This article has been published in a revised form in NATURAL LANGUAGE ENGINEERING [Volume 23 , Issue 4 , July 2017 , pp. 485 – 506 https://doi.org/10.1017/S1351324916000292].

One, No One and One Hundred Thousand Events: Defining and Processing Events in an Inter-Disciplinary Perspective

R. SPRUGNOLI[1–2], S. TONELLI[1]
[1]Fondazione Bruno Kessler,
Via Sommarive 18, Trento, Italy
[2]Università di Trento,
Via Sommarive 9, Trento, Italy
{sprugnoli,satonelli}@fbk.eu

(*Received X; X*)

Abstract

We present an overview of event definition and processing spanning 25 years of research in NLP. We first provide linguistic background to the notion of event, and then present past attempts to formalize this concept in annotation standards to foster the development of benchmarks for event extraction systems. This ranges from MUC-3 in 1991 to the Time Track challenge at SemEval 2015. Besides, we shed light on other disciplines in which the notion of event plays a crucial role, with a focus on the historical domain. Our goal is to provide a comprehensive study on event definitions and investigate which potential past efforts in the NLP community may have in a different research domain. We present the results of a survey, where the notion of event for historians is put in relation to the NLP perspective.

1 Introduction

In the last 25 years, several systems performing event extraction have been presented within the NLP community. Diverse approaches aimed at building timelines from large document collections have been implemented, and technologies to support automatic storytelling have become a relevant research topic in the AI community (Ashish et al., 2006). Event processing has been addressed from a variety of perspectives, from data visualization to knowledge representation and modelling. However, the notion of event has been revised several times and often tailored to the task of interest, so that a number of different definitions of event has been introduced since the first MUC evaluation campaign. Furthermore, the notion of event has been studied also in other disciplines, such as philosophy, cognitive science and history, which the NLP community has hardly taken into account.

A further distinction concerns two different research areas within NLP: in the

field of Topic Detection and Tracking (TDT), the identification of events is assimilated to the identification of topics within a stream of texts and the clustering of documents by topic[1]. Instead, in the field of Information Extraction (IE), the aim is to extract events expressed by words or phrases in a text. In this paper, we focus mainly on the latter perspective, since it has led to several standardisation proposals and evaluation campaigns, and to the creation of a wide community of researchers working at temporal processing tasks. However, we are aware that TDT is going to attract more and more attention, because it is particularly suitable to perform coarse-grained event detection on large streams of documents, for instance on social media data.

With this survey, we aim at providing a comprehensive overview of the way events have been defined in linguistics (Section 2) and in IE (Section 3), with a focus on the different evaluation campaigns organized over the years (Section 3.2). We also account for multilingual event processing, presenting tasks and corpora that cover languages other than English, and for new domains involved in recent event definition efforts (Section 3.3). Finally, we present a case study in Section 4, taking the perspective of history scholars, i.e. researchers from another area that typically deal with events in their daily activity. We try to address the following questions: was all the work devoted to event processing with IE techniques useful to serve real historical investigation? Were the various definitions of events provided over the years compatible with research practices adopted in other communities? How should events be defined to be processable with NLP tools but also to comply with historical research? We shed light on such questions by means of an online survey, in which historians were involved in an 'event definition' exercise.

2 Background: Definition of Events in Linguistics

Temporal Information Processing is an NLP task that aims at automatically detecting and interpreting events (e.g. *to live / the war*), temporal expressions (e.g. *20/05/2015 / this summer*) and temporal relations within texts (e.g. in *Waters recede before a tsunami* the event *recede* happened BEFORE the event *tsunami*).

Although event identification and processing may appear an easier task than the classification of temporal relations and expressions, which are often vague or implicit in natural language, this is still very challenging due to the ambiguous nature of the concept of event. The term 'event' itself has many readings: some authors use it to refer only to dynamic actions, others to refer also to static situations (Sasse, 2002). This terminological confusion mirrors the inherent complexity of the concept of event: in fact, an event may designate both an ontological and a linguistic category. However, between the ontological level and the linguistic one there is no one-to-one mapping because the same event may be expressed using various types of linguistic elements. As a matter of fact, even if verbs prototypically denote events

[1] According to the LDC annotation guidelines of the TDT task, "a topic is defined as an event or activity, along with all directly related events and activities", see https://catalog.ldc.upenn.edu/docs/LDC2005S11/tdt4guidelines v1 5.pdf

whereas nominals denote objects, this distinction is not clear-cut in natural language (Hagège, 1996)[2]. In particular, nominals exhibit a strong semantic ambiguity due to polysemy, showing alternations between eventive and non-eventive readings (Pustejovsky, 2005): for example, *administration* denotes an event in *spending grew during his administration* and a human group in *this administration is doing well*.

The best-known classification of events is the one proposed by (Vendler, 1967), who distinguishes between states (non-dynamic situations persisting over a period of time and without an endpoint, e.g. *believe*), activities (open-ended dynamic processes, e.g. *walk*), accomplishments (processes with a natural endpoint and an intrinsic duration, e.g. *build a house*), and achievements (almost instantaneous events with an endpoint, e.g. *find*). The Generative Lexicon theory revisits Vendler's classification introducing a three-way taxonomy of event types including states, processes, and transitions: in the latter category, accomplishments and achievements are collapsed (Pustejovsky, 1991). Moreover, in the literature, all types of actions, states and processes often fall under the cover term "eventualities", coined by (Bach, 1986) in his work on the algebra of events.

3 IE Perspective on Events

Starting from 1991, several evaluation campaigns and workshops dedicated to various aspects of temporal information processing and in particular to the analysis of the notion of event have been organized and have fostered the creation of a research community around the topic of event detection and processing. The timeline in Figure 1, built by collecting information from websites and proceedings, summarizes the history of workshops, in the lower part, and evaluation campaigns, in the upper part, related to temporal processing and organized starting from MUC-3.

We describe them in detail in the following subsections.

3.1 First studies on Events in the NLP community

In 2001, during the Workshop "Temporal and Spatial Information Processing" (TASIP), three relevant works dealing with event annotation and processing were presented, each of them relying on a different notion of event. (Filatova and Hovy, 2001), whose system assigns a position on a timeline to events in newspaper articles, define events as propositions that contain a subject and a predicate. The system achieves a precision of 0.55 and a recall of 0.60. (Schilder and Habel, 2001) present a tool for the automatic annotation of temporal expressions and events in news. The authors define events as expressions that have an implicit time dimension and are either verbs or noun phrases. The list of markable nouns is limited to those directly connected to a temporal expression or a temporal preposition (e.g. *after the*

[2] "(...) le verbe et le nom comme deux poles (...), constituer une sorte de champe magnétique où les catégories oscillent en subissant l'attracion soit de l'un soit de l'autre" trad. *the verb and the noun acting as poles constitute a kind of magnetic field where the categories fluctuate as they are attracted either by one or the other pole*.

