Toward Real Event Detection

Michael Färber[*] and Achim Rettinger

Karlsruhe Institute of Technology (KIT), Institute AIFB, Karlsruhe, Germany
{michael.faerber,rettinger}@kit.edu

Abstract. News agencies and other news providers or consumers are confronted with the task of extracting events from news articles. This is done i) either to monitor and, hence, to be informed about events of specific kinds over time and/or ii) to react to events immediately. In the past, several promising approaches to extracting events from text have been proposed. Besides purely statistically-based approaches there are methods to represent events in a semantically-structured form, such as graphs containing actions (predicates), participants (entities), etc. However, it turns out to be very difficult to automatically determine whether an event is real or not. In this paper, we give an overview of approaches which proposed solutions for this research problem. We show that there is no gold standard dataset where *real events* are annotated in text documents in a fine-grained, semantically-enriched way. We present a methodology of creating such a dataset with the help of crowdsourcing and present preliminary results.

Keywords: Event Detection, Information Extraction, Factuality.

1 Motivation

News agencies and other digital media publishers publish each day news articles in the magnitude of dozens of thousands. They also process the news for further business tasks such as trend prediction and market change detection. This is still mainly done manually today. Even if knowledge workers at news agencies have access to all this information, it is infeasible for them to read all the news and to determine, whether the articles contain information which is not only interesting for people in their domains, but which contain real events and, hence, have a significant, immediate impact on business such as financial operations (shares) and political happenings. Consider for example the first sentence of a news article:

$$\text{``Apple may acquire Beats Electronics next week.''} \qquad (1)$$

Here, it remains unclear whether Apple is really going to acquire Beats (and does not cancel it in the last minute) or whether this is just a rumor. The sentence

$$\text{"Apple confirmed that it acquired Beats Electronics on Wednesday." \qquad (2)}$$

in contrary, reveals that the acquisition already happened (besides the confirmation which is an event per se). This demonstrates the differentiating characteristic between real events and events in general. As humans we can estimate that the first article is not a trigger for immediate shifts in the stock market (besides psychological effects), but maybe the second mentioned article. Machines, in contrast, have their difficulties in distinguishing real events from other events.

We envision building a decision support tool for agents like stockbrokers. The aim of the system is to inform the user quickly and automatically when some detected event has really happened and hence might influence the invested assets of the user. The user should also have the possibility to store purely real events in his database. For such purposes, an event extraction system would consist of two steps: i) It extracts events in a structured, semantically enriched representation and ii) determines based on linguistic cues whether the event is real or not.

Research on real event detection has been very limited so far. In this paper, we present an approach to define events and real events in a setting as described. Since no suitable gold standard for evaluating a real event detection system exists, we present our setting of creating one using crowdsourcing. Preliminary results regarding this gold standard are presented, as well as challenges which we came across.

The remainder of this paper is organized as follows: First we present definitions of event detection in Section 2, before considering definitions of real event detection in Section 3. After discussing our setup of creating a gold standard for real event detection in Section 4, we conclude in Section 5.

2 General Event Definitions

2.1 Event Definitions in Use

We can distinguish between the following classes of event representation (see also Fig. 1 for examples):

1. *Something happened*: In this event representation, events are only roughly covered. There are no types and deeper meanings gathered, only what topic the document/sentence is about. This topic is often characterized by the words occurring in the document (bag-of-words model) and/or by the set of recognized named entities.
2. *This happened*: For this representation, the event type of the event is detected. The event type can be quite generic such as *earthquake*. The number of events which can be detected is often very limited. Events may
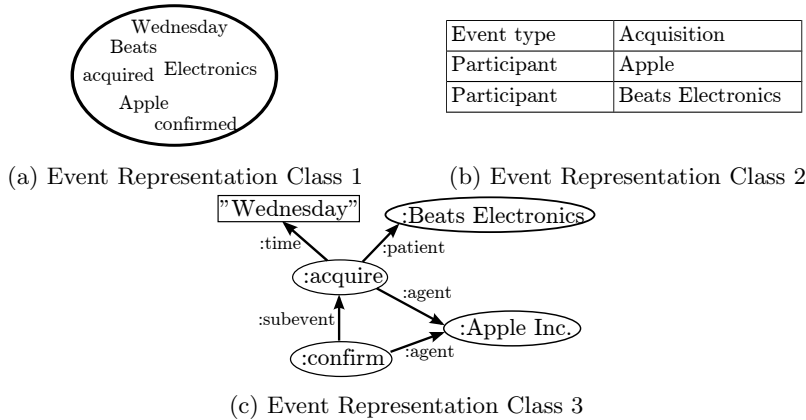
(a) Event Representation Class 1   (b) Event Representation Class 2

(c) Event Representation Class 3

Fig. 1: Examples of event representations for the different event representation classes regarding the example sentence "Apple confirmed that it acquired Beats Electronics on Wednesday."

have attributes or slots which are pre-defined for the single event types. Instead of predefined entity types such as *earthquake* or *accident* sometimes only the entity types PER, LOC, ORG, and MISC are used.

