Adapting TimeML to Basque: Event Annotation

Begoña Altuna, María Jesús Aranzabe, and Arantza Díaz de Ilarraza

Universidad del País Vasco/Euskal Herriko Unibertsitatea,
Manuel Lardizabal pasealekua 1, 20018 Donostia, Spain
{begona.altuna,maxux.aranzabe,a.diazdeilarraza}@ehu.eus

Abstract. In this paper we present an event annotation effort following EusTimeML, a temporal mark-up language for Basque based on TimeML. For this, we first describe events and their main ontological and grammatical features. We base our analysis on Basque grammars and TimeML mark-up language classification of events. Annotation guidelines have been created to address the event information annotation for Basque and an annotation experiment has been conducted. A first round has served to evaluate the preliminary guidelines and decisions on event annotation have been taken according to annotations and inter-annotator agreement results. Then a guideline tuning period has followed. In the second round, we have created a manually-annotated gold standard corpus for event annotation in Basque. Event analysis and annotation experiment are part of a complete temporal information analysis and corpus creation work.

Keywords: EusTimeML, events, event extraction, event classification, temporal information, corpus creation

1 Introduction

Events—situations that happen or occur—can only be perceived through time, by means of the changes in the reality or the absence of those. Humans conceptualise time as points and intervals which are employed to locate those events in a chronology, through past, present and future, and events express an action or state located in a certain time or period. Apart from the events, there are time expressions that express those points and intervals in the temporal continuum and some structures, implicit or explicit, which convey temporal relation information such as "before", "after" or "simultaneous". These all help the speakers to situate the events in time, with a direct reference or one relative to another.

Event analysis is a major issue on natural language processing (NLP) as a part of temporal information analysis and processing, since they are the core of the discourse: the actions and situations we are talking about. Many evaluation challenges such as SemEval 2015 Task 4 [1] make us aware of its relevance. In order to take advantage of event information, experimentation (*e.g.* extraction, analysis, annotation) on events has to be done. Event information has to be made machine readable by means of a mark-up language. For example, TimeML [2] is a mark-up language to code events and temporal expressions, their features

and the relations among them based on XML. The information will then be represented in temporally annotated corpora like TimeBank [3] or WikiWars [4], which may be used to train machine-learning based tools such as TIPSem [5]. The information extracted from events and saved in corpora can be useful in many NLP tasks such as event forecasting [6] or timeline creation [7].

This paper has two main parts. First, an adaptation of TimeML temporal mark-up language [8] for event annotation in Basque. This has been done through an analysis of the event expressions in Basque of which we have analysed the main features. Second, the annotation of events following EusTimeML [9].

The article is structured as follows: in section 2 we give a definition of event and we classify events according to their lexical content to give a theoretical basis for the annotation of two most relevant event features. In section 3, we present the linguistically based decisions on event annotation for the experiment presented in section 4. To conclude, we sum up our main ideas and propose further research on the temporal information processing field in section 5.

2 Definition of Events

Event is a cover term for situations that happen, occur, hold, or take place and states and circumstances in which something obtains or holds true [10]. This definition already shows a difference between actions and states, but a more thorough classification can also be done.

Apart from their meaning, events convey different linguistic information. Some features such as the class are semantic, whereas features like the part of speech category are grammatical. These features can be normalised through attributes and a set of values.

2.1 Event Categorisation

Events can be classified according to their semantic features. For the annotation of Basque events, the event classification described in TimeML annotations have been followed:

- **Occurrence**: these are dynamic events that happen or occur, *e.g. salto egin* (to jump), *dantzatu* (to dance) or *ibili* (to walk).
- **State**: these are events describing circumstances in which something obtains or holds true and do not vary over time, *e.g. egon* (to be) or *geratu* (to remain).
- **Reporting**: reporting events describe the utterance, narration, description, etc. of an event, *e.g. esan* (to say) or *iragarri* (to announce).
- **Aspectual**: aspectual events indicate the beginning, continuity or end of an event: *e.g. hasi* (to begin), *jarraitu* (to continue) or *amaitu* (to end).
- **Perception**: these events describe the physical perception of another event, *e.g. ikusi* (to see), *entzun* (to hear) or *sumatu* (to perceive).