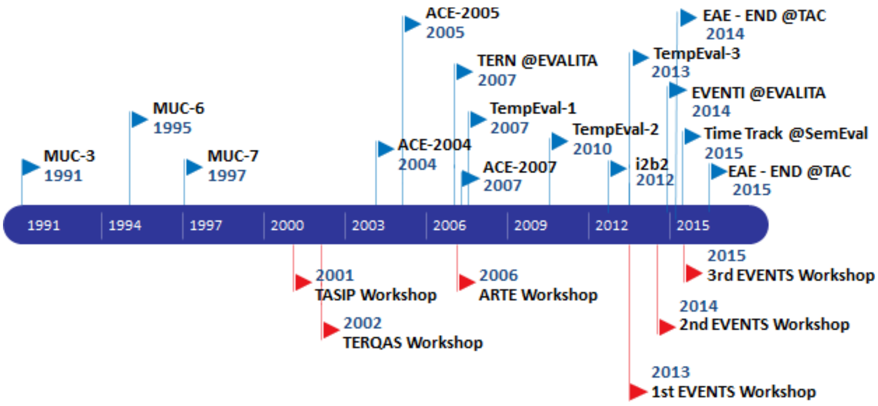

Fig. 1. Timeline of evaluation campaigns (above) and workshops (below) in the field of event detection and processing.

election in May) and belonging to the domain of interest (i.e. finance, *opening of the stock exchange*). In a further extension of the system, the authors perform event recognition through an ontology containing event-denoting nouns in the financial domain and information on event types (Schilder and Habel, 2003). Finally, (Katz and Arosio, 2001) propose a method to annotate temporal relations at sentence level limiting events to those realized by verbs. The three works highlight the need to achieve a consensus on a definition of event, aimed also at making automatic approaches comparable.

In that same year, (Setzer, 2001) presents STAG (*Sheffield Temporal Annotation Guidelines*), the first annotation scheme that takes into account all temporal information elements (i.e. events, temporal expressions, temporal relations and event identity). The author defines an event as something that happens, must be anchorable in time, can be instantaneous or may last for a period of time. States are therefore not taken into consideration and, from the linguistic point of view, candidate events include nominalizations, finite and non-finite verbs. Each event is associated with attributes giving grammatical and semantic information (e.g. aspect) .

Built upon STAG, TimeML (Pustejovsky et al., 2003a) is a scheme for the annotation of events, temporal expressions and relations between events and/or temporal expressions (i.e. temporal, aspectual and subordination relations). Following Bach's broad notion of event, TimeML identifies a wide range of linguistic expressions realizing events, i.e. tensed and untensed verbs (e.g. *was captured*, *to thank*), adjectives (e.g. *sick*), nominals (e.g. *strike*), and prepositional phrases (e.g. *on board*). The consolidation of TimeML as an international standard (ISO, 2008) has facilitated its adaptation to different languages, such as Spanish (Saur´ı, 2010) and Korean (Im et al., 2009), and the release of annotated data, such as the English TimeBank (Pustejovsky et al., 2003b) .

3.2 Evaluation Campaigns

Parallel to the works reported in the previous Subsection, several evaluation campaigns on temporal information extraction and processing have been carried out. As shown in Fig. 1, such campaigns have become very frequent in the last decade, with some years characterized by multiple evaluations.

The first campaign was the Message Understanding Conference (MUC-3) in 1991. It hosted the "Scenario Template" task, in which systems were required to identify information about a given event (e.g. an air vehicle launch) and relate such information to the entities involved in it. Thus, an event was considered as a set of relationships between participants, time and space: from a practical point of view, it was seen as a template with slots to be automatically filled. Low systems performance registered in the task throughout the years demonstrates its difficulty: the best result was recorded in MUC-6, where a system achieved an F-measure of 0.57, while in MUC-7 the best system achieved an F-measure of 0.51.

In the "Event Detection and Recognition" task, run for three years in the context of the ACE (Automatic Content Extraction) program, an event is a specific occurrence involving participants, something that happens and can often be described as a change of state (Linguistic Data Consortium, 2005). According to the ACE approach, extracting an event means marking up both the verb, noun, pronoun or adjective that most clearly expresses its occurrence (i.e. the event *trigger*) and the entire sentence containing that word (i.e. event *mention*). However, not all events are taken into account but only those belonging to a list of predefined types, each with a number of subtypes (e.g. the event type Conflict has two subtypes: Attack and Demonstrate). Each event is associated with the entities playing a role in it (e.g. the location target of an Attack event) and values are assigned to a set of attributes (e.g. genericity and tense). It is not possible to make a precise comparison between ACE and MUC results because the former adopted a different evaluation measure called Value Score (Doddington et al., 2004). However, the two initiatives share the same limitation: they were both designed around specific domains and too limited types of events (Grishman, 2010). Therefore, the proposed systems could hardly be adapted to different domains and applications. Moreover, the complexity of ACE annotation makes the creation of consistent labeled data very challenging.

In order to address this last shortcoming, the ERE (Entities, Relations, Events) scheme has been developed within the DARPA DEFT program (Aguilar et al., 2014), with the goal to propose a lighter-weight version of ACE. ACE and ERE share the same definition of events but ERE simplifies the annotation by collapsing tags, accepting a looser event extent and reducing the set of attributes and values. Recently, a transition between this simple scheme (also known as Light ERE) towards a more sophisticated representation of events has been proposed under the name of Rich ERE (Song et al., 2015). In Rich ERE, the event ontology and the number of attributes are expanded and more attention is devoted to event coreference. These DEFT ERE standards are the basis of the novel Event Nugget annotation scheme (Mitamura et al., 2015). An event nugget is a semantically meaningful unit referring to an event and linguistically represented not only by a single word but also by a continuous or discontinuous multi-token expression. The Knowledge Base Population evaluation track of the Text Analysis Conference (TAC KBP) conducted a pilot task on event nugget detection in 2014[3]: this same task is included also in the Event Track of TAC KBP 2015 [4].

Although the TAC KBP campaigns have been successful, their impact at large has been limited because the annotated datasets were distributed only to tasks participants. A different approach was adopted instead by TempEval organizers, who greatly contributed to improving state-of-the-art technologies in the field of Temporal Processing by making the data freely available after the campaigns. This consolidated also the success of TimeML annotation standard.

