errusiako presidentearekin negoziazioetan <EVENT aspect1='PERFECT' aspect2='-FUTURE'>izanda<\EVENT> Poroxenko.

Aurreko adibidean (75) ikus daitekeenez, aspektua bi atributuren bidez adierazi da, aspektuak bi “dimentsio” baititu: “FUTURE”(aspect2) eta “PERFECT”(aspect1).

Aurretik aipatu bezala, euskaraz lehenaldia eta orainaldia baino ez dira denboratzat hartzen geroaldia aspektuaren bidez adierazten baita. Hauek dira euskarazko aditz denbora adierazteko erabiliko diren atributuak:

- PRESENT Orainaldia adierazten duena.

(76) Unai janaria *prestatzen ari da*.

- -PRESENT Orainaldia adierazten ez duena.

(77) Enpresak 3.500 langile *zitu*en.

- PAST Lehenaldia adierazten duena.

(78) Lehen langileek eskubide gehiago *zuten*.

- -PAST Lehenaldia adierazten ez duena.

(79) Gontzal egunero *joaten da* metroz lanera.

Aspektuarekin gertatzen den bezala, tempusa ere bi dimentsiotan uler daiteke. Tempus edo aditz-denboraren balioa ere gurutzatu egiten da euskaraz (2. taula) eta horregatik **tense1** (+/- PRESENT) eta **tense2** (+/- PAST) atributuak proposatu dira:

tense2/tense1	PRESENT	-PRESENT
PAST		sartu ziren, egin zuten
-PAST	sartzen dira, egiten dute	sartuko lirateke, egingo lituzkete

2. taula: Euskarazko aditz-denborak

Ondoko (80) adibidean, tempus atributuen adibidea aurki daiteke:

- (80) Vladimir Putin Errusiako presidentearekin negoziazioetan <EVENT tense1='PRESENT' tense2='-PAST'>izanda<\EVENT> Poroxenko.

2.3.5 Polaritatea

Negatiboa (NEG) ezezka agertzen diren gertaerentzat (81) eta positiboa (POS) beste guztientzat (82).

- (81) Aitor Kerejeta Hondarribiko alkatea ez <EVENT polarity='NEG'>da joan<\EVENT>, eta Eusko Jaurlaritzak ere ez <EVENT polarity='NEG'>du bidali<\EVENT> beste ordezkariarik.
- (82) Ukrainak eta matxinoek su-etena urratzea <EVENT polarity='POS'>egotzi diote<\EVENT> elkarri.

2.3.6 Modalitatea

Aditz modalei buruzko informazioa adierazteko erabiltzen da. Aditz modalik badago, aditz nagusiaren atributu moduan emango da horren berri. Aditz modalak *ahal* (*ezin* polaritate negatiboaz adieraziko da), *behar* eta *nahi* dira. Ez da ahaztu behar ere euskaraz ahalera aditz multzoan bertan adieraz daitekeela beste aditz bat behar izanik gabe. *Daiteke* moduko aditzen ahalera ere atributu honen bidez adieraziko da.

Ondorioz, beste hizkuntzetarako TimeMLn ez bezala, euskaraz, (83) adibidean ageri denez, modalitatea balio finkoak hartuko ditu: AHAL, BEHAR, NAHI.

- (83) Beren habitatean <EVENT mod='NAHI'>ikusik</EVENT> eta <EVENT mod='NAHI'>ikasi nahi izan zituen </EVENT> animaliak.

3 DENBORA-ADIERAZPENAK

Denbora-adierazpenek denbora esanahia dute eta perpausaren denbora-kokapena eskaintzen dute, hau da, eurek seinatzen dute perpausoko 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.

Hizkuntzaren prozesamenduari (HP) begira, horiek ISO-TimeML denbora-etiketatzeko lengoaiatik ateratako <TIMEX3> etiketaren bidez etiketatuko ditugu, euren informazio eta ezaugarriak prozesatzailearentzat erabilgarri bihurtzeko.

3.1 <TIMEX3>

<TIMEX3> etiketa ISO-TimeML markaketa-lengoaiatik (ISO-TimeML working group, 2008) aterata dago eta iraupenak (*5 ordu*) zein uneak (*2014ko otsailaren 12an*) etiketatzeko erabiltzen da. Denbora-adierazpenak absolutuak (*2014ko otsailaren 12an*) zein erlatiboak (*gaur*) izan daitezke. Gertaera baten araberrakoak (*etorri eta [bost minutura]*) edo hainbat uneren multzoa ere izan daitezke (*astelehenetan*).

3.1.1 Hedadura

Orokorrean <TIMEX3> etiketa hartuko duen denbora-adierazpenaren hedadura ahalik eta txikiena izan behar da, zehatzago esanda, abiarazle lexikoa duten adierazpenak bakarrik:

- (84) [*Aste*] honetan ez da idazkaritza zerbitzurik izango Fakultatean.
(85) Otso bat atzeman zuten baserri batzuen inguruan [*bart*] [*arratsean*].
(86) [*2014ko*] [*otsailaren*] [*12an*] hasi zuen Jennifer McArthurrek munduko bira.
(87) Ez da [*orain*] *arte* inolako aldaketarik ikusi ikaslearen jarreran.

Aurreko adibideetan (84–87) kortexte artean agertzen direnak abiarazle lexiko deitu ditugunak dira, denbora-adierazpena, ordea, letra etzanez agertzen den guztia. Abiarazle lexikoak semantikoki denbora adierazten duten hitz edo zenbaki-egiturak dira, hala nola *aste* (84), *bart* (85), *otsail* (86) edo *orain* (87). Abiarazle horiek ez dira beti denbora-adierazpenaren pareko izango: (84) adibidean “aste” izango da abiarazlea, baina “aste honetan” denbora-adierazpena. Denbora abiarazle lexikoak,

beraz, denbora-adierazpenak aurkitzeko pista izango dira soilik. Orotara, honakoak izan daitezke abiarazle lexiko (Ferro et al., 2005)-en oinarrituta (3):

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 arabeko abiarazle lexikoen sailkapena

Baina gerta daiteke abiarazle lexikorik gabeko egiturek ere denbora adieraztea: *batzuetan, gehietan, handitan*. . . Ondorioz, denbora-adierazpenak honako luzera eta forma har ditzake abiarazle lexikoa izan ala ez:

- izen sintagmak: *goizean, astearte arratsaldean, azken bi urteak, etorkizuna, batzuetan, handitan*.
- adjektibo sintagmak: *eguneroko, hilabetekari, goiztiar, zinez berantiar*.
- adberbio sintagmak: *sarri, gaur, berandu, beti, berriki, oso berandu, oso maiz*.
- postposizio askedun izen sintagmak: *9ak irian, bostak inguru, zazpiak aldera*.
- perpaus bereziak: *badira hiru aste, duela bost minutu, orain dela 25 urte, joan den astean, datozen egunetan*.
- zenbakizko egiturak: *2014/05/05, 14:33*.

Edo kasu urri batzuetan abiarazle lexikoa agertu arren, denbora-adierazpena ez izatea (88) adibidean nabarmendutako kasuan, adibidez:

(88) Aizkorrira igo zen Iñaki atzo. *Atzokoak* neke-neke eginda utzi zuen

Denbora-adierazpen batzuek adierazpen konposatu itxura izan dezakete. Adierazpenok etiketa bakarra edo bi etiketa erabiltzea ondoko irizpideen bidez erabakiko da. Denbora adierazpen bat edo bi diren jakiteko, osagai sintaktikoen *testa* egingo da:

(89) Batzarra arratsaldeko 5etan izango da.

(90) *Arratsaldeko, batzarra 5etan izango da.

(91) *5etan izango da, arratsaldeko izango den batzarra.