- **Intensional action**: are dynamic events that select for an event-denoting argument which is explicitly in the text, *e.g. saiatu* (to try), *agindu* (to order) or *aztertu* (to analyse).
- **Intensional state**: these are states that, as intensional actions do, select for an event-denoting argument which is explicitly in the text, *e.g. pentsatu* (to think), *gorrotatu* (to hate) or *prest egon* (to be ready).

This categorisation fulfils two major objectives. First it serves to determine whether an event may be taken as an argument for another event. Secondly, the difference between dynamic and stative events offers a preliminary view on how events happen in time, that is to say, whether they are punctual or last through a period of time.

The categorisation presented in this section is intended to cover all the different event types and give relevant semantic information of them. The different events described in this section can be represented by means of the expressions described in section 3.

3 Event Expressions in Basque: a Syntactic Perspective

Events can be expressed by more than one grammatical category. As in many other languages, mainly verbs (1), nouns (2), adjectives (3) and adverbs (4) (in bold) can express events in Basque:

(1) Hor ez **dira sartuko** Edesako langileak.
There no **AUX enter.FUT** Edesa.REL workers.ABS
'Edesa workers **will** not **enter** in there'.

(2) Fagor Etxetresnak enpresak **konkurtsora** joko du.
Fagor Etxetresnak company.ERG **tender.ALL** go.FUT AUX
'Fagor Etxetresnak company will go out to a **tender**'.

(3) Sartu den emakumeak **gaztea** dirudi.
Come.in AUX.REL woman.ERG **young** looks
'The woman who has come in looks **young**'.

(4) **Txaloka** egin dute ibilaldi guztia.
Clapping do AUX walk all.DET
'They have done all the walk **clapping**'.

The events in bold in (1–4) express a single event: a single action or state. For the annotation experiments, only the lexical head of each event expression will be marked, although all morpho-syntactic information contained in the phrase (auxiliaries, demonstratives, etc.) will be taken into account. For events expressed by nouns, adjectives and adverbs, those will be considered the lexical heads. For events expressed by verbs, instead, only the lexical head, *sartuko* (1), will get an event tag.

The linguistic analysis and decisions we have taken have been extracted from Basque grammars ([11], [12] and [13]), a classification of complex predicates [14] and the decisions taken for temporal annotation in other languages ([8], [15]). These expressions are described and examples for each are given below.

3.1 Events Expressed by Verbs

Events in Basque are mainly expressed by verbs. In the following sections the verb forms that may express an event are presented.

Synthetic Forms Synthetic forms are one-word units. The lexical root conveys the semantic information and a compound of morphemes add aspect, tense, person and mood information. Only a handful of verbs possess synthetic forms and their extension in the verbal paradigm is also restricted to some tenses. As a consequence, the majority of verbal events in Basque will be expressed by means of periphrastic forms.

(5) Zientzialariek urteak **daramatzate** fusiozko energia
Scientists.ERG years.ABS **have.been** fussion.INS power
merkearen bila.
cheap.GEN looking.for.
'Scientists **have** long **been** looking for cheap fusion power'.

Periphrastic Forms For periphrastic forms (also called analytical) we stick to the traditional Basque definition. These forms are formed by a lexical head which bears aspectual information and a mood, person and tempus carrying auxiliary. All verbs in Basque have periphrastic forms. The main formal variation happens in auxiliaries, which drastically reduces the mechanisms of morphological creation since lexical heads do not largely vary. This phenomenon makes the creation of periphrastic forms an easy language resource and leads to the reduction of synthetic form use.