TempEval-1 (Verhagen et al., 2007) was the first open and international evaluation competition that used TimeBank as a benchmark. TempEval-1 avoids the complexity of complete temporal annotation focusing only on the identification of temporal relations between given pairs of temporal expressions and events. TempEval-2 (Verhagen et al., 2010) was a more complex campaign than the previous one: it was multilingual and consisted of 6 subtasks including event extent identification and classification of event attributes. This subtask was proposed also in TempEval-3 (UzZaman et al., 2013). Only one out of seven participants in the Event extraction and classification subtask uses a rule-based approach (Zavarella and Tanev, 2013). The best performing systems rely on a supervised approach both for event extraction and event type classification: TIPSem (Llorens et al., 2010), ATT1 (Jung and Stent, 2013) and KUL (Kolomiyets and Moens, 2013) are based on CRF, MaxEnt classification and Logistic Regression respectively. They all take advantage of morphosyntactic information (e.g. POS) and semantic features at both the lexical and the sentence level (e.g. WordNet synsets and Semantic Role labels). Best results in event extraction are around 0.80 F1-score. However, when dealing with the classification of event types, system performances drop by almost 10 points, with F1-scores all below 0.72.

At SemEval-2015[5] 3 tasks related to temporal processing have been proposed with a focus on new challenges, new evaluation approaches and new domains. The *TimeLine* task addressed coreference resolution of events and temporal relation extraction at a cross document level with the aim of build timelines (Minard et al., 2015). *QA TempEval* introduced an extrinsic evaluation that took into consideration a specific end-user application, i.e. question answering (Llorens et al., 2015). *Clinical TempEval* moved past TempEval efforts from the news to the clinical domain (Bethard et al., 2015).

As a wrap-up of the different annotation schemes described in this section, we present in Figure 2 the same sentence annotated according to ACE, Light ERE, Event Nugget, and TimeML guidelines. Differences in event types among ACE, Light ERE and Event Nugget are minimal (in this example are even null), while there is more variation concerning extension. ACE, Light ERE and TimeML anno-

[3] http://www.nist.gov/tac/2014/KBP/Event/index.html
[4] http://www.nist.gov/tac/2015/KBP/Event/index.html
[5] http://alt.qcri.org/semeval2015/

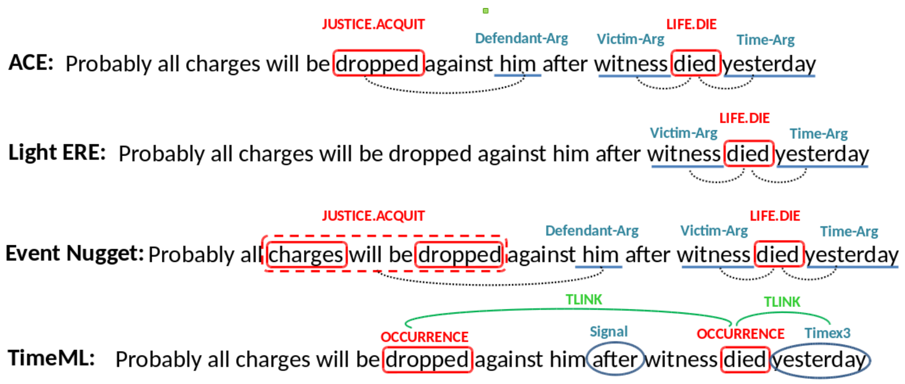

Fig. 2. Comparison of different event annotations. Red squares highlight event triggers while blue underlinings identify other annotated elements that in ACE, Light ERE and Event Nugget constitute event arguments. Connections between events and arguments are displays in dotted lines. For TimeML, temporal links are in green.

tate only events as single tokens, while Event Nugget annotation accounts also for multi-token and discontinuous expressions (*charges...dropped* in the third example). Moreover, in Light ERE only actual events are eligible to be annotated (this is why *dropped* is not annotated in the second example). All the other schemes, instead, include the annotation of probable, possible and negated events. In ACE, Light ERE and Event Nugget events are connected to their arguments, i.e. entities such as *him* and *witness*. In TimeML, instead, the focus is on temporal links between two events (e.g. *dropped* and *died*) or between an event and a temporal expression (e.g. *died* and *yesterday*). In general, ACE, Light ERE and Event Nugget combine information on events with their argument structure, while in TimeML the temporal dimension acquires more relevance, having its roots in Allen's interval algebra (Allen, 1984).

3.3 Adaptation of Event Processing to New Domains

Most evaluation exercises presented so far were concerned with event processing in the news domain. Only recently, NLP researchers have started to look at different domains and develop domain-specific annotation guidelines and systems. For instance, following an increased interest in the temporal processing of clinical records, ISO-TimeML has been adapted to the **clinical domain**: as a result, the THYME annotation guidelines have been developed[6]. Following such guidelines, an event is "anything relevant to the clinical timeline" (Styler et al., 2014), for example diseases, medical treatments and all actions and states related to the patient's clinical timeline. THYME guidelines formed the basis of both the i2b2 shared task in 2012 (Sun et al., 2013) and of the Clinical TempEval evaluation, organized within SemEval 2015 and aimed at assessing the performance of temporal information

[6] http://clear.colorado.edu/compsem/documents/

extraction systems on clinical notes and pathology reports[7]. An extension of the THYME guidelines integrating ISO-TimeML, the Stanford Event coreference (Lee et al., 2012) and the CMU Event coreference guidelines (Hovy et al., 2013) has been proposed by the University of Colorado at Boulder under the name of Richer Event Description (RED)[8]. RED adopts the TimeML wide definition of events and annotates events, temporal expressions and entities, as well as temporal, causal and coreference relations.

Systems for extracting events from **biomedical data** were evaluated in three editions of the BioNLP shared task in 2009, 2011, and 2013. In this field, event definition is strongly domain-dependent and datasets are annotated by biologists. During the last evaluation campaign, different tasks were proposed: systems were required to detect trigger words expressing molecular and sub-cellular events (e.g. *mutation*), assign a type to each event (e.g. *anatomical* or *pathological*), link events to their arguments (e.g. a molecule) and identify speculated and negated events (e.g. the failure of a mutation). The F-score of the best systems in event extraction was above 0.50 (Nédellec et al., 2013).

Event extraction from **social media** is another emerging area of research. Although most of the works in this field address the task as a clustering problem following the TDT approach mentioned in Section 1 (Petrović et al., 2010), (Ritter et al., 2012) apply IE techniques to identify events in a stream of tweets. The authors annotated manually event-referring phrases in a corpus of 1,000 tweets following TimeML event definition and developed an automatic tagger that deals with the complexity of Twitter language (i.e. informal and ungrammatical style) achieving an F-score of 0.64.

As for Humanities studies, there is a large research community which would benefit from the availability of temporal processing systems to extract events from textual data, especially in the **history domain**. However, NLP tools and methods have been applied so far to historical texts only partially, and mostly to analyze lexical and syntactic aspects of the language (Piotrowski, 2012).

In the literature, a number of works that try to tackle the semantics of historical texts using a combination of Semantic Web technologies and NLP approaches has been presented (Meroño-Peñuela et al., 2014). However, NLP techniques specifically developed for event processing have not been fully exploited and the current standardization efforts have received little attention in this domain. For example, in the *Agora project* (Van Den Akker et al., 2010), aimed at enriching museums metadata through the extraction of historical event names from unstructured texts, event extraction is assimilated to the recognition of named entities. Therefore, only named events, such as *French Revolution*, are taken into account.