Aurreko adibideetan “arratsaldeko 5etan” denbora-adierazpena adierazpen bakarra edo bitan deskonposa daitekeena den frogatu da. (89) adibidean denbora-adierazpen osoa dago, “arratsaldeko 5etan”, eta (90–91) adibideetan denbora-adierazpena banandu egin da. Adierazpena banantzean

esaldi ez-gramatikalak sortu direnez, “arratsaldeko 5etan” denbora-adierazpen bakarra dela eta etiketa bakarra hartuko duela ondoriozta daiteke. Hala, *testa* egin ostean honako sailkapena lortu dugu:

- ETIKETA BAKARRA: adierazpen-unitate bakarra osatzen dutenean:
 - (92) 2014ko martxoaren 4a
 - (93) arratsaldeko 4etan
 - (94) aurtengo uda, aurtengo lehen hiruhilekoa
 - (95) 3 urte, 5 hilabete eta 24 egun
 - (96) ordu biak eta 5 minutu
 - (97) atzo arratsaldez, gaur goizean
 - (98) Ostiralean 8etan
 - (99) orain dela 30 urte
- BI ETIKETA: bi adierazpen-unitate bereiz direnean (100–101) edo (102) adibidean bezala, denboran kokatuta dagoen iraupena bada.
 - (100) [Atzo] baino [bi egun] lehenago, [iraungitze data] baino [bi egun] beranduago
 - (101) [1984tik] [1992era]
 - (102) [Urteko] [lehen 6 hilabeteak] - denboran kokatuta dagoen iraupena baita.

Bi etiketa hartzen duten denbora-adierazpen konplexuen kasuan, bi adierazpenak denbora-erlazio (TLINK) baten bidez lotuko dira (aurrerago azalduko da hau TLINKak definitzen direnean).

ADI! (102) adibidean denbora baten kokatuta dagoen iraupen bati buruz ari gara. Ez da berdina izango “Urteko lehen seihilekoa”ri buruz hitz egitea (114), hau une bat izango baita eta etiketa bakarra hartuko du.

3.1.2 Atributuak

Denbora-adierazpen horien guztien informazioa azalarazteko, hurrengo azpiataletako atributuak definitu dira. Hona hemen etiketa honek hartzen dituen atributuen laburpena:

- tid: makinak automatikoki sortzen duen identifikatzailea
- type: mota. Denbora-adierazpenaren mota adierazteko 4 balioen bidez: DATE (eguna baino handiagoa), TIME (eguna baino txikiagoa), DURATION (irupena) eta SET (errepikapena)
- value: balioa. ISO-8601 estandarrean oinarritutako balioa
- mod: denbora-modifikatzaileak adierazteko
- temporalFunction: adierazpen baten balioa denbora-funtzio baten bidez kalkulatu behar den esateko
- anchorTimeID: denbora-adierazpen baten erreferentzia puntuaren id-a adierazteko

- `functionInDocument`: denbora-adierazpenaren funtzioa adierazteko. `CREATION_TIME` (sorrrera data) edo `NONE` (beste guztientzat)
- `valueFromFunction`: `valueFromFunction`-en balioa dagokion `temporalFunction` baten IDaren balioa da eta horrek `<TIMEX3>`-ak adierazten duen denbora denbora-funtzio horren *outputa* dela adieraziko du
- `beginPoint` eta `endPoint`: iraupenaren gaineko informazioa zabaltzeko
- `freq`: gertaeraren maiztasuna adierazteko
- `quant`: gertakaria denbora tarte batean zenbatetan gertatu den adierazteko
- `comment`: komentarioak

Jarraian `<TIMEX3>` etiketaren atributuen Backus-Naur notazioa azalduko da:

TIMEX3 etiketaren atributuen BNFa

```

attributes ::= tid type [functionInDocument] [beginPoint] [endPoint] [quant] [freq]
              [temporalFunction] (value | valueFromFunction) [mod] [anchorTimeID]
tid ::= ID
tid ::= TimeID
TimeID ::= t<integer>
type ::= 'DATE' | 'TIME' | 'DURATION' | 'SET'
beginPoint ::= IDREF
beginPoint ::= TimeID
endPoint ::= IDREF
endPoint ::= TimeID
quant ::= CDATA
freq ::= CDATA
functionInDocument ::= 'CREATION_TIME' | 'EXPIRATION_TIME' | 'MODIFICATION_TIME'
                      | 'PUBLICATION_TIME' | 'RELEASE_TIME' | 'RECEPTION_TIME'
                      | 'NONE' (default, if absent, is 'NONE')
temporalFunction ::= 'true' | 'false' (default, if absent, is 'false')
{temporalFunction ::= boolean}
value ::= CDATA
valueFromFunction ::= IDREF
{valueFromFunction ::= TemporalFunctionID}
TemporalFunctionID ::= tf<integer>}
mod ::= 'BEFORE' | 'AFTER' | 'ON_OR_BEFORE' | 'ON_OR_AFTER' | 'LESS_THAN'
        | 'MORE_THAN' | 'EQUAL_OR_LESS' | 'EQUAL_OR_MORE' | 'START' | 'MID'
        | 'END' | 'APPROX'
anchorTimeID ::= IDREF
anchorTimeID ::= TimeID

```

Atributuen deskribapen osoagoa egingo da jarraian. Atributuen definizioaz aparte, etiketatze adibideak ere emango dira euskarazko denbora-adierazpenak nola etiketatu azaltzeko.

tid Denbora-adierazpen guztiei automatikoki esleituko zaien identifikatzailea.

type (mota) atributuak lau balio posible ditu:

- **date**: eguna bezain luze edo luzeagoak diren egutegiko datak adierazteko: *ostirala, otsaila, aurtengo uda* eta abar.
- **time**: eguna baino laburragoak diren une edo tarte eta erlojuko orduentzat: *goizean, 12:32, bostetatik, zortzietara* eta abar.
- **duration**: denbora tarteentzat. Egutegiko informaziorik badago denbora-adierazpenean, TIME edo DATE moduko adierazpentzat hartu behar dira. Adibidez, 1986 ezin da DURATION bezala markatu testuinguruko ekintzak urte osoan iraun duela ulertzen bada ere. Iraupenak, ondorioz, honelakoak izango dira: *bi hilabetez, 48 orduan, lau aste* eta abar.
- **Set**: errepikatzen diren denbora-adierazpentzat. *Astean birritan, astero, 3 egunetik behin* eta abar.

2. irudian denbora-adierazpenen atributuak osatzeko etiketatze-tresnak duen leihoa ikus daiteke. Atributu batzuk, jarraian ikusiko denez, automatikoki esleituko dira, baina besteak leiho horren bidez eskuz osatu behar dira.

value (balio) atributuak denbora-egituraren esanahia adierazten du egiturari esleitutako mota atributuaren arabera eta ISO 8601 formatuan oinarrituz. Ondoko zerrendan azalduko da mota bakoitzeko denbora-adierazpenei balioa nola esleitu.

- Dataren (DATE) formatua YYYY-MM-[WW]-DD da (hau da, urtea, hilabetea, astea (aukerazkoa) eta eguna). Eguna bezain luzeak edo handiagoak diren “uneak” kronologian kokatzeko erabiliko da.

Daturen bat osatu ezin izanez gero, erabili “X” hutsunea betetzeko. Definitzen den unitatea (*hilabetea*, esaterako) baino txikiagoak diren hutsuneak ez dira “X”z bete behar; adibidez *Ekainean* “XXXX-06” izango da eta ez “XXXX-06-XX”.

(103) Orokorrean inauteriak `<TIMEX3 type='DATE' value='XXXX-02'>` otsailean `<\TIMEX3>` izaten dira.

Asteak adierazteko hilabeteen kokalekua erabili, aurretik “W” astearen markatzailea adieraziz.

(104) `<TIMEX3 type='DATE' value='2014-W45'>` Datorren astean `<\TIMEX3>` ez da Ligako partidurik izango.

“Asteburua” adierazteko egunaren lekuan “WE” koka ezazu.

ADI! Hilabetea ezagututa ere, ez jarri, asteak berak kokapen zehatza adierazten baitu, eta asteek eta hilabeteek leku bera hartzen dutenez, prozesamendu akatsak izango lirakeelako.