(6) 8etan **atera dira** mendizaleak mendi tontorrerantza.
8.PL.LOC **leave AUX** hikers.ABS mountain summit.DIR
'Hikers **have left** at 8 towards the summit'.

Non-finite Forms Verbal expressions may also appear in Basque texts as non-finite forms (radical, participles and verbal nouns). These forms are used on their own as sentence heads in contexts such as fossilised expressions, exclamatory sentences and questions.

(7) Akordiorik **lortu** ezean, grebari eutsiko diote.
Agreement.PART **reach** NEG, strike.DAT continue AUX
'If no agreement **is reached**, (they) will continue on strike'.

3.2 Events Expressed by Nouns

Some nouns may express events. These can be verb nouns (8), common nouns (9) or proper nouns denoting a particular event (10).

(8) Derrigorrezkoa da Greziari zorraren zati bat
Compulsory is Greece.DAT debt.SG.GEN part.ABS a
barkatzea.
condone.ABS
'It is compulsory to **condone** a part of the Greek debt'.

(9) Lau **eskaera** nagusi egin dituzte.
Four **request** major.ABS do AUX
'(They) have done four major **requests**'.

(10) 2016an **Olinpiar Jokoak** Rio de Janeiron ospatuko dira.
2016.LOC **Olympic Games**.ABS Rio de Janeiro held.FUT AUX
'In 2016 **Olympic Games** will be held in Rio de Janeiro'.

We follow a test proposed in [8] to decide whether a noun refers to an event. A noun may express an event if it fits at least in two of the presented settings.

- NOUN lasted for several minutes/days/years/...
- NOUN was very fast/immediate/...
- NOUN took/takes/will take place *in temporal expression.*
- NOUN began/continued/ended *in temporal expression.*

3.3 Events Expressed by Adjectives

Adjectives express the qualities of the entity they refer to. Although they may appear in many contexts, we only consider events the adjectives acting as predicate adjectives.

(11) Zer egin behar da enpresa **bideragarria** egiteko?
What do have.to AUX business **profitable** make.FIN
'What has to be done to make the business **profitable**?'

3.4 Events Expressed by Adverbs

Adverbs will be considered event expressions when they accompany verbs to create a more complex event construction (section 3.6). These will mainly be adverbs of manner.

(12) Arrakasta **harro** egoteko modukoa da
Success.ABS **proud** be.FIN likely.ABS is
'Success is big enough to be **proud** of it'.

3.5 Events Expressed by Pronouns

Pronouns themselves do not express events, but may have a deictic value when they corefer with another event in the text.

(13) Bihar egingo da mozorro desfilea. **Horretarako**
Tomorrow do.FUT AUX.3.SG costume parade.ABS .PUN
erdigunea itxiko dute.
That.FIN centre.ABS close.FUT AUX
'Costume parade will be done tomorrow. For **that** the centre will be closed'.

3.6 Complex Structures

Complex predicates conform a non-homogeneous yet gradated group [14]. We will now present Basque complex predicate structures, based on Jedzrejko's [14] classification for Polish complex predicates. We have adapted this list to accommodate Basque predicate features as can be seen in the following lines:

- **Standard nominal predicates** are constructions with a basic auxiliary verb (*izan* (to be), *ukan* (to have)). The verbs in these constructions are semantically or referentially empty and the nominal or adjectival predicate carries all the predicative information.

  (14) Ordenagailua **geldoa da**.
  Computer.ABS **slow.ABS is**
  'The computer **is slow**'.

- **Modal predicates** are complex constructions formed by a lexical verb in its participle form and a conjugated modal verb (*nahi* (to want), *behar* (to have to, must) and *ahal* (to can)) or derived nouns expressing modality (15) (*nahi* (wish), *behar* (need, obligation), *ahal* (possibility)) and a participle.

  (15) Jende nagusiak noizbehinka **jesarri beharra du**.
  People elderly.ERG sometimes **sit.down need.ABS has**
  'Elderly people **has the need to sit down** sometimes'.