Another choice usually made in projects dealing with historical document processing is that the extraction of events is limited to a set of specialized types. For example, in the *FDR/Pearl Harbor project* (Ide and Woolner, 2004) research was

7 http://alt.qcri.org/semeval2015/task6/

8 https://github.com/timjogorman/RicherEventDescription

carried out only on *communication* events. This choice was driven by the goal of the project, which was to help historians of WWII to search and retrieve information from documents (e.g. government correspondence and memoranda) written before the Pearl Harbor attack in 1941. This specific set of events was categorized based on FrameNet (Baker et al., 1998) by assigning verbs in the corpus of reference to the "Communication" frame and its sub-frames. Another project, the *Semantics of History*, focuses only on *conflict-related* and *motion* actions (Cybulska and Vossen, 2011).

Unlike what happened in the clinical domain, no attempt was made to find a domain-specific definition of event combining the historical perspective and ongoing research in the NLP field. Another weak point of current NLP research for historical texts is the scarcity of corpora fully annotated with temporal information. For example, files tagged within the projects described above have never been publicly released. Two notable exceptions are the ModeS TimeBank, containing Spanish texts from the 18th century, and the De Gasperi corpus, a collection of documents written by the Italian statesman Alcide de Gasperi and dating back to the beginning of the 20th century (see Table 2). Both were manually annotated following a language-specific adaptation of TimeML. ModeS TimeBank was employed for theoretical studies on the evolution of the Spanish language (Nieto et al., 2011), while the De Gasperi corpus was used to measure the performance of event extraction systems on historical texts within the EVENTI evaluation exercise[9].

In order to measure how an event extraction system trained on news performs on historical data, we present in Table 1 two comparisons. In the first row, we report the performance of the state-of-the-art Italian system *FBK-HLT-Time* (Mirza and Minard, 2014) obtained on news and on the De Gasperi corpus in the framework of the EVENTI evaluation campaign for event extraction. In the second row, we measure the performance of the Spanish version of TIPSem (Llorens et al., 2010) on the Modes TimeBank, and we compare it with the TempEval 2013 results of the same system (UzZaman et al., 2013). Both systems are supervised, were trained on news data and are available as off-the-shelf tools, thus they are run with the same settings on the two domains.

Table 1. *Comparison between the results obtained by two event extraction systems on contemporary news articles and historical texts.*

	News			**History**		
	P	R	F1	P	R	F1
FBK-HLT-Time	0.88	0.85	0.87	0.89	0.78	0.83
TIPSem	0.92	0.86	0.89	0.27	0.72	0.39

We observe that the two systems have a comparable performance on news, while

9 https://sites.google.com/site/eventievalita2014/

their behaviour is significantly different in the historical domain. This is due mainly to the characteristics of the historical corpora taken into account: the language in the De Gasperi corpus is very similar to contemporary Italian, and tokens corresponding to events are generally easy to recognise, therefore the drop in performance of FBK-HLT-Time is limited to 0.7 points of recall. The language of the ModeS TimeBank, instead, shows many diachronical language variations, which makes it difficult to achieve a good performance, especially as regards precision. This comparison shows that event recognition systems trained on news in some cases would be suitable for investigations in new domains, given that the event definition framework and the language of the documents to be processed are similar to the ones used for training. For instance, historians analysing corpora dating back to the previous century may still achieve satisfactory system performance. In contrast to the common belief in the NLP community, the main issue related to using text processing systems in different research domains does not lie in software adaptation, but rather in the portability of annotation schemes, since scholars in specific research areas often have their own frameworks to model phenomena such as events and temporal flows. This is particularly evident in the history domain, and will be discussed in detail in Section 4.

In order to account for all corpora annotated so far with event information in different domains and languages, we report a summary in Table 2. The information presented in the table was gathered through the direct analysis of the resources downloaded from the Web and merging data from scientific papers. Resources listed in the table have been annotated following different schemes and cover five domains, with a prevalence of the news domain. The number of corpora in the list shows the interest of the NLP community in event processing, and the most recent corpora confirm the growing attention for new domains, new languages and more complex tasks integrating event extraction.

Table 2. *Corpora including event annotation in different domains. For each corpus the language, number of tokens, number of files and number of annotated events are provided. The symbol "-" is used in case of missing information. Resources in boldface are available online at the moment of writing.*

Domain	Corpus	Lang	#Tokens	#Files	#events
NEWS	**ACE 2005 (training)**[a]	EN	259,889	599	4167
		ZH	307,991	633	3332
	French TimeBank (Bittar et al., 2011)	FR	15,423	109	2,115
	Romanian TimeBank (Forascu and Tufi, 2012)	RO	65,375	181	7,926
	TimeBankPT (Forascu and Tufi, 2012)	PT	69,702	182	7,887
	Persian TimeBank (Yaghoobzadeh et al., 2012)	FA	26,949	43	4,237
	Catalan TimeBank 1.0[b]	CA	75,800	210	12,342
	Spanish TimeBank 1.0[c]	ES	75,800	210	12,641
	BCCWJ-TimeBank (Asahara et al., 2013)	JA	56,518	54	3,824
	EVENTI corpus (Caselli et al., 2014)	IT	130,279	366	21,633
	TempEval 1 (training)[d]	EN	52,740	162	5,150
		ZH	32,788	61	1,204
	TempEval 2 (training+test)[e]	EN	62,613	184	2,256
		IT	31,995	66	1,036
		FR	13,387	98	248
		KO	16,900	28	602
		ES	56,880	212	2,129
	TempEval-3 (AQUAINT+TimeBank+Platinum)[f]	EN	102,375	276	12,534
	FactBank (Saurí and Pustejovsky, 2009)	EN	77,000	208	9,500
	2012 EventCorefBank (ECB) (Lee et al., 2012)	EN	-	482	2,533
	ECB+ (Cybulska and Vossen, 2014)	EN	377,367	982	15,003
	Light ERE[g]	ZH	200,000	-	-
		EN	570,000	-	-
		ES	100,000	-	-
	Rich ERE (Song et al., 2015)	EN	32,240	91	-
	Event Nugget (training+test) (Mitamura et al., 2015)	EN	336,126	351	10,719
	TimeLine (Minard et al., 2015)	EN	29,893	90	915
CLINICAL	i2b2 (Sun et al., 2013)	EN	178,000	349	30,000
	Clinical TempEval (Train+Dev) (Bethard et al., 2015)	EN	533,393	440	59,864
BIOMEDICAL	**GENIA** (Kim et al., 2008)	EN	-	1,000	36,114
SOCIAL MEDIA	Twitter NLP (Ritter et al., 2012)	EN	19,484	1,000	-
HISTORY	**ModeS TimeBank**[h]	ES	25,611	102	1,261
	De Gasperi Corpus (Caselli et al., 2014)	IT	5,671	10	1,195