(105) `<TIMEX3 type='DATE' value='XXXX-XX-WE'>` Asteburuetan `<\TIMEX3>` mendira joaten gara.

Markable Attributes

Current Extent: gaur egun

Markable type: TIMEX3

type: DATE

value: PRESENT_REF

functionInDocument: NONE

anchorTimeID: [drag markable here to link](#)

beginPoint: [drag markable here to link](#)

endPoint: [drag markable here to link](#)

*unsaved values

Save

2. irudia: TIMEX3 atributuak txertatzeko leihaxka

Asteko egunak YYYY-Www-D formatuaren bidez adieraziko dira, eguna 1etik 7rako balio batekin adieraziz, 1a astelehena eta 7a igandea izanik. Formatu hori adierazpen generikoak adierazteko baino ez da erabiliko.

(106) <TIMEX3 type='DATE' value='XXXX-WXX-1'> Astelehenetan <\TIMEX3> pilates eskolak izaten dira polikiroldegian.

Adierazpen baten balioa hilabete edo asteetan oinarrituta adieraz badaiteke, hilabeteetan oinarritutakoa aukeratu.

(107) <TIMEX3 type='DATE' value='2014-03-14'> Datorren ostiralean <\TIMEX3> Willis Drummondek joko du auzoko jaietan.

Hilabeteak YYYY-MM eta urteak YYYY formatuen bidez adieraziko dira.

(108) <TIMEX3 type='DATE' value='2016'> 2016 <\TIMEX3>

(109) <TIMEX3 type='DATE' value='2016-04'> 2016ko apirilean <\TIMEX3>

Hamarkadak YYYY, mendeak YY eta milurtekoak Y formatuaren bidez adieraziko dira.

(110) <TIMEX3 type='DATE' value='196'> 60ko hamarkada <\TIMEX3>

(111) <TIMEX3 type='DATE' value='19'> XX. mendea <\TIMEX3>

(112) <TIMEX3 type='DATE' value='2'> hirugarren milurtekoa <\TIMEX3>

Urtaroak token bidez adierazi behar dira: SP (udaberria), SU (uda), AU (udazkena) eta WI (negua). Eta hilabeteen tokian kokatuko dira.

(113) 68ko uda - value="1968-SU" <TIMEX3 type='DATE' value='1968-SU'> 68ko uda <\TIMEX3>

Hiruhilekoak (Q1, Q2, Q3 eta Q4) ere adierazi ahal dira, hala nola, seiuhilekoak (H1 eta H2), baita urte fiskalak ere (FY).

(114) <TIMEX3 type='DATE' value='2013-H2'> Aurtengo bigarren seiuhilekoan <\TIMEX3> Eroskik % leko irabaziak izan ditu.

(115) <TIMEX3 type='DATE' value='FY2012'> 2012 urte fiskala <\TIMEX3>

ADI! (114) adibidean "Aurtengo bigarren seiuhilekoa" agertzen da. Kasu horretan garrantzitsuagoa da unea iraupena baino eta denbora-adierazpen bakar moduan etiketatuko da. "Aurtengo azken sei hilabeteetan" agertuko balitz, ordea, baliokideak badira ere, bi denbora-adierazpen bereiztuko dira: "aurtengo" eta "azken sei hilabeteetan". Bigarren motako horiek (102) adibidean deskribatu dira.

Adierazpen zehaztugabeei ere balio bat eman ahal izango zaie 'PAST_REF', 'PRESENT_REF' eta 'FUTURE_REF' balio zehaztugabeen bidez.

(116) <TIMEX3 type='DATE' value='PRESENT_REF'> Orain <\TIMEX3>

(117) <TIMEX3 type='DATE' value='PAST_REF'> Lehen <\TIMEX3>

(118) <TIMEX3 type='DATE' value='FUTURE_REF'> Gero <\TIMEX3>

Datak iraupen baten hasiera eta amaiera puntuak izan daitezke, adibidez *2012tik 2015era*, eta hala, bi une independente modura, markatu behar izango dira. Gauza bera egingo da ere TIME motako adierazpenekin.

ADI! Kristo aurreko urteei zein balio eman definitzeke dago oraindik.

- Egunak baino laburragoak diren denborak (TIME) adierazteko orokorrean erabiliko den formatua THH:MM:SS da. Orduak adierazteko THH eta minutuak adierazteko THH:MM ere erabil badaitezke ere. Ikus daitekeenez, orduak "T" markatzailea dute datetatik bereizi ahal izateko.

Esku artean dugun testuak data bati erreferentzia zuzena egiten badio, orduaren balioan aipatu behar da:

(119) <TIME3 type='TIME' value='2014-02-13T10:19:43'> Gaur goizeko 10:19:43 <\TIME3> heldu da zunda Martera.

(119) adibidean ikus daiteke nola data eta eguneko une zehatza agertzen diren testuan. Kasu horretan data eta ordua “T” adierazlearen bidez bereiziko dira. Balioaren amaieran “Z” bat ezarriz ordua Greenwicheko Meridianoaren denboraren (GMT) edo Denbora Koordinatu Unibertsalaren (UTC) arabera adieraziko da:

(120) 11:13 GMT - T11:13Z <TIME3 type='TIME' value='T11:13Z'> 11.13 GMT <\TIME3>

Beste ordu-eremuak adierazi nahi izanez gero, ordu diferentzia ordu osoen bidez adieraziko da:

(121) <TIME3 type='TIME' value='T11:13-05'> 11:13 ET (Eastern Time) <\TIME3>

ADI! (121) adibideko “-k balio negatiboa adierazi nahi du. Greenwichetik eskuinera dauden ordu eremuek “+” hartuko dute. Hau, dena den, ez dago TimeMLn definituta eta hedapena ere aztertu behar izango litzateke.

Egunaren atalak adierazteko honako tokenak erabiliko dira: MO goizarentzat, AF eguerdiarentzat (12:00–16:00), EV arratsaldearentzat, NI gauarentzat eta GA goizaldearentzat. Eguna adierazteko DT (*daytime*) erabiliko da.

(122) <TIME3 type='TIME' value='2014-02-18TEV'> Gaur arratsaldean <\TIME3> kafea hartzeko geratu dira.

Gogoratu denbora zehatza agertzen bada, ez direla aurreko tokenak erabili behar.

(123) <TIME3 type='TIME' value='T05:00'> Goizaldeko bostetan <\TIME3> esnatzen da lanera joateko.

ADI! “Gaur goizean”, “bihar arratsaldean” eta horrelakoak denbora-adierazpen bakarra dira. Ikus daitekeenez, bigarren elementua “eguna baino laburragoa” da eta, ondorioz, egitura bakarra izanik, adierazpenari “TIME” balioa emango zaio mota atributuan.

- Iraupenak (DURATION) denbora tarte bat adierazteko erabiliko dira. PnYnMnDTnHnMnS edo PnW formatuen bidez adieraziko dira, “P” iraupenaren marka izanik eta Y urtea, M hilabetea, D eguna, T ordu marka, H ordua, M minutua, S segundoa eta W astea. Beste denbora-unitate batzuk ME (milurtekoa), CE (mendea) eta DE (hamarkada) izango dira.

(124) <TIME3 type='DURATION' value='P3M15DT14H'> 3 hilabete, 15 egun eta 14 ordu <\TIME3>

(125) <TIME3 type='DURATION' value='PT25M'> 25 minutu <\TIME3>

Iraupen bat frakzio dezimal bidez adieraz badaiteke, hala egingo da:

(126) <TIME3 type='DURATION' value='P0.5Y'> Urte erdia <\TIME3>

(127) <TIMEX3 type='DURATION' value='PT0.25H'> Ordu laurden <\TIMEX3>

Baina “sei hilabete” edo “15 minutu” etiketatu behar izanez gero, “P6M” eta “PT15M” erabiliko dira.

Egunaren atalak (MO, AF, EV, NI, GA eta DT), asteburuak (WE), urtaroak (SP, SU, AU, WI), hiruhilekoak (Q), seihilekoa (H) eta urte fiskalak (FY) adierazteko tokenak erabiliko dira horien adinako luzerako iraupenak adierazteko.