- **Aspectual predicates** are formed by an aspectual verb or noun and a verbal or nominal predicate. The aspectual expression in the construction marks the phase of its argument.

  (16) **Beherapenak amaitu dira**
  Sales.ABS end AUX
  '**Sales have ended**'.

- **Generic verb constructions**. Generic verbs are those which are used to give predicative properties to nouns. Therefore, generic verb constructions are formed by a noun that carries the lexical meaning of the event and a verb that provides the syntactic information.

(17) Ona da lagunek elkarri **musu ematea**.
Good is friends.ERG each.other.DAT **kiss give.ABS**
‘ It is good for friend to **kiss** each other.,

– **Metaphors**. In these constructions a fully predicating verb is used next to a noun phrase with a meaning other than its main meaning. The verb, apart from adding the grammatical and syntactic information, modifies the conceptual information of the noun it accompanies.

(18) Entzuleak **barrez lehertu ziren**.
Hearers.ABS **laugh.INS explode AUX**
‘Hearers **laughed** a lot’.

– **Idiomatic expressions** are formed by a noun and a verb which carries the grammatical information of the construction. Nevertheless, idiomatic constructions cannot be seen as a simple sum of the meanings of its parts and can only be understood as a single meaning unit.

(19) Adierazpenek **hautsak harrotu zituzten**.
Statements.ERG **dusts.ABS raise AUX**
‘The statements **caused a commotion**’.

Although these complex constructions express a single complex event, they may contain more than one event expressing form and all those forms will be annotated as single events according to the EusTimeML guidelines to show that complexity.

4 Experimentation

In order to prove the correctness and universality of the EusTimeML mark-up language and annotation guidelines, we have conducted a two-round annotation experiment on event identification and feature extraction. The first was a preliminary experiment to evaluate and discuss the guidelines [16]. There was a guideline tuning period following this first round in which the annotating team added or corrected annotation features. Once the new guidelines [9] were finished, a second annotation round was used to annotate a gold standard corpus of verbal event expression in Basque.

Both annotation efforts have been done using the CELCT Annotation Tool [17], which is easily customizable and offers a range of interesting features for textual annotation such as inter-annotator agreement metrics.

4.1 First Annotation Round

For this first experiment about 172 events[1] were annotated. The annotated documents are part of a 25 article corpus that contains news on the closure of a

[1] The amount of events varies among the annotations.

company extracted from a Basque newspaper. The events were annotated according to the EusTimeML guidelines [16], a set of guidelines for Basque temporal annotation based on the TimeML annotation scheme. Three annotators (A, B and C) took part in this annotation effort.

In this annotation round the agreement on event identification and extension were evaluated. The annotations of the three annotators were evaluated in pairs. Agreement levels ranged between 0.864 and 0.947 in weighted Dice's coefficient [18] depending on the annotator pair. The agreement level on the part of speech category, modality and whether events are aspectual were also evaluated.

We found that events expressed by a single token were unanimously annotated in most of the cases. We also discovered an unexpectedly high agreement on events expressed by nouns and adjectives. However, although agreement in general was high, some annotation features were troublesome; we list them below:

- Some tokens were incorrectly considered events; mainly verbs taking part in time expressions and discourse markers.
- Some events on complex structures were neglected.
- Event expressions derived from verbs, were not consistently given the same part of speech category.

In order to overcome these disagreement issues in the forthcoming annotation experiment, discussion on the annotation and guidelines among annotators was crucial; mainly in what referred to obscure annotation guidelines and ambiguous categories (namely, grammatical categories). Then we revisited Basque grammars and we updated the annotation guidelines adding more accurate information.