[a] https://catalog.ldc.upenn.edu/LDC2006T06
[b] https://catalog.ldc.upenn.edu/LDC2012T10
[c] https://catalog.ldc.upenn.edu/LDC2012T12
[d] http://www.timeml.org/tempeval/
[e] http://timeml.org/tempeval2/
[f] http://www.cs.york.ac.uk/semeval-2013/task1
[g] Light ERE, Rich ERE and Event Nugget corpora include both news and discussion forum data
[h] https://catalog.ldc.upenn.edu/LDC2012T01

4 What is an Event in History?

As shown in the previous Section, past projects trying to apply NLP techniques to historical investigation have adopted heterogeneous approaches, and there has been no real effort among history scholars to standardize event definition taking into account the proposals made in the NLP community and presented in the previuos sections. However, researchers in history are daily faced with issues related to the observation, analysis and interpretation of events. This gap between the two research communities may depend on a lack of communication and cross-fertilization, but also on the fact that events as defined in IE do not fully satisfy requirements from other disciplines. In order to clarify the reasons of this gap, we ran an investigation involving historians based on an online questionnaire. The questionnaire was distributed via social media (i.e. Twitter and LinkedIn), mailing-lists (e.g. the Humanist Discussion Group) and targeted emails to individual historians, professional associations (e.g. the Australian Historical Association) and research centers (e.g. Institute of Historical Research). After two months from its launch, 74 historians participated in the survey.

The general goal of this analysis was to leverage knowledge about the way events are defined in historical research and to compare it with ongoing standardization efforts in the NLP community. To the best of our knowledge, this is the first questionnaire on this topic, whose outcome can potentially enrich the current theoretical discussion on the nature of events. Besides, it can be seen as a preliminary step towards the definition of annotation guidelines for developing NLP tools in this domain.

4.1 Questionnaire Description and Results

The questionnaire was composed of two parts. In the first one, we collected participants' demographic information and assessed their general interest in NLP. The second part aimed at shedding light on the notion of "event" for historians based on three questions.

In the **first question**, participants were asked to list all the single words or expressions encoding events (if any) in three given sentences, without providing any definition of what an event is. The aim was to indirectly leverage an operational definition of events based on historians' knowledge.

The sentences, shown in Table 3, are taken from J. K. Kennedy's public speeches[10] and are rich of linguistic phenomena: negated verbs (e.g. *has not been prepared*), nominalizations (e. *disarmament*), aspectual nominals (e.g. *end*), cognitive verbs (e.g. *think*), named events (e.g. *Korean war*), nominals expressing states (e.g. *independence*), and phrasal verbs (e.g. *taken place*). A high percentage of respondents did not detect any event in the first and the third sentence containing a state (i.e. *independence*) and an opinion (i.e. *I think...*) respectively. The majority of the identified events are non-verbal. This contrasts with the outcome of the experiment

[10] Available at http://www.presidency.ucsb.edu/1960_election.php

Table 3. *Sentences annotated by the questionnaire participants. For each sentence, we report the percentage of responders who did not detect any event and the percentage of participants who annotated at least one of the given event extent.*

Sentences	NO Events	Non Verbal Events	Multi Token Events	Clauses	Most Frequent Extents
Today, once again, the independence of the Western Hemisphere is menaced from abroad	40%	75%	32%	4%	today menaced independence
This country has not been prepared for any disarmament, arms control or atomic testing conference that has taken place since the end of the Korean war	3%	81%	68%	11%	conference end of Korean war disarmament
I think we can work that out with the advice of the Ways and Means Committee	60%	52%	44%	4%	advice work that out think

reported by (Hatzivassiloglou and Filatova, 2003), in which nouns such as *war* and *earthquake* were never identified as events by a group of students. Events consisting of more than one token are annotated very frequently in all the sentences and some of them correspond to entire clauses (e.g. *This country has not been prepared*). Such choice goes against the TimeML and RED minimal chunk rule for tag extent, according to which only single tokens are to be annotated as events. The distinction made in ACE and ERE between event mention, that is usually the entire sentence, and event trigger seems to better meet historians' needs. Moreover, ACE, ERE and Event Nugget allow the annotation of multi-token event triggers (the latter also discontinuous cases).

Conclusion 1. The notion of event is seen as independent from its grammatical relation, in line with TimeML. However, the minimal chunk annotation used in TimeML is not optimal. Among the considered standards, the multi-token annotation of continuous and discontinuous multi-token expressions proposed in Event Nugget addresses best historians' view on events.

In the **second question**, we asked participants to rate the relevance of a list of properties to define when a word or expression can be labeled as an event. Such properties included for instance impact, cause and frequency, and were inspired by the essay "What is an Event?" [11] written by the history scholar Robert Bedrosian. The ratings included 4 possible values, i.e. "very important", "somewhat important", "not important", and "don't know". Figure 3 presents the value distribution

[11] http://rbedrosian.com/event.htm

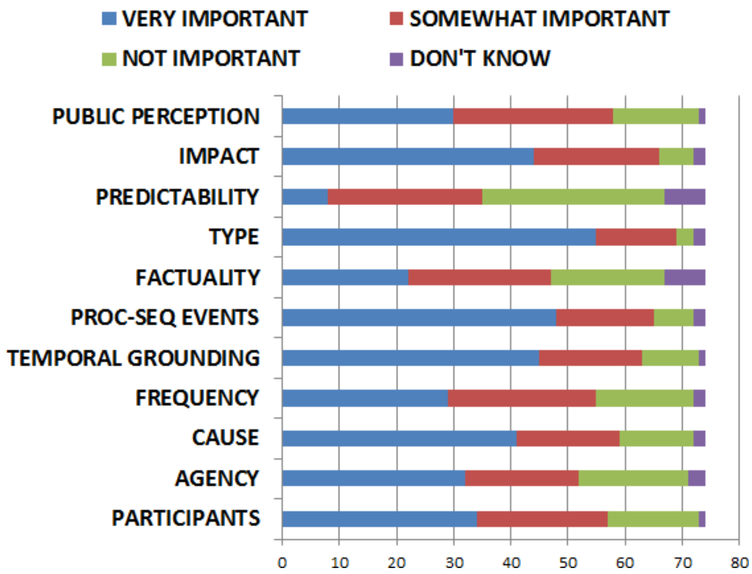

Fig. 3. What are the most important properties for a historian in order to understand if a word (or a set of words) expresses a relevant event.