3. *This happened to these objects in this way*: If we use this representation format, we have a deeper understanding in the actual event. Events of this class are quite specific and include not only specific actions, but also participants, and maybe time, place, and manner of the action. Often linguistic theories such as Semantic Role Labeling provide the basis for event representations of this class.

Related work using event definitions corresponding to the first event representation class do not define events at all [1,2,3,4,5]. This is due to the fact that here it must be only known *that* something happened (something that is, for instance, different to what has been seen so far), but not *what*. Events do not need to be represented on its own; instead, events are indirectly represented by the document in which they are expressed. Documents are compared against each other, either by using the bag-of-words model [2,3,4] or in addition by taking detected named entities (with the classical entity types PER, LOC, ORG, MISC) into account [1,5].

Approaches using the second event definition have in common that coarse-grained events such as accidents or earthquakes are represented. Each event has therefore an event type. Property-value-pairs can be assigned to the events, whereas the assignable properties are pre-defined for all event types. Often templates are used for storing the information about events [6].

In case of event representations of the third kind, structural representations of fine-grained events are extracted from text – here, typically from single sentences or clauses. Research based on this event class usually does not

introduce a new definition of events, but instead either uses linguistic definitions of events where events consist of happenings with agents, locations, time, etc. [7,8,9] or abstracts from it to a certain, but limited extend [10]. Bejan [10] characterizes an event as a happening at a given location and in a specific time interval. Each event has semantic relations to agents, to a location, time, etc. as parts of the event. These are the semantic/thematic roles of an event in the linguistic understanding. Events can contain several sub-events. Events of an event scenario (as higher-order structure) are connected by event relations. An example is the *cause* relation where one event causes another event. Xie et al. [7] propose two approaches which are based on Semantic Frames – constructed by the tool *SEMAFOR*. Also, Wang et al. [8] use semantic parsing which is based on PropBank in order to represent events. Yeh et al. [9] regard events as similar to frames in FrameNet. Each event encodes knowledge about the participants, where (and when) the event occurred and the events which are caused by this event. A *buy* event, for instance, is about the object bought, the donor, and the recipient.

2.2 Event Definition

In this paper we focus on the detection and semantically-structured representation of real events of the third-mentioned event class, which is the most expensive one. More specifically, an event in our scenario is characterized by

- specific participants (agents or objects)
- situations (events or states) which are described within the event
- taking place at a specific place and/or time
- being not a state.

States are hereby defined as lasting for an indefinite period of time and which are not really observable. Given the example sentence 2 in Section 1 we can extract two events from it: i) The event that Apple confirmed something (which is an event itself) and ii) the event that Apple acquired Beats Electronics.

Fig. 1c shows how these events can be represented as a semantically-structured graph. Hereby, Event ii) can either be part of Event i) (as depicted in the figure) or be stored as a separate graph. Nodes in each event graph can be either predicate nodes (representing actions), entity nodes (representing participants), or literal nodes (representing the time, etc.). Predicate and entity nodes can be linked to entries in knowledge bases such as DBpedia (for entities) and WordNet (for predicates). This enables having unique identifiers for resources and to resolve ambiguities. The edges in these event graphs arise from the semantic roles assigned by a Semantic Role Labeling tool. In the depicted figure, the semantic roles are grounded as RDF predicates.

3 Real Event Detection

3.1 Definitions of Real Events

We define *real event detection* as the task of determining whether a given event expressed in text is real. Real events are events according to the definition in Section 2.2 and have already happened or are happening. Thus, the definition of events is extended by this aspect. We can split the task of real event detection therefore into two subtasks: 1. Determining if the situation described in the text is about an event according to our definition. 2. Determining if the event already happened or is currently happening.

Regarding the first subtask, we can refer to two areas of linguistic work: i) The distinction of events from states, and ii) the identification of factuality of events. In the following, we amplify these two areas with respect to our goal of real event detection. We hereby use the term *situation* as a generic concept which encompasses both event and state (cf. [11]).