Iraupena zehaztu ezin bada, “X” balioa emango zaio:

(128) hilabete batzuetan zehar - value=“PXM” <TIMEX3 type='DURATION' value='P3M15DT14H'>
Hilabete batzuetan zehar <\TIMEX3>

(129) Gaur goizean - value“2014-03-12TMO”

Epe eta *tarte* moduko hitzak agertzen badira, etiketaren barruan sartuko dira (130). Ekintza bat + “epe” agertuz gero, ekintza ebento moduan markatuko da (horiek etiketatzeo gidale-
roak definitzen direnean) eta “epe” hitza <TIMEX3> moduan (131).

(130) Bahitzaileek 48 orduko epea eskaini zioten gobernuari.

(131) Helegite epea maiatzaren 23an zabalduko da.

Gerta daiteke iraupenaren unitatea definituta ez agertzea (132) eta (133) adibideetan bezala:

(132) Ikasketak amaitu eta gero *bolada bat* atzerrian igaro nahi dut.

(133) Maitek eta Andonik harremana eten zutenean, *tarte batez* elkarri hitz egin gabe egon ziren.

Kasu horietan iraupenari unitaterik logikoena esleituko zaio ahal izanez gero. (132) adibideko “bolada bat” hilabetetan neurtuko da ziurrenik eta (133) adibideko “tarte batez” asteak izan daitezke, adibidez.

- SET motak gertaera edo denboren iterazioa gertatzen den denbora tartea adierazten du. Alde batetik, balioa adierazi behar izango dugu, denbora-unitatea adierazteko eta beste alde batetik, maiztasuna edo kantitatea adieraziko ditugu, hau da, zenbat aldiz errepikatzen den definitutako unitatea.

(134) <TIMEX3 type='SET' value='XXXX-WXX-1' quant='EVERY'> Astelehenero <\TIMEX3>

(135) <TIMEX3 type='SET' value='XXXX-02' quant='EVERY'> Otsailero <\TIMEX3>

(136) <TIMEX3 type='SET' value='P1W' freq='2'> Astean birritan <\TIMEX3>

(137) <TIMEX3 type='SET' value='P3D' freq='1'> Hiru egunean behin <\TIMEX3>

SET motako denbora-adierazpenetan “quant” eta “freq” atributuak erabiliko ditugu. Lehenak gertaera zenbat aldiz gertatzen den adieraziko du eta “EVERY” (guztiak), “EACH” (bakoitz) edo “SOME” (batzuk) bezalako balioak hartuko ditu. Bigarrenak, bestalde, gertaeraren maiztasuna adieraziko du; hau da, definitutako denbora tartean zenbat aldiz gertatu den.

(134) eta (135) adibideetan ikustenenez, balioa puntu moduan adierazita dago egutegiko informazioa eskuratu ahal izateko. Arau modura, kontuan izan, egutegiko data agertuz gero, puntu moduan emango dela errepikapenaren balioa eta, ez badago, iraupen modura.

mod Denbora-modifikatzaileen informazioa batzeko balio du, hala nola, “hasieran”(“START”), “gutxi gorabehera”(“APPROX”), “inguru”(“APPROX”) edo “baino gehiago”(“MORE_THAN”). Modifikatzaileok ere denbora-adierazpenaren esparruaren barruan egongo dira.

- (138) `<TIMEX3 type='DATE' value='2014-03' mod='START'>` Hilabete hasieran `<\TIMEX3>` langabezia % 25ekoa zen.
- (139) `<TIMEX3 type='DURATION' value='P1DE' mod='MORE_THAN'>` Hamarkada bat baino gehiago `<\TIMEX3>` behar izan zuten eraikin hura altxatzeko.
- (140) Filma `<TIMEX3 type='TIME' value='T22H' mod='APPROX'>` hamarrak inguruan `<\TIMEX3>` amaitu zen.

Atributu honen balioak hauek izango dira: 'BEFORE (lehenago), 'AFTER (geroago), 'ON_OR_BEFORE (une horretan edo lehenago), 'ON_OR_AFTER (une horretan edo geroago), 'LESS_THAN (baino gutxiago), 'MORE_THAN (baino gehiago), 'EQUAL_OR_LESS (berdin edo gutxiago), 'EQUAL_OR_MORE (berdin edo gehiago), 'START (hasiera), 'MID (erdialdea), 'END (amaiera) eta 'APPROX (gutxi gorabehera).

temporalFunction Adierazpen baten balioa denbora-funtzio baten bidez kalkulatu behar den esaten duen atributu binarioa da (TRUE, FALSE). Balioa positiboa izango da baldin eta ezinezkoa bada denbora-adierazpenaren balioa uneko informazioaren bidez guztiz zehaztea. Espezifikatu gabeko `<TIMEX3>`-ak kontuan hartzeko baino ez du balio. TRUE balioa emango zaio atributu honi denbora-adierazpena espezifikatu gabea dela esateko.

TemporalFunction balioak postprozesamenduan esleituko dira; beraz, anotatzaileak ez du atributu horri balioa emateaz arduratu behar.

- (141) `<TIMEX3 type='DURATION' value='PXM' temporalFunction='TRUE'>` Azken hilabeteetan `<\TIMEX3>`
- (142) `<TIMEX3 type='DATE' value='2014-03' temporalFunction='FALSE'>` 2014ko martxoan `<\TIMEX3>`

(141) adibidean, testutik ezin da balio oso bat inferitu eta, ondorioz, denbora-funtzioa positiboa izango da. (142) adibidean, ordea, balio osoa dugu eta horregatik `temporalFunction`-en balioa “FALSE” da.

Hori automatikoki esleituriko etiketa da eta prozesamendu automatikoan baino ez du garrantzirik izango.

anchorTimeID Denbora-adierazpen batzuek beste batzuen erreferentzia behar dute euren balioa zehazteko. Erreferente diren adierazpenen identifikatzailea (ID) adierazteko erabiltzen da hau.

- (143) `<TIMEX3 type='DATE' value='2014-02-27' anchorTimeID='t00'>` Atzo `<\TIMEX3>` euria gogotik ari zuen.

(143) adibidean “t00” identifikazioa duen denbora-adierazpen bat (adibidean agerten ez dena, baina existitzen dena) da adibideko adierazpenaren balioaren erreferentzia emango duena. Adierazpen guztiek ez dute erreferentzia punturik (*anchor timerik*) izango, ez dutelako behar (144) edo defini ezin daitekeelako (145).

- (144) Bigarren Mundu Gerra [1939ko irailaren 1ean] hasi zen.

(145) Gazteen langabezia arazo larria da [gaur egun].

ADI! DCTak (Document Creation Time) ez du INOIZ `anchorTimeID`rik hartuko.

functionInDocument Atributu honek denbora-adierazpen bat beste batentzat erreferentzia puntua den adierazten du. Gure kasuan bi balio baino ez dizkiogu emango atributu horri:

- `CREATION_TIME`: testuaren sorrera unea.
- `NONE`: testuko beste guztientzat, inolako erreferentziarik adierazten ez duena.

Gerta liteke beste meta informaziorik aurkitzea, hala nola publikazio ordua edo azken eguneratze ordua. Balego, horientzat balioa sortu behar izango litzateke, baina gure kasuan sorrera unea baino ez dugu kontuan izango. Sorrera une hori testutik bereizita agertuko da erraz identifikatzeko.

valueFromFunction Atributu hau ez da kontuan hartu behar eskuz etiketatzean, automatikoki esleitzen baita.

beginPoint eta endPoint Atributu `DURATION` motako adierazpenentzat dira, adierazitako iraupenaren hasiera eta amaiera unek markatzeko. Ezin izango dira beti bete, informazioa ez baita testuan agertuko. Batzuetan, ordea, testuan agertzen den iraupen bati hasiera eta amaiera unek esleitu ahal izango zaizkio, horretarako etiketa huts batzuk sortuko dira (ikus. 3.1.5).