across the properties. *Public Perception* and *Impact*, i.e. the degree to which an event affects the society or the nature, are proprieties not related to the linguistic analysis of texts but to the historian's interpretative work. Both were considered quite relevant, especially the latter. *Predictability* is the only property in which the value "not important" prevails. On the contrary, *Type* has the highest positive consensus. In TimeML, event type information is conveyed by seven possible values of the class attribute, where both semantic (e.g. STATE) and syntactic criteria (e.g. I_STATE) are taken into account. On the other hand, the event ontology of ACE, ERE and Event Nugget is made of a list of types and subtypes which limits the annotation to a specific set of categories (e.g. type: JUSTICE, subtype ACQUIT in Figure 2). *Factuality*, i.e. the distinction between actual real facts and imaginary, future, avoided and prevented events, has a limited interest for historians, while it is more relevant from a linguistic perspective. In fact, this property is encoded in TimeML (through SLINKs) as well as in other annotation schemes, where it appears as an attribute attached to the event (Saurí and Pustejovsky, 2009; van Son et al., 2014)[12]. *Preceding and consequent events* appear to be very important for historians, and this is in line with the ongoing effort in NLP to encode intra- and cross-document event ordering. In TimeML, this information is conveyed by TLINKs (i.e. Temporal LINK), corresponding to 13 types of binary temporal relations, inherited from Allen's interval algebra (Allen, 1984). Besides, the challenge of cross-document event ordering has been recently addressed by the "TimeLine" task

[12] See also the FactBank framework, Table 2.

at SemEval-2015[13]. In TimeML, the TLINK tag is also employed to link events to points in time (e.g. *25/12/2014*), durations (e.g. *3 month*) and temporal expressions denoting recurring times (e.g. *every month*): this corresponds to the *Temporal Grounding* property, that is the degree to which an event can be pinpointed to a particular time or period, and the *Frequency* property. In the MUC Scenario Template as well as in ACE, ERE and Event Nugget, temporal relations between events or between an event and a temporal expression are not explicitly addressed. The link between an event and a temporal expression is encoded in the form of a temporal slot in case of MUC or of a temporal argument in case of ACE, ERE and Event Nugget (e.g. the *Time-Arg* argument "yesterday" of the event trigger "died" in Figure 2). The property of an event being the cause or the effect of another event (i.e. *Cause*) is strictly connected to the *Agency* property, i.e. who/what caused such event. TimeML does not include a specific relation for causative constructions but causes and effects denoted by events are temporally ordered using a TLINK (a cause always precedes the effect). However, attempts have been made to explicitly annotate causal relations as an extension of TimeML (Mirza and Tonelli, 2014). In ACE, ERE and Event Nugget, *Agency* is annotated as event argument for several event types, e.g. in the sentence "his father-in-law killed him", *father-in-law* is the Agent argument of the trigger event *killed* of type LIFE. Event-event causality relations are planned as future development of the Rich ERE annotation but they are currently not included in the guidelines. On the contrary, causal relations play an important role in the RED guidelines (Hovy et al., 2013). As for *Participants*, TimeML does not foresee the annotation of the entities involved in an event, even if the historians' responses suggest that this information is quite relevant. Attempts have been made to add participants' information to events (Pustejovsky et al., 2007), but this has not led to the extension of TimeML specifications. On the contrary, participants annotation is crucial in MUC, ACE, ERE and Event Nugget, in which several arguments have to be identified (e.g. *Victim-Arg* in Figure 2). In addition, research on semantic roles can provide much guidance in this respect, for example by taking inspiration from PropBank (Palmer et al., 2005) or FrameNet frameworks as proposed in the NewsReader project[14].

Conclusion 2. An event is a complex information object characterized by many properties. A new framework for the annotation of events in historical texts should take advantage of the temporal dimension as defined in TimeML but also look at other annotation efforts (e.g. semantic roles in FrameNet, participants' information in Event Nugget) to cover all important properties.

Finally, in the **third question**, participants were asked to choose between two linguistic annotations of the following passage taken from a speech uttered by J.F.Kennedy:

After the key African state of Guinea, now [*voting*]$_{1-2}$ *with the Soviet Union in Communist foreign policy, after it* [*gained*]$_{1-2}$ *its* [*independence*]$_{2}$*, a Russian Ambassador* [[*showed*]$_{1}$ *up*]$_{2}$ *the next day. Our Ambassador did not* [[*show*]$_{1}$ *up*]$_{2}$ *for 9 months.*

13 http://alt.qcri.org/semeval2015/task4/
14 http://www.newsreader-project.eu/

The annotation marked with $[...]_1$ follows TimeML specifications, thus only single tokens are annotated as events and states are annotated only if temporally bound to a particular point or period of time (therefore *independence* is not annotated). The option marked with $[...]_2$ proposes looser criteria, annotating multi-token event expressions and also states in every context of occurrence. Only 5% of participants preferred the strict TimeML annotation, 61% chose the second option and the rest did not give preference to any of the two annotations. We asked the motivations behind this choice: answers highlighted some weak points of the TimeML annotation, for instance that a broader context is needed to represent events ("An event is not one word, it's syntactical, inter-relation between agent and object/patient"). Besides, states and conditions are important even if not bound to a temporal expression ("I feel that the state/condition is important."). In ACE, ERE and Event Nugget, states that result from actions, such as being *dead*, *married* or *retired*, are included in the annotation, but disagreement is an open issue for human annotators (Mitamura et al., 2015).

Conclusion 3. TimeML could not be applied to a new domain as is. States/conditions are important and should be considered in the annotation of historical documents.

5 Conclusions

This paper presents a survey of the state of the art in event definition and processing in NLP, adopting an inter-disciplinary perspective. In the last 25 years, thanks to many workshops and evaluation exercises dedicated to the semantic and linguistic analysis of events, research has moved forward. However, a careful adaptation of existing annotation schemes is necessary to apply the outcome of these research activities to new domains. On the basis of the analysis of the state of the art and of the historians' replies to our questionnaire, we can now answer the questions posed in Section 1:

(i) *Was all the work devoted to event processing with IE techniques useful to serve real historical investigation?* NLP methods and technologies have not been fully exploited yet in the domain of history. Existing annotation schemes and systems constitute an important starting point but a careful adaptation is necessary to meet the requirements of domain experts.

(ii) *Were the various definitions of events provided over the years compatible with research practices adopted in other communities?* Several event definitions have been proposed over the years, each showing specific strengths and weaknesses. TimeML event definition relies on the broad notion of eventuality: the fact that it includes states as well as processes and actions is compatible with historians' needs. On the other hand, states should be taken into consideration even if not bound to a specific point or period of time. Allowing only single token events does not meet research practices adopted in other domains. The multi-token choice proposed in the Event Nugget initiative addresses better this need.

(iii) *How should events be defined to be processable with NLP tools but also to comply with historical research?* Events can be defined as complex information objects characterized by many properties. These can be cast by combining different

NLP analyses providing rich semantic information, such as semantic role labeling, causality detection and temporal relation processing.