4.2 Second Annotation Round

After the grammatical reanalysis, a second annotation round has been conducted. This second time, four annotators have taken part; three of them were familiar with EusTimeML and the CAT annotation tool and the fourth one had a deep knowledge on temporal annotation as well as the guidelines and the annotation tool. The annotation has been done on 15 documents of the Basque version of the MEANTIME corpus [19] used in NewsReader project [20]. The first three annotators have annotated 115 sentences and their annotations have been compared to the fourth annotator's.

The number annotations for each annotator (A, B, C) and super-annotator (fourth annotator) and a counting of unanimously annotated events is given in table 1. The numbers already show a relatively high agreement.

The main reason for disagreement has been the difficulty to class some entities as events. In example (20) there is a linguistic form which expresses an event in the MEANTIME corpus and in example (21) there is the same form not expressing any event. This phenomenon has been more pronounced in the cases in which a form refers to a process and the final product of that process.

Table 1. Annotated events by each annotator and agreed events

Document sets	Annotator	Super-annotator	Agreed events
First (Ann. A)	96	74	69
Second (Ann. B)	394	418	358
Third (Ann. C)	95	99	84

(20) **Ekoizpena** AEBra ekartzeko asmoa du.
Production.DET.ABS USA.ALL bring.FIN intention.ABS has
'(He/She) intends to bring the **production** to the USA'.

(21) Nekazariek euren **ekoizpena** salgai jarriko dute.
Farmers.ERG their **production.ABS** to.be.sold put.FUT AUX
'Farmers will put their **production** on sale'.

State denoting events have also been a disagreement point. It has been sometimes difficult to decide whether they are events as they do not always express an ongoing state but a very generic situation.

(22) Oso **desengainatuta** gaude
Very **disappointed** are
'(We) are very **disappointed**'.

In Basque the verb *egin* (to do) is used to focus events expressed by verbs. This verb, does not offer any event information and, although it was stated not to annotate it, it has sometimes been annotated.

Table 2. Event extent agreement results

Annotator pairs	Micro-average (Markable)	Micro-average (Token)	Macro-average (Markable)	Macro-average (Token)
A – SA	0.812	0.812	0.819	0.819
B – SA	0.877	0.877	0.875	0.875
C – SA	0.866	0.866	0.883	0.883

Results in table 2 show a high agreement [18] on markable extent between annotators (the first three and the super-annotator). Markable extent agreement refers to the perfect overlap of the tags of two different annotators. Token extent agreement, instead, refers to the markable extent considering only the overlapping tokens. In our case both, markable and token extent, agreement results get the same values as markables have always a single-token extent. One may

Table 3. Unanimously annotated POS

Event annotation	**A-GS**	**B-GS**	**C-GS**
Verbs	45	242	51
Nouns	14	58	33
Adjectives	1	3	1
Advebrs	0	9	2
Pronouns	0	1	0
Other	0	0	0
TOTAL	60 (87%)	311 (87%)	77 (92%)

consequently deduce that all annotators have respected the single-token rule for event annotation in EustimeML guidelines.

As shown in table 3, a rather high agreement on the grammatical category of events has been reached. Most of the disagreement is due to one of the annotators not giving any value to an event or forgetting to change the default value. However some other disagreement is due to grammatical reasons.

- Verbal nouns ended with *-tea/-tzea* have been annotated as nouns and verbs.
- Participles with a relative mark *-tako/-riko* have been annotated as adjectives and verbs.
- Some adverbs have been considered part of the verb form and have been given a verb value or an "other" value.

Table 4. Modality agreement results

Modal event annotation	**A-GS**	**B-GS**	**C-GS**
BEHAR	0	5	2
NAHI	0	3	1
AHAL	3	4	0
TOTAL	3	12	3

The modal verbs unanimously annotated by the first three annotators and the gold standard can be seen in table 4. Modal events have been easy to identify, since there is little variation on the modality expressing forms. Moreover, there is virtually no possibility of confusedly giving a wrong value to a modal event expression as they have very distant meanings. Although the number of modal events is low, the result analysis has shown that mistakes in the annotation were due to annotators' mistakes during the annotation; not to wrong perceptions of those events.