(146) Izena emateko `<TIMEX3 type='DURATION' value='PXD' beginPoint='2014-04-30'>`
epea `<\TIMEX3>` bihar zabalduko da.

(147) `<TIMEX3 type='DURATION' value='P3H' beginPoint='2014-04-30T13:20' endPoint='2014-04-30'>`
Hiru ordu `<\TIMEX3>` daramat iturginaren zain.

(146) adibidean suposatzen dugu erreferentzia unea 2014-04-29 dela eta horregatik gara gai “epea”ren hasiera unea 2014-04-30 dela ondorioztatzeko.

ADI! `TIME`, `DAT` eta `SET` motako `<TIMEX3>`ek ezin dituzte `beginPoint` eta `endPoint` hartu.

freq eta quant `SET`, errepikapenen kasuan, `quant`-en bidez testuan esplizituki agertzen den kuantifikatzailea adierazten da. `freq`-en bidez, zenbakizko balio bat eta denbora-granularitate bat baliatuz, denbora-egitura baten gauzatze frekuentzia adierazten da:

(148) `<TIMEX3 type='SET' value='P1W' freq='2X'>` astean bi aldiz `<\TIMEX3>`

Goiko (148) adibidean maiztasunaren balioa “2X” da X balio zehaztugabea izaki.

(149) `<TIMEX3 type='SET' value='XXXX-WXX' quant='EVERY'>` astero `<\TIMEX3>`

(150) `<TIMEX3 type='SET' value='P1W' quant='EACH' freq='3D'>` asteko hiru egu-
nean `<\TIMEX3>` (3 days each week)

comment Anotatzaileak etiketari buruzko komentarioak gehi ditzan. Atributu hau ez da `TimeMLn` agertzen, baina anotatzaileek iruzkinak testuan bertan egiteko oso baliagarria da.

3.1.3 Kulturak zehaztutako adierazpenak

Hizkuntza bakoitza kultura baten isla da eta kultura bakoitzak egun berezi eta gertaera berezi batzuk dituzenez, horien gaineko aipamenak etiketatu ahal izateko ezagutza espezifiko behar da. Horrelakoak dira jaien izenak (*Gabonak*, *Aste Santua*), adibidez. *Euskararen egunak*, adibidez, *egun* abiarazle lexikoa du, baina ez da beti hala gertatzen: *Maritxu Kajoi*. Jai izen horiek etiketa daitezke beti ere euren balioa testutik inferi badaiteke. Ezin izanez gero, XXXX-XX-XX balio zehaztugabea esleituko zaie. Egutegi gregoriarraz besteko denbora-adierazpenak (*eskola-urtea* edo *txotx-denboraldia*) bakarrik markatuko dira baldin eta abiarazle lexikoa badute.

(151) Jende pila batu zen Plaza Berrian <TIMEX3 type='DATE' value='2014-04-20'> pazko-igandean <\TIMEX3>.

(152) Beherapen handiak espero dira <TIMEX3 type='DATE' value='2014'> aurtengo Gabonetan <\TIMEX3> salmentak onak izan baitira.

(152) adibidean ikus daitekeenez, ezin defini daitezke guztiz *aurtengo gabonak*, beraz, testutik atera daitekeen beste informazio baliatuko dugu.

3.1.4 Dokumentuaren sorrera-unearen etiketatzea

Dokumentuaren sorrera-unea (DCT) esplizituki agertzen bada dokumentuan, <TIMEX3> bezala etiketatu behar da eta `functionInDocument` atributuan "CREATION_TIME" balioa hartuko du. DC-Taren `tid="t00"` izango da. Sorrera unea bera izango da besteentzat erreferentzia, beraz, ez du INOIZ erreferentzia punturik (`anchorTimeID`) hartuko.

3.1.5 <TIMEX3> etiketa hutsak

TimeMLk testua kontsumitzen ez duten denbora-etiketak sortzeko aukera ere ematen du, denbora-adierazpen bat testua kontsumitzen duen adierazpen batetik inferitu badaiteke. Horrelakoak askotan gertatzen dira iraupenekin, adibidez.

Ainguratze-erreferentzia implizituak dituzten iraupen ainguratuak: aingura testuinguruaren bidez interpretatu ahal izanez gero, etiketa huts bat sortuko da aingura data implizituari erreferentzia eginez.

(153) <TIMEX3 type='DURATION' value='P2M' beginPoint='t1'> duela bi hilabete <\TIMEX3> (etiketa hutsa <TIMEX3 type='DURATION' value='2014-01' tid='1'> orain 2014-03an baikaude).

Ainguratze-data esplizitua duten iraupen ainguratuak: ainguratze-elementua testuan agertzen da eta beste etiketa huts bat sortuko da egitura osoa adierazteko.

(154) The earthquake happened [two years ago] [today] (bi egiturak lotuko dituen etiketa hutsa sortu). *Lurrikara [orain dela 2 urte] gertatu zen*

Tartea adierazten duten adierazpenak: hauek iraupen implizitu baten hasiera eta amaiera adierazten duten bi denbora-adierazpen (DATE edo TIME motakoak) hartzen dituzte.

(155) [1998tik] [2003ra] inmigrazioa % 10 hasi zen. (etiketa hutsa: <TIMEX3 type='DURATION' value='P5Y'> 5 urte <\TIMEX3>)

Markodun iraupenak: marko-erlazioek puntu bat eta iraupen adierazpen bat dituen denbora-tarte bat adierazten dute. Hau da, iraupena eta iraupenaren mugetako bat agertzen dira. Kasu honetan iraupenaren hasiera eta amaiera adieraziko dituzten bi etiketa huts sortuko dira.

(156) [Urteko] [lehen sei hilabeteak]

ADI! (114) adibideko “aurtengo lehen seihilekoa” ez da kokatutako iraupena, une bat da; denbora-adierazpen bakarra une bati erreferentzia egiten diona.

4 SEINALEAK

Seinaleak denbora-egiturak, izan denbora-adierazpenak edota gertaerak, elkarren artean lotzeko balio duten esapideak dira (157–158). Seinaleek jarraian agertzen den <SIGNAL> etiketa hartuko dute.

(157) Bigarren singlea “Badok.info” atarian izango da entzungai urriaren 6tik *aurrera*.

(158) Haria bira amaitu *orduko* kantu berriak lantzen hasi ginen.

4.1 <SIGNAL>

<SIGNAL> etiketa ere ISO-TimeMLtik (2008) hartua da eta denbora-erlazio esplizitua adierazten duten testuko elementuak etiketatzeko erabiltzen da. Erlaziook bi gertaera-aipamenen, bi denbora-adierazpen edota gertaera-aipamen eta denbora-adierazpen baten artekoak izango dira eta ondoko elementuen bidez gauza daitezke euskaraz:

- denborazko postposizio askeak: *arte*, *bitarte*, *ondotik*, *aurre*.
- denborazko lokailuak: *orduko*.
- denborazko eta adberbio lokuzioak: *ondoren*, *eta gero*, *baino lehenago*.²
- karaktere bereziak: - eta /

Gogoratu behar da seinaleon hedadura funtzio-hitz edo karaktereari mugatuko zaiola. Postposizio egituretan denbora-adierazpena badago, hori <TIMEX3> etiketaren bidez markatuko da eta postposizioa <SIGNAL>-en bidez.

Karaktere berezien kasuan, kontuan hartzekoa da orokorrean egitura baten barruan agertzen direla (2014/05/17) eta horrelakoetan data osorik hartuko da; ez da barra bakoitza seinaleztat hartuko.

(159) Garazik astelehenera *arte* du erantzuna emateko.

(160) Zazpiak *eta gero* ez da banaketa zerbitzurik eskaintzen.

(161) Bostak *orduko* dokumentuak bukatuta egon behar du.

² *oraindik*, *jadantik*, *jada* etab. modukoak ez dira inolaz etiketatuko.

Seinale izango dira, beraz, postposizio askeak (*arte* (159)), adberbio lokuzioak (*baino lehen, eta gero*), lokailuak (*orduko* (161)) eta karaktereak (/,-).