References

Aguilar, Jacqueline and Beller, Charley and McNamee, Paul and Van Durme, Benjamin and Strassel, Stephanie and Song, Zhiyi and Ellis, Joe 2014. A Comparison of the Events and Relations Across ACE, ERE, TAC-KBP, and FrameNet Annotation Standards. In Proceedings of the Second Workshop on EVENTS: Definition, Detection, Coreference, and Representation, pages 45-53, ACL.

J.F. Allen. 1984. Towards a general theory of action and time. *Artificial Intelligence*, 23(2):123–54.

Masayuki Asahara, Sachi Yasuda, Hikari Konishi, Mizuho Imada, and Kikuo Maekawa. 2013. BCCWJ-TimeBank: Temporal and Event Information Annotation on Japanese Text. In *Proceedings of PACLIC 27*, pages 206–214.

Naveen Ashish, Doug Appelt, Dayne Freitag, and Dmitry Zelenko. 2006. Papers from the AAAI Workshop on Event Extraction and Synthesis. Technical Report WS-06-07, American Association for Artificial Intelligence.

E. Bach. 1986. The algebra of events. *Linguistics and Philosophy*, 9:5–16.

Collin F Baker, Charles J Fillmore, and John B Lowe. 1998. The Berkeley FrameNet project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*, pages 86–90.

Steven Bethard, Leon Derczynski, Guergana Savova, James Pustejovsky, and Marc Verhagen. 2015. SemEval-2015 Task 6: Clinical TempEval. In *Proceedings of SemEval 2015*. ACL.

André Bittar, Pascal Amsili, Pascal Denis, and Laurence Danlos. 2011. French TimeBank: An ISO-TimeML Annotated Reference Corpus. In *Proceedings of ACL*, pages 130–134, Portland, Oregon, USA. ACL.

Tommaso Caselli, Rachele Sprugnoli, Manuela Speranza, and Monica Monachini. 2014. EVENTI EValuation of Events and Temporal INformation at Evalita 2014. In *Proceedings of the Fourth International Workshop EVALITA 2014*.

Agata Cybulska and Piek Vossen. 2011. Historical event extraction from text. In *Proceedings of the 5th ACL-HLT LaTeCH Workshop*, pages 39–43. ACL.

Agata Cybulska and Piek Vossen. 2014. Using a sledgehammer to crack a nut? lexical diversity and event coreference resolution. In *Proceedings of LREC2014*, pages 26–31.

George R Doddington, Alexis Mitchell, Mark A Przybocki, Lance A Ramshaw, Stephanie Strassel, and Ralph M Weischedel. 2004. The Automatic Content Extraction (ACE) Program-Tasks, Data, and Evaluation. In *Proceedings of LREC 2004*.

Elena Filatova and Eduard Hovy. 2001. Assigning time-stamps to event-clauses. In *Proceedings of the ACL-EACL 2001 Workshop for Temporal and Spatial Information Processing*.

Corina Forascu and Dan Tufi. 2012. Romanian timebank: An annotated parallel corpus for temporal information. In *Proceedings of LREC'12*. European Language Resources Association (ELRA).

Ralph Grishman. 2010. The impact of task and corpus on event extraction systems. In Proceedings of *LREC 2010*.

Hagège, Claude. 1996. L'homme de paroles: contribution linguistique aux sciences humaines. Fayard.

Vasileios Hatzivassiloglou and Elena Filatova. 2003. Domain-independent detection, extraction, and labeling of atomic events.

Hovy, E., Mitamura, T., Verdejo, F., Araki, J., Philpot, A.. 2013. Events are not simple: Identity, non-identity, and quasi-identity. In Proceedings of the The 1st Workshop on EVENTS: Definition, Detection, Coreference, and Representation, ACL.

Nancy Ide and David Woolner. 2004. Exploiting Semantic Web Technologies for Intelligent Access to Historical Documents. In *Proceedings of LREC 2004*.

Ikuta, R., Styler IV, W. F., Hamang, M., OGorman, T., Palmer, M.. 2014. Challenges of Adding Causation to Richer Event Descriptions. In Proceedings of the The 2nd Workshop on EVENTS: Definition, Detection, Coreference, and Representation, ACL.

Seohyun Im, Hyunjo You, Hayun Jang, Seungho Nam, and Hyopil Shin. 2009. Ktimeml: specification of temporal and event expressions in korean text. In *Proceedings of the 7th Workshop on Asian Language Resources*, pages 115–122. Association for Computational Linguistics.

SemAf/Time Working Group ISO, 2008. *ISO DIS 24617-1: 2008 Language resource management - Semantic annotation framework - Part 1: Time and events*. ISO Central Secretariat, Geneva.

Hyuckchul Jung and Amanda Stent. 2013. Att1: Temporal annotation using big windows and rich syntactic and semantic features. In *Proceedings of * SEM*, volume 2, pages 20–24.

Graham Katz and Fabrizio Arosio. 2001. The annotation of temporal information in natural language sentences. In *Proceedings of the ACL 2001 Workshop on Temporal and Spatial Information Processing*.

Jin-Dong Kim, Tomoko Ohta, and Jun'ichi Tsujii. 2008. Corpus annotation for mining biomedical events from literature. *BMC bioinformatics*, 9(1):10.

Oleksandr Kolomiyets and Marie-Francine Moens. 2013. KUL: A data-driven approach to temporal parsing of documents. In *Proceedings of SemEval 2013*, pages 83–87.

Heeyoung Lee, Marta Recasens, Angel Chang, Mihai Surdeanu, and Dan Jurafsky. 2012. Joint entity and event coreference resolution across documents. In *Proceedings of EMNLP-2012*, pages 489–500.

Linguistic Data Consortium. 2005. Ace (automatic content extraction) english annotation guidelines for events, version 5.4.3 2005.07.01.

Hector Llorens, Estela Saquete, and Borja Navarro. 2010. TIPSem (English and Spanish): Evaluating CRFs and Semantic Roles in TempEval-2. In *Proceedings of SemEval-2010*, pages 284–291. ACL.

Llorens, Hector and Chambers, Nathanael and UzZaman, Naushad and Mostafazadeh, Nasrin and Allen, James and Pustejovsky, James. 2015. SemEval-2015 Task 5: QA TEMPEVAL-Evaluating Temporal Information Understanding with Question Answering. In *Proceedings of SemEval 2015*. ACL.

Albert Meroño-Peñuela, Ashkan Ashkpour, Marieke van Erp, Kees Mandemakers, Leen Breure, Andrea Scharnhorst, Stefan Schlobach, and Frank van Harmelen. 2014. Semantic technologies for historical research: A survey. *Semantic Web Journal*.

Anne-Lyse Minard, Manuela Speranza, Eneko Agirre, Itziar Aldabe, Marieke van Erp, Bernardo Magnini, German Rigau, and Ruben Urizar. 2015. Semeval-2015 task 4: Timeline: Cross-document event ordering. In *Proceedings of SemEval 2015*. ACL.