Finally we have measured the agreement on event category. The results are not as high as expected (A-SA: 58%, B-SA: 56% and C-SA: 49%), however, it

is to mention that the agreement strongly varies between categories. Reporting and aspectual events have been easily identifiable, despite the fact that some have been incorrectly annotated presumably by mistake in many of the cases. Occurrence and intensional actions, instead, have been a major matter for disagreement.

From a thorough analysis of the agreement, we have noticed that the event documents that have been annotated later get higher agreement in event categorisation. Therefore, one may deduce that the more the training the better results in categorisation.

4.3 Final Guideline Tuning

Once we have analysed the annotation results, we have dropped some conclusions and have made some decisions:

- The more trained and familiar with the task an annotator is, the less mistakes will make and the higher agreement will achieve.
- In order not to forget filling or saving the attribute values, a means for it will be designed.
- Event identification and part-of-speech categorisation do not seem difficult to master.
- Although modal events have been correctly annotated in general, we expect further discussion and training on them to improve the results.
- Event categorisation agreement has been lower than expected. Although some categories seem easier to assign, we will set a new analysis guideline tuning period for the most conflictive.

After the corrections to the 15 annotated documents are done, the trained annotators will continue enlarging the gold standard corpus, as well as annotating more temporal structures such as time expressions, temporal linking constructions and temporal relations.

5 Conclusions and Future Work

The analysis and processing of event information is a very relevant task in text processing. Some information can be extracted, analysed and processed language-independently, but other needs a previous analysis on the forms of a certain language. In this paper, we have offered a summary of the linguistic forms in Basque that can express event information based on Basque grammars and we have also highlighted their main features. We have also classified the different eventive forms according to their semantics following the classification proposed in TimeML.

This information has been made explicit in EusTimeML, a temporal mark-up language for Basque, and the manual annotation guidelines have been written. A series of experiments to evaluate our linguistic decisions and create a corpus

annotated with temporal information have been run. First annotation effort served to train annotators and tune the annotation guidelines. The second has led to the creation of a gold standard corpus of event information.

This work is part of a complete analysis of temporal information and corpus creation for Basque. After having analysed temporal expressions and events, our ongoing research is focused on the analysis of the temporal relations between those events and time expressing constructions. With each analysis and annotation effort, we are building a gold standard corpus for temporal information in Basque. This corpus is expected to be used for the evaluation of automatic information extraction tools in a first instance.

Aknowledgement

This research is funded by the Basque Government PRE_2015_2_0284 grant.

References

1. Anne-Lyse Minard, Manuela Speranza, Eneko Agirre, Itziar Aldabe, Marieke van Erp, Bernardo Magnini, German Rigau, and Ruben Urizar. SemEval-2015 Task 4: TimeLine: Cross-Document Event Ordering. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 778–786, Denver, Colorado, June 2015. Association for Computational Linguistics.
2. James Pustejovsky, José M. Castaño, Robert Ingria, Roser Saurí, Robert J. Gaizauskas, Andrea Setzer, Graham Katz, and Dragomir R. Radev. TimeML: Robust Specification of Event and Temporal Expressions in Text. *New directions in question answering*, 3:28–34, 2003.
3. Branimir Boguraev, James Pustejovsky, Rie Ando, and Marc Verhagen. TimeBank evolution as a community resource for TimeML parsing. *Language Resources and Evaluation*, 41(1):91–115, 2007. `"http://dx.doi.org/10.1007/s10579-007-9018-8"`.
4. Paweł Mazur and Robert Dale. WikiWars: A New Corpus for Research on Temporal Expressions. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, EMNLP '10, pages 913–922, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics. `http://dl.acm.org/citation.cfm?id=1870658.1870747`.
5. Hector Llorens, Estela Saquete, and Borja Navarro. TIPSem (English and Spanish): Evaluating CRFs and Semantic Roles in TempEval-2. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, SemEval '10, pages 284–291, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics. `http://dl.acm.org/citation.cfm?id=1859664.1859727`.
6. Kira Radinsky and Eric Horvitz. Mining the web to predict future events. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 255–264. ACM, 2013.
7. Sandro Bauer, Stephen Clark, and Thore Graepel. Learning to identify historical figures for timeline creation from wikipedia articles. In *Proceedings of HistoInformatics2014 - the 2nd International Workshop on Computational History*, Barcelona, Spain, 2014.