<SIGNAL> seinaleek ID (sid) identifikazio etiketa automatikoa baino ez dute hartuko.

Astelehena [baino] bi egun [lehenago] modukoetan “lehenago” (beranduago, goizago e.a.) etiketatuko da seinaleztat. “Eta gero” lokailua, ordea, bere osotasunean etiketa bakarrarekin markatuko da.

ADI! mod atributua azaltzean “astelehena baino lehen” modukoak nola etiketatu azaldu da. Horrelakoetan “baino lehen” denbora-adierazpenaren parte izango da eta ez da azaldu berri ditugun seinaleekin konfunditu behar.

Gogoratu *orain dela hiru egun* modukoak (*duela, badira*) denbora-adierazpentzat hartuko direla:

(162) Etxerako lan hori *duela bi egun* bidali zigun irakasleak.

4.2 Aditzak eta seinaleak

Euskaraz oso arruntak dira denborazko mendeko perpausak. Horietan aditzari denborazko menderagailu bat, *-(e)nean* edo *-(e)la(rik)*, gehitzen zaio eta aditz nagusiak adierazten duen ekintza edo egoera eta mendeko perpausako aditzaren artean denborazko erlazio bat sortzen da.

(163) Menditik *zatozenean* eros ezazu mesedez ogia.

(164) Bahituak aurkitu *zituztenean*, oso egoera txarrean zeudela konturatu ziren.

(165) Bazkaldu *eta gero*, harrikoa egiten dugu beti.

(163–165) adibideetan nabarmendu diren hitzek denbora erlazioa adierazten dute. (163) adibidean aditz nagusia eta erlazio-atzizkia token berean agertzen dira; (164) adibidean denbora-atzizkia aditz laguntzailearen baitan agertzen da eta (165) adibidean seinalea aditzetik guztiz bereiztuta agertzen da. Lehenengo kasuan token bakarrak bi etiketa hartu behar izango lituzke, baina, hori ezinezkoa denez, horrelako kasuetan gertaera bakarrik etiketatzea erabaki da, gerora denbora-erlazioak egin ahal izango baitira, erlazio hori adierazten duen elementua markatuta ez badago ere.

Ondorioz, honako erabakia hartu da denborazko perpausak etiketatzeko: Erlazio-atzizkia aditz nagusi edo laguntzaileari loturik agertuz gero, EZ da etiketatuko. (165) adibidean bezala, ordea, guztiz bananduta agertzen bada, BAI.

Hala, honela etiketatu behar izango lirateke aurreko (163–165) adibideak:

(166) Menditik <EVENT> zatozenean <\EVENT> eros ezazu mesedez ogia.

(167) Bahituak <EVENT> aurkitu<\EVENT> zituztenean, oso egoera txarrean zeudela konturatu ziren.

(168) <EVENT> Bazkaldu <\EVENT> <SIGNAL>eta gero<\SIGNAL>, harrikoa egiten dugu beti.

5 ERLAZIOAK

TimeMLk gertaeren arteko erlazioak sortzeko aukera ere eskaintzen du. TLINK denbora-erlazioak adierazteko erabiliko da, SLINK mendekotasuna adierazteko erabiliko da esaldi bereko bi aditzen artean eta ALINK aspektuzko predikatua eta horren mendeko gertaera lotzeko erabiliko da.

5.1 TLINK

Arestian esan bezala, TLINKak denbora-erlazioak adierazteko baliatu dira. Erlaziook gertaeren (<EVENT>) edota denbora-adierazpenen (<TIMEX3>) artean gauzatzen direnak dira.

TimeMLk 13 denbora-erlazio mota bereizten ditu eta motok `relType` atributuaren bidez adieraziko dira. Atributu hori izango da garrantzitsuena eta etiketatzailerak eskuz bete behar izango duten bakarra.

Jarraian aurkeztuko dira TLINK denborazko loturak hartu ahal izango dituzten atributuak.

5.1.1 TLINKen BNFa

```
attributes ::= [lid] [origin] (eventInstanceID | timeID) [signalID] [syntax]
              (relatedToEventInstance | relatedToTime) relType

lid ::= ID
{lid ::= LinkID
LinkID ::= l<integer>}
origin ::= CDATA
eventInstanceID ::= IDREF
{eventInstanceID ::= EventInstanceID}
timeID ::= IDREF
{timeID ::= TimeID}
signalID ::= IDREF
{signalID ::= SignalID}
relatedToEventInstance ::= IDREF
{relatedToEventInstance ::= EventInstanceID}
relatedToTime ::= IDREF
{relatedToTime ::= TimeID}
relType ::= 'BEFORE' | 'AFTER' | 'IBEFORE' | 'IAFTER' | 'INCLUDES' |
            'IS_INCLUDED' | 'MEASURE' | 'SIMULTANEOUS' | 'BEGINS' | 'BEGUN_BY' |
            'ENDS' | 'ENDED_BY' | 'IDENTITY'
syntax ::= CDATA
```

5.1.2 relType atributua

`relType` atributua izango da eskuz bete behar den bakarra, beste guztiak saguarekin lotu behar diren tokenak klikatzean edo automatikoki beteko baitira. 13 balio posible izango ditu: 5 bikote konplementario eta 3 balio independente. Etiketatzean, gertaera eta denbora-adierazpen baten arteko erlazioetan gertaera izango da iturria eta denbora-adierazpena xedea (`MEASURE` motako erlazioetan salbu). Bi gertaera edo bi denbora-adierazpen badaude, ordea, esaldi nagusikoa izango da iturri. Maila bereko bi esalditako gertaera edo denbora-adierazpenen kasuan, azkenik, lehenengo agertzen dena izango da iturria.

Jarraian datozen adibideetan nabarmendu nahi ditugun elementuak baino ez dira kontuan hartuko argitasunaren izenean. Denbora-erlazioak adierazten dituzten seinaleak kortxete bidez eta letra arruntez (ez kurtsibaz) adieraziko dira adibideotan. Erlazioaren bi elementuak jarraian agertzen direnean, hauek kurtsibaz emango dira, baita kortxetez bereizita ere, bien mugak ezagutarazteko.

- BEFORE: gertaera edo denbora bat beste baten aurretik gertatzen dela adierazteko.

- (169) Bilerara *sartu* [aurretik], Pavlo Klimkin Ukrainako Atzerri ministroak babesa *eskatu du*.
- AFTER: gertaera edo denbora bat beste baten ondoren gertatzen dela adierazteko.
- (170) Bilerara *sartu* [ondoren], Pavlo Klimkin Ukrainako Atzerri ministroak babesa *eskatu du*.
- IBEFORE (immediately before): gertaera edo denbora bat justu beste baten aurretik gertatzen dela adierazteko.
- (171) Ganadua zeraman kamioi bat kontrola *galdu* eta txirrindulari talde baten gainera *erorri da*.
- IAFTER (immediately after): gertaera edo denbora bat justu beste baten ondoren gertatzen dela adierazteko.
- (172) NA-1240 errepideko lehen kilometroan bi txirrindulari *hil dira*, kamioi batek azpian *harrapatuta*.
- INCLUDES: gertaera edo denbora batek beste bat bere barne hartzen duela adierazteko.
- (173) Horietako bat segizioa *pasatzen zen* [bitartean] “Gora Errepublika! *oihukatzeagatik* eraman du poliziak.
- IS_INCLUDED: gertaera edo denbora bat beste baten barne dagoela adierazteko.
- (174) Petro Poroxenko Ukrainako presidentea astebeteko su-etena [iragarri zuen] [joan den *astea*].
- BEGINS: denbora edo gertaera baten hasiera adierazten duten denbora edo gertaerak adierazteko.
- (175) Grebalariak *astelehenerik* daude lantegian itxita.
- BEGUN_BY: denbora-adierazpen baten unean edo beste gertaera bat gertatzen den unean hasten diren gertaerak adierazteko.
- (176) Grebalariak [astelehenerik] [daude] lantegian itxita.
- ENDS: denbora edo gertaera baten amaiera adierazten duten denbora edo gertaerak adierazteko.
- (177) Grebalariak *astelehenera* [arte] *egon dira* lantegian itxita.
- ENDED_BY: denbora-adierazpen baten unean edo beste gertaera bat gertatzen den unean amaitzen diren gertaerak adierazteko.