Mirza, Paramita and Minard, Anne-Lyse. 2014. FBK-HLT-time: a complete Italian Temporal Processing system for EVENTI-Evalita 2014. In Proceedings of the Fourth International Workshop EVALITA 2014. Pisa University Press.

Paramita Mirza and Sara Tonelli. 2014. An analysis of causality between events and its relation to temporal information. In *Proceedings of COLING 2014*, pages 2097–2106. Dublin City University and ACL.

Mitamura, Teruko and Yamakawa, Yukari and Holm, Susan and Song, Zhiyi and Bies, Ann and Kulick, Seth and Strassel, Stephanie. 2015. Event Nugget Annotation: Processes and Issues. In Proceedings of the The 3rd Workshop on EVENTS: Definition, Detection, Coreference, and Representation, pages 66–76, ACL.

Claire Nédellec, Robert Bossy, Jin-Dong Kim, Jung-Jae Kim, Tomoko Ohta, Sampo Pyysalo, and Pierre Zweigenbaum. 2013. Overview of BioNLP shared task 2013. In *Proceedings of the BioNLP Shared Task 2013 Workshop*, pages 1–7.

Nieto, Marta Guerrero, Roser Saur´ı, and Miguel ngel Bernab Poveda. 2011. ModeS TimeBank: A Modern Spanish TimeBank Corpus. *Procesamiento del lenguaje natural*, 47 (2011): 259-267.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, 31.

Saša Petrović, Miles Osborne, and Victor Lavrenko. 2010. Streaming first story detection with application to twitter. In *Proceedings of NAACL 2010*, pages 181–189.

Michael Piotrowski. 2012. Natural Language Processing for historical texts. *Synthesis Lectures on Human Language Technologies*, 5(2):1–157.

James Pustejovsky, José M. Castaño, Robert Ingria, Roser Saurí, Robert Gaizauskas, Andrea Setzer, and Graham Katz. 2003. TimeML: Robust Specification of Event and Temporal Expressions in Text. In *Proceedings of IWCS-5*.

James Pustejovsky, Patrick Hanks, Roser Saur´ı, Andrew See, Robert Gaizauskas, Andrea Setzer, Dragomir Radev, Beth Sundheim, David Day, Lisa Ferro and Marcia Lazo. 2003. The TIMEBANK Corpus. In *Proceedings of Corpus Linguistics 2003*, , pages 647–656.

James Pustejovsky. 2005. A survey of dot objects. Manuscript.

James Pustejovsky, Jessica Littman, and Roser Saur´ı. 2007. Arguments in timeml: events and entities. In *Annotating, Extracting and Reasoning about Time and Events*, pages 107–126. Springer.

J. Pustejovsky. 1991. The syntax of event structure. *Cognition*, 41(1-3):47–81.

Alan Ritter, Oren Etzioni, Sam Clark, et al. 2012. Open domain event extraction from Twitter. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1104–1112. ACM.

Sasse, Hans-Jürgen. 2002. Recent activity in the theory of aspect: Accomplishments, achievements, or just non-progressive state. *Linguistic Typology*, 6(2):199–271.

Roser Saur´ı and James Pustejovsky. 2009. Factbank: A corpus annotated with event factuality. *Language resources and evaluation*, 43(3):227–268.

Roser Saurı. 2010. Annotating temporal relations in catalan and spanish timeml annotation guidelines.

F. Schilder and C. Habel. 2001. From temporal expressions to temporal information: Semantic tagging of news messages. In *Proceedings of the ACL-EACL 2001 Workshop for Temporal and Spatial Information Processing*.

F. Schilder and C. Habel. 2003. Temporal information extraction for temporal question answering. In *New Directions in Question Answering*, pages 35–44.

Andrea Setzer. 2001. *Temporal information in newswire articles: An annotation scheme and corpus study*. Ph.D. thesis, University of Sheffield.

Song, Zhiyi and Bies, Ann and Strassel, Stephanie and Riese, Tom and Mott, Justin and Ellis, Joe and Wright, Jonathan and Kulick, Seth and Ryant, Neville and Ma, Xiaoyi. 2015. From Light to Rich ERE: Annotation of Entities, Relations, and Events. In Proceedings of the The 3rd Workshop on EVENTS: Definition, Detection, Coreference, and Representation, pages 89–98, ACL.

William F. Styler, IV, Steven Bethard, Sean Finan, Martha Palmer, Sameer Pradhan, Piet C. de Groen, Brad Erickson, Timothy Miller, Chen Lin, Guergana Savova, and James Pustejovsky. 2014. Temporal annotation in the clinical domain. *Transactions of the Association for Computational Linguistics*, 2:143–154.

Weiyi Sun, Anna Rumshisky, and Ozlem Uzuner. 2013. Evaluating temporal relations in clinical text: 2012 i2b2 challenge. *Journal of the American Medical Informatics Association*, pages amiajnl–2013.

Naushad UzZaman, Hector Llorens, Leon Derczynski, James Allen, Marc Verhagen, and James Pustejovsky. 2013. Semeval-2013 task 1: Tempeval-3: Evaluating time expressions, events, and temporal relations. In *Proceedings of SemEval 2013*, pages 1–9. ACL.

Chiel Van Den Akker, Lora Aroyo, Agata Cybulska, Marieke Van Erp, Peter Gorgels, Laura Hollink, Cathy Jager, Susan Legene, Lourens van der Meij, Johan Oomen, et al. 2010. Historical event-based access to museum collections. In *Proceedings of EVENTS2010*.

Chantal van Son, Marieke van Erp, Antske Fokkens, and Piek Vossen. 2014. Hope and fear: Interpreting perspectives by integrating sentiment and event factuality. In *Proceedings of LREC2014*, pages 26–31.

Z. Vendler, 1967. *Linguistics and philosophy*, chapter Verbs and times, pages 97–121. Cornell University Press, Ithaca, NY.

Marc Verhagen, Robert Gaizauskas, Frank Schilder, Mark Hepple, Graham Katz, and James Pustejovsky. 2007. Semeval-2007 task 15: Tempeval temporal relation identification. In *Proceedings of SemEval-2007*, pages 75–80. ACL.

Marc Verhagen, Roser Sauri, Tommaso Caselli, and James Pustejovsky. 2010. Semeval-2010 task 13: Tempeval-2. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 57–62. ACL.

Yadollah Yaghoobzadeh, Gholamreza Ghassem-Sani, Seyed Abolghassem Mirroshandel, and Mahbaneh Eshaghzadeh. 2012. ISO-TimeML Event Extraction in Persian Text. In *Proceedings of COLING 2012*, pages 2931–2944.

Vanni Zavarella and Hristo Tanev. 2013. FSS-TimEx for TempEval-3: Extracting Temporal Information from Text. In *Proceedings of SemEval 2013*, pages 58–63, Atlanta, Georgia, USA. ACL.