Ad i) The classification of situations can be traced back to Aristotle who distinguished between verbs that have a defined end or result, and others that do not [12]. Vendler [13] distinguished situations into four *aspectual classes* (also called *aktionsarten*) and performed empirical experiments. The aspectual classes are based on the temporal structure of events. These classes are namely: *state*, *activity*, *accomplishment*, and *achievement*. A state is something in which an entity remains for a longer, often unspecified period of time (e.g., "Jack knows the answer"). The three other classes in the aspectual classification cover different types of events in the narrower sense. An event is characterized as something which happens or occurs in a definite time interval or at a specific point in time. It often comes along with predicates such as "write", "push", etc. An event usually causes some state change.

To determine which aspectual class a given situation belongs to, we can differ between telic, dynamic, and durative situations (see Table 1). Telic situations always have a culmination point beyond which the situation cannot continue. Dynamic situations consist of internal sub-events which change over time and are, hence, intrinsically heterogeneous. For instance, walking consists of several alternating subevents. Durative situations (e.g., eating) last for a specifiable period in time and are not punctual.

Table 1: Vendler's four-way distinction between verbs based on their aspectual features [13].

Class	Telic	Dynamic	Durative
state	-	-	✓
activity	-	✓	✓
accomplishment	✓	✓	✓
achievement	✓	✓	-

In our case we want to distinguish events from states. But how can we determine which aspectual class holds for a given situation? For Vendler [13] and others who worked on top of his theories it became apparent that it is not trivial to determine the class automatically. See [13,11,14,15,16] for more details on linguistic rules for that purpose.

Moens and Steedman [14] propose another classification of situations. Here, situations are also either states or events. Events are sub-classified by two dimensions: 1. Events are either atomic or durative events. 2. Entities of events are in a consequent state or not. We refer to [14] for more information.

Ad ii) Other researchers have focused on determining the factuality of events, i.e. to recognize whether events are presented in the sentences as corresponding to real situations in the world, as situations that have not happened, or as situations of uncertain status. The focus is, hence, the trustfulness of events in text. Factuality can be characterized by two dimensions: Polarity and epistemic modality. Polarity – more concrete: polarity on actuality and not subjective polarity – is a discrete category and can be either positive or negative. Epistemic modality, in contrast, expresses the speaker's degree of commitment to the truth of a proposition [17]. It ranges from uncertain (also called "possible") to absolutely certain (also called "necessary"). According to Horn [18], modality is a continuous category. Sauri [19] spans the factuality values space from positive, negative, to unknown for the polarity dimension, and certain, probable, possible, to unknown for the modality dimension. *Unknown* is true for cases of uncommitment. In this way, a tuple of polarity value and epistemic modality value states the factuality of the event.

How is factuality expressed in the text? This is done by lexical markers as well as syntactic markers. Lexical modal markers are modal auxiliaries (e.g., "could", "may", "must"), as well as clausal/sentential adverbial modifiers (e.g., "maybe", "likely", "possibly"). Examples of lexical polarity markers are adverbs (e.g., "not", "until"), quantifiers (e.g., "no", "none"), and pronouns (e.g., "nobody"). Syntactic constructs are necessary to consider since often one clause is embedded in another. Considerable are in this context especially relative clauses and that-clauses as in the example sentences.

What are the challenges to determine the factuality? Factuality markers interact with each other. The local modality and polarity operators (e.g., of the current clause) are therefore not enough. Instead, a global consideration is necessary. For instance, in case of that-clauses, the factuality of the inner event is dependent on the factuality of the outer event. Furthermore, what makes the factuality much more complex is the fact that the source of an event is often not only the author. These additional sources are introduced by means of predicates of reporting (such as "say" or "tell"), knowledge and opinion (such as "believe", "know"), psychological reaction (such as "regret"), etc. Sauri and Pustejovsky [19] calls these predicates due to their role *Source Introducing Predicates (SIPs)*. The difficulty is that the status of the other sources often differs from the author. The reader does not have direct access to the factual assessment of these other sources. In the sentence, "The Guardian wrote that the G-7 leaders pretended everything was OK in Russia's economy.", the reader cannot assess directly the "frame of mind" of The Guardian with respect to the factuality of the event of "pretended". However, the factuality assessment has to be relative to the relevant sources.