(178) Eusko Legebiltzarrak *urrira* [arte] *atzeratu* berri du Durangoko Azokari diru-laguntza handiagoa emateko erabakia.

- IDENTITY: gertaera berbera adierazten duten bi testu zati lotzeko.

(179) Fundazio horrek Madrilen baso-eskola *irekitzeko* asmoa du. 2015–2016 ikasturtean *abiatuko* litzateke Bosquescuola ekimena.

- MEASURE: DURATION motako <TIME3> bat gertaera batekin lotzen denean. Iraupenak ez dira beti TLINK baten parte izango; baina izanez gero, MEASURE motako erlazioa sortuko dute. ADI! MEASURE motako erlazioetan denbora-adierazpena izango da iturria eta gertaera xedea.

(180) Petro Poroxenko Ukrainako presidentea [astebeteko] [su-etena] iragarri zuen joan den astean

- SIMULTANEOUS: bi denbora edo gertaera une berean (edo ia une berean) gertatzen direla adierazteko.

(181) Leihoak *zabaltzean* haize bortitza *sartzen da* haietatik.

(182) Sorginak [gauerdian] [irteten dira] akelarreetarantz.

5.2 Beste atributuak

relType atributuaz aparte beste atributu batzuk ere hartzen ditu TLINKak. Atributu horiek, aitzitik, automatikoki esleituko dira, anotatzaileak aukeratzen dituen markableen arabera.

- eventInstanceID edo timeID: erlazioan parte hartzen duten elementuen IDak. Nahitaezkoa.
- signalID: denbora-erlazioa adierazten duen seinale bat testuan agertzen bada, horren identifikatzailea markatzeko. Aukerazkoa.
- relatedToEventInstance edo relatedToTime: erlazioan parte hartzen duen gertaerari edo denbora-adierazpenari lotutako entitatea adierazteko. Nahitaezkoa.
- origin: denbora-erlazioa eskuz edo automatikoki sortu den adierazteko. Aukerazkoa.

5.3 TLINKak egiteko protokoloa

<TLINK>ak sortzeko ondoko prozedura jarraituko da. Jarraituko den prozedura NewsReader proiektutik hartu da, hor erabiltzen dena oso egokia baita. Hala ere, etiketatzean aurrera joan ahala, aldaketak izan daitezke beharizanen arabera.

1. Gertaera eta DCTaren (dokumentuaren sorrera data) arteko loturak.

Aditz jokatuak baino ez dira lotuko DCTarekin. DCTa betiere esateko unetzat hartuko dela ere kontuan izan behar da.

2. Perpausetako gertaera nagusien artean:

Perpau bateko gertaera nagusiaren (aditz nagusiak adierazten duena) eta hurrengo perpau-seko aditz nagusiaren arteko erlazioa adierazteko.

3. Perpaus bereko aditz nagusiaren eta mendeko perpausetako aditzen arteko denbora-erlazioak:
 Beti ere mendeko perpausoko gertaera ADITZEZKOA bada, bestela ez da loturarik sortuko.
 Badira hiru kasu non ez den perpaus nagusiko gertaeraren eta mendeko perpausoko gertaeraren arteko loturarik egingo:

- GERTAERA NAGUSIA + HELBURUZKO PERPAUSEKO GERTAERA

(183) Produkzioa *geratu zuten* lantegian makinei errebisio sakona *egiteko*

- PREDIKATU OSAGARRIA + IRUDITU, IZAN edo antzekoak

(184) Gu pairatzen ari garenak *krisi* latza *dirudi*

- GERTAERA NAGUSIA + ERLATIBOZKO PERPAUSEKO GERTAERA

(185) Hitzarmena, gaur zortzi *sinatuko dena*, ez du inork *aurrikusi*

4. Perpaus bereko gertaera eta denbora adierazpenak:

- Aditzezkoak ez diren gertaerak:

Denbora-adierazpenak modifikatzen dituenen.

(186) 1992ko Olinpiar Jokoak

Denboran mugimendua adierazten duen aditz batekin agertzen direnean.

(187) *Batzarra* zuzendariak *arratsaldera* arte *atzeratzea* erabaki du.

- Aditzezko gertaerak:

– denbora-adierazpena perpaus nagusian badago, gertaera nagusiaren eta denbora-adierazpen horren arteko erlazioa sortu.

(188) *Bihar* alkateak dimisioa *emango* du.

– denbora-adierazpena mendeko perpausuan badago, mendeko perpausoko gertaeraren eta denbora-adierazpen horren arteko erlazioa sortu.

(189) Jaurlaritzak uste du *datorren urtean* langabeziak behera *egingo* duela.

– denbora-adierazpena perpaus nagusian badago eta mendeko perpausa denborazkoa bada, mendeko perpausoko aditza + denbora-adierazpena eta perpaus nagusiko aditza + denbora adierazpena loturak egingo dira.

(190) *Gaur* zuzendariak akziodunei bisita *egingo* die hitzarmena *sinatu* baino lehen.
 egingo-gaur (IS_INCLUDED)
 sinatu-gaur (IS_INCLUDED)
 egingo-sinatu (BEFORE)

5. Perpaus bereko bi denbora-adierazpen lotuta daudenean:

(191) *2014ko* [*lehen sei hilabeteetan*] langabeziak behera egin du

6 CAT Tresna

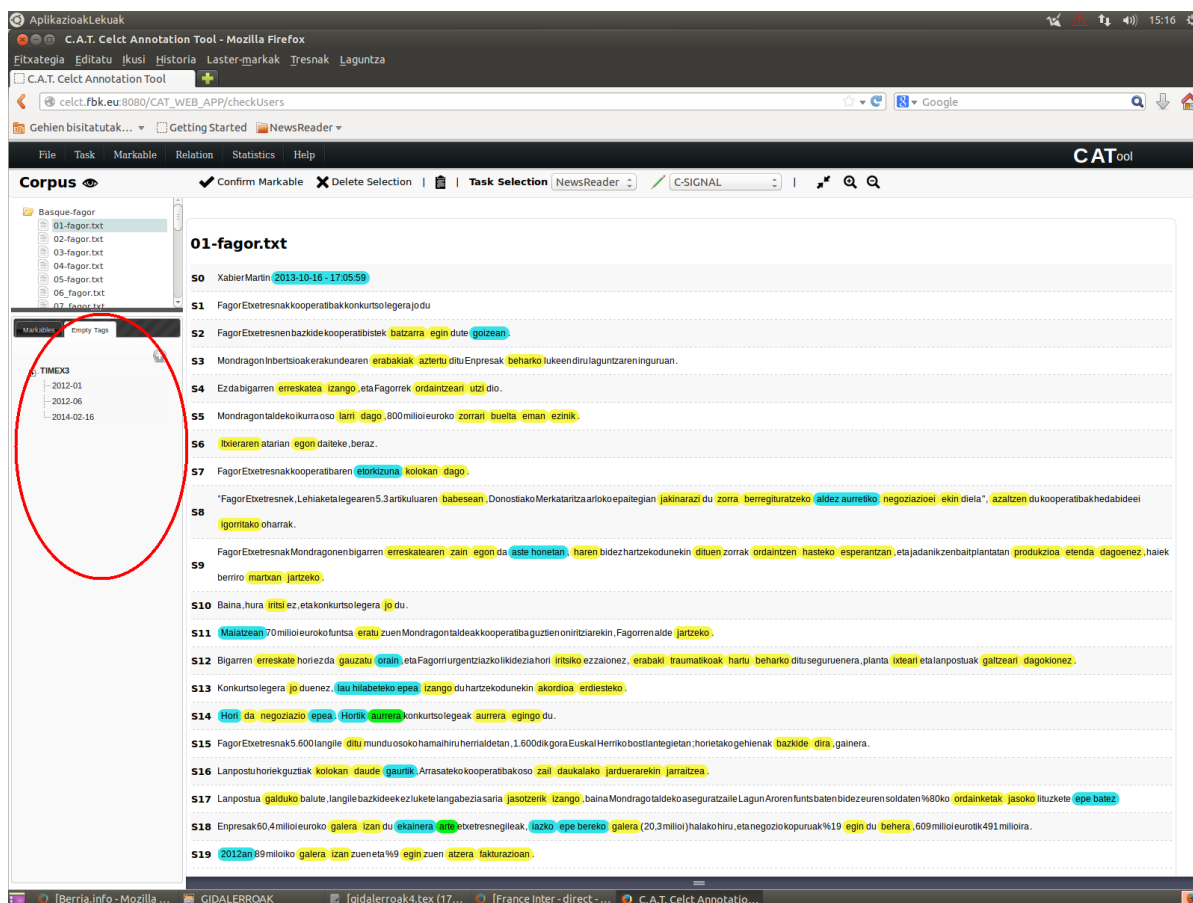
CAT tresna (Bartalesi et al., 2012) erabiliko da denbora informazioa etiketatzeko. Tresnak online bertsioa³ du eta hori erabiliko da bertsiorik berriena baitu. Jarraian azalduko dira tresna erabiltzeko urratsak. Ondoren azaldutakoak Linux eta Firefoxekin erabiltzekoak dira, gerta liteke beste sistema edo nabigatzaile batekin gauzaren bat aldatzea. Ondoko azalpenekin batera etiketatzetresnaren interfazearen irudia (3. irudia) ikus daiteke, pausoetako batzuk non egin azaltzeko.