8. TimeML Working Group. TimeML Annotation Guidelines version 1.3. Manuscript. Technical report, Brandeis University, 2010.
9. Begoña Altuna, María Jesús Aranzabe, and Arantza Díaz de Ilarraza. Euskarazko denbora-egiturak etiketatzeko gidalerroak v2.0. Technical report, Lengoaia eta Sistema Informatikoak Saila, UPV/EHU. UPV/EHU/LSI/TR;01-2016, 2016. `https://addi.ehu.es/handle/10810/17305`.
10. Roser Saurı, Olga Batiukova, and James Pustejovsky. Annotating Events in Spanish. TimeML Annotation Guidelines. Technical report, Technical Report Version TempEval-2010., Barcelona Media-Innovation Center, 2009.
11. Patxi Altuna, 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, Beat Oyharabal, and Karmele Rotaetxe. *Euskal Gramatika Lehen urratsak (EGLU) I*. Euskaltzaindiko Gramatika batzordea, Euskaltzaindia, Bilbo, 1985.
12. Patxi Altuna, 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, Beat Oyharabal, and Karmele Rotaetxe. *Euskal Gramatika Lehen Urratsak (EGLU) II*. Euskaltzaindiko Gramatika Batzordea, Euskaltzaindia, Bilbao, 1987.
13. José Ignacio Hualde and Jon Ortiz de Urbina. *A grammar of Basque*, volume 26. Walter de Gruyter, Boston, USA, 2003.
14. Ewa Jędrzejko. The problematics of describing periphrastic predication between word and image. *Studies in Polish Linguistics*, 6(1):27–44, 2011.
15. Tomasso Caselli, Valentina Bartalesi Lenzi, Rachele Sprugnoli, Emanuele Pianta, and Irina Prodanof. 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, 2011.
16. Begoña Altuna, María Jesús Aranzabe, and Arantza Díaz de Ilarraza. Euskarazko denbora-egiturak etiketatzeko gidalerroak. Technical report, Lengoaia eta Sistema Informatikoak Saila, UPV/EHU. UPV/EHU/LSI/TR;01-2016, 2014. `http://ixa.si.ehu.es/Ixa/Argitalpenak/Barne_txostenak/1414871293/publikoak/Denbora-egiturak%20etikeatzeko%20gidalerroak`.
17. Valentina Bartalesi Lenzi, Giovanni Moretti, and Rachele Sprugnoli. CAT: the CELCT Annotation Tool. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, pages 333–338, Istanbul, Turkey, 2012. European Language Resources Association (ELRA).
18. Lee Raymond Dice. Measures of the Amount of Ecologic Association between Species. *Ecology*, 26:297–302, 1945.
19. Anne-Lyse Minard, Manuela Speranza, Ruben Urizar, Begoña Altuna, Marieke van Erp, Anneleen Schoen, and Chantal van Son. MEANTIME, the NewsReader Multilingual Event and Time Corpus. In *Proceedings of LREC 2016*, 2016.
20. Rodrigo Agerri, Eneko Agirre, Itziar Aldabe, Begoña Altuna, Zuhaitz Beloki, Egoitz Laparra, Maddalen López de Lacalle, German Rigau, Aitor Soroa, and Rubén Urizar. Newsreader project. *Procesamiento del Lenguaje Natural*, 53:155–158, 2014.