1. Sartu webgunera eta erabiltzailea eta pasahitza idatzi.
2. Goiko barran anotazio ataza (gure kasuan: NewsReader) aukeratu eta ezkerreko zutabearen etiketatu nahi den artxiboa hautatu eta, klik bikoitza eginez, zabaldu.
3. Goiko barran erabili nahi dugun etiketa mota aukeratu.
4. Markatu nahi dugun elementua hautatu. Horretarako `shift` zapaldu eta tokenaren (edo tokenen gainean) klik egingo dugu. Berde jarriko da hautatutakoa. Orduan goiko barran “Confirm markable” zapaldu.
5. *Markable*aren (etiketatu daitekeneenaren) atributuak betetzeko, klik egin hautatutako elementuaren gainean eta aterako den leihaxkan datuak bete. GOGORATU: datuak sartu ondoren, “Save” botoia sakatu behar izango da, bestela datuak galduko dira. Datuak gorde direla, leihaxkan atributuak gorri izatetik beltz jartzera pasatuko direlako jakingo dugu.
6. Bada *markable* ez jarraituak egiteko aukera. Horretarako lehen markablea normal (`shift+click`) hautatuko da. Bigarrena, ordea, `ctrl+shift+click` eginez hautatuko da. Gero, “Confirm Markable” sakatzearekin nahiko izango da.
7. Etiketa hutsak sortzeko, ezkerreko zutabearen “Empty tags” zerrendan “+” sinboloa sakatu eta etiketa berria sortuko da (4. irudia). Horri, eskuineko botoia sakatuta, izena aldatuko zaio eta klik eginda atributuak bete ahal izango dira.
8. Erlazioak sortzeko (TLINK, ALINK, SLINK) azpiko menua zabaldu eta azpiko atal horretan dagoen laukitxoan erlazio mota hautatu eta alboan duen “+” ikurrean klikatu.
9. Erlazio hutsa agertzen denean (“+” ikurrean klikatu ondoren), erlazioaren iturria aukeratu nahi dugun markablearen gainean `ctrl+alt+click` eginez hautatu. Erlazioaren xedea `ctrl+shift+click` eginez aukeratu.
10. Erlazioaren ezaugarriak adierazteko, beheko menuan, sortutako erlazioaren gainean klik egin, atributuak bete eta gorde.

³http://celct.fbk.eu:8080/CAT_WEB_APP/checkUsers

The screenshot displays the CAT web application interface. At the top, there is a navigation bar with options like 'Fitxategia', 'Editatu', 'Ikusi', 'Historia', 'Laster-markak', 'Tresnak', and 'Laguntza'. Below this is a browser address bar showing 'celct.fbk.eu:8080/CAT_WEB_APP/checkUsers'. The main interface includes a 'Corpus' sidebar on the left with a file tree and a 'Markable' section containing 'SIGNAL' and 'TIMEX3' checkboxes. The central area shows a document titled '03-fagor.txt' with a list of numbered text segments (50-517). A 'Markable Attributes' dialog box is open over segment 56, showing fields for 'Current', 'Extent', 'Markable type', 'type', 'value', 'functionInDocument', 'anchor/TimeID', 'beginPoint', and 'endPoint'. Red numbers 1 through 7 are overlaid on the interface to highlight specific features: 1 points to the 'Task Selection' menu, 2 to the document title, 3 to the search bar, 4 to the 'Confirm Markable' button, 5 to the 'Save' button in the dialog, 6 to the 'Markable Attributes' dialog title, and 7 to the 'Markable' sidebar.

3. irudia: CAT tresnaren interfazeko atalak



4. irudia: Etiketa hutsak sortzeko esparrua

7 OHARRAK

Euskarazko denbora-adierazpenak etiketatzea ez da lan zaila, postposizioak (atzizkiak edo askeak) alde batera utzita egiten baita etiketatzea. Denbora-erlazioak, <TLINK>, osatuko direnean erreparatu behar izango zaie postposizioi, baina ez momentuz, adierazpenon balio kronologikoa eta mota baino ez baitzaizkigu ardura.

Ferro et al. (2003)-ren arabera *beti*, *batzuetan*, *askotan* modukoak ez dira markagarriak, guk etiketatuko ditugun erabaki behar dugu.

Erreferentziak

P. Altuna, P. Salaburu, P. Goenaga, M. P. Lasarte, L. Akesolo, M. Azkarate, P. Charriton, A. Eguskitza, J. Haritschelhar, A. King, J. M. Larrarte, J. A. Mujika, B. Oyharabal, and K. Rotaetxe. *Euskal Gramatika Lehen Urratsak II*. Euskaltzaindia, Bilbao, 1987.

- Valentina Bartalesi, Giovanni Moretti, and Rachele Sprugnoli. Cat: the celct annotation tool. In *Proceedings of LREC 2012*, Istanbul, Turkey, 2012.
- L. Ferro, L. Gerber, J. Hitzeman, E. Lima, and B. Sundheim. Ace time normalization (tern) 2004 english training data v1.0. <http://www ldc.upenn.edu/Catalog/catalogEntry.jsp?catalogId=LDC2005T07>, 2005. Accessed: 2013-07-03.
- Lisa Ferro, Laurie Gerber, Inderjeet Mani, Beth Sundheim, and George Wilson. Tides 2003 standard for the annotation of temporal expressions. Technical report, MITRE, McLean, USA, September 2003.
- ISO-TimeML working group. Language resource management semantic annotation framework (semaf) part 1: Time and events. International Standard ISO/CD 24617-1(E), ISO, 2008.
- E. Pociello. *Euskararen ezagutza-base lexikala: Euskal WordNet*. PhD thesis, Euskal Filologia Saila (Euskal Herriko Unibertsitatea), Leioa, 2008.
- J. Pustejovsky, J. Castao, R. Ingria, R. Saur, R. Gaizauskas, A. Setzer, and G. Katz. Timeml: Robust specification of event and temporal expressions in text. In *Proceedings of IWCS-5, Fifth International Workshop on Computational Semantics*, pages 1–11, Tilburg, Netherlands, 2003.
- TimeML Working Group. Timeml annotation guidelines version 1.3. manuscript. Technical report, Brandeis university, 2010.
- Sara Tonelli, Rachele Sprugnoli, and Manuela Speranza. Newsreader guidelines for annotation at document level. version 4.1. Technical report, Fondazione Bruno Kessler, 2014.