3.2 Requirements of a Gold Standard for Real Event Detection

According to our event definition in Section 2.2 and the additional aspect of factuality addressed in Section 3.1 we can list the following requirements a gold standard dataset for the evaluation of a real event detection system must fulfill:

1. Each mention of an action within an event (e.g., "wrote") is annotated.
2. There is a distinction between events and states, so that all events in the strict sense are annotated.
3. There is no restriction to specific event types.
4. The factuality of the event is annotated (being positive or negative).
5. All participants and participating objects are annotated.
6. All participants and participating objects are linked to prevalent knowledge bases.
7. Subevents of events are annotated and linked.
8. Mentions of place and time of each event are annotated.

This gold standard is also suitable when it comes to extracting real events according to the Event Representation Classes 1 and 2 (see Section 2.1). In these cases, the information about the structural representation of events can be neglected. Additional filtering can achieve that only events of specific types such as accidents are detected.

3.3 Datasets for Real Event Detection

In the following, we review existing corpora where event factuality was annotated to some degree.

The *Multi-Perspective Question Answering (MPQA) corpus* [20] provides news articles annotated for opinions and other private states such as beliefs or thoughts. It was designed for subjectivity and sentiment research and does not provide any structured representation of (real) events. At most, it might be applicable as negative corpus in a scenario where situations written in text are approved to be not real events.

The *Penn Discourse TreeBank (PDTB)* [21] is a corpus where discourse connectives are annotated along with their arguments (e.g. $arg1 "– even though" $arg2). On top of the original annotation scheme, an extended annotation scheme was released for marking the attribution of abstract objects such as propositions, facts and eventualities associated with discourse relations and their arguments annotated in the PDTB. The events described in the arguments are, however, not transformed into a structured event representation.

TimeBank 1.2 [22] is a corpus which was annotated with TimeML [23]. TimeML is a language for representing temporal and event information. TimeBank is suitable for event factuality learning since it uses grammar markers as well as annotations of predicates. Events are classified into *occurrence*, *state*, *reporting*, *immediate-action*, *immediate-state*, *aspectual*, and *perception*. TimeBank does not contain a structured event representation where all

participating objects are annotated. In addition, the event definition is somehow different to our proposed definition: A huge fraction (25,7%) of phrases annotated as events are not verbs, but nouns, adjectives, etc. Not all phrases that should be regarded as event predicates are annotated.

FactBank [19] is a corpus which was built on top of TimeBank and a subset of the documents in the AQUAINT TimeML Corpus (A-TimeML Corpus). It comes along with annotations of explicitly factual information about events. FactBank has the same obstacles as TimeBank.

ACE 2005 [24] from the *Automatic Content Extraction (ACE) technology evaluation* is a dataset dedicated to the detection of events in text. The task was limited to the detection of specific event types which are: *Life*, *Movement*, *Transaction*, *Business*, *Conflict*, *Contact*, *Personnel*, and *Justice*. Each type has one to 13 subtypes so that each event is assigned to one main event type and one subtype of it. The limitation to these event types is the main obstacle why ACE 2005 cannot be used in our setting directly. Four attributes are attached to each annotated event: *Modality*, *Polarity*, *Genericity*, and *Tense*. In accordance with the event type, specific slots (*argument roles* called here; such as entities, values, and times) can be assigned. ACE entities are categorized in specific classes (namely, Person, Organization, Location, Geo-political entity, Facility, Vehicle, and Weapon) and their subclasses, but are not linked to any knowledge base.

In summary, we can state that none of the mentioned corpora contains semantically-structured representations of events to the extent it is needed to evaluate a real event detection system where events are defined as in Section 2.2. Thus, in the following section we provide experiments on how to build a gold standard which fulfills all our requirements.

4   Experiments for Building a Gold Standard Dataset

Very first crowdsourcing experiments revealed that letting users annotate real events as described in Section 3.2 at once is too complex for any crowdsourcing job. Therefore, we arranged subtasks where the following questions are answered separately for each event:

1. Which are the actions/predicates inducing a real event?
2. Which are the participating objects?
3. What is the time and place?
4. Which sub-events are contained?

In the following we present our approach regarding the first subtask, namely identifying *real events* and naming the central predicates of them. We performed two crowdsourcing jobs which differ in their methodology.[1]

Run 1 The crowd was asked to read a given sentence, to look for real events (as defined above), and to enter the action verbs of these events as written in the sentence.

---

[1] The crowdsourcing job descriptions and evaluation data is available online at `http://www.aifb.kit.edu/web/Toward_Real_Event_Detection`

Run 1: "Find real actions"
187 sentences, 8 test questions, 12¢ per task,
5 users per judgment

Our gold
standard:
205 verbs inducing
real events

224 verbs judged by
crowd as inducing
real events

152/224 (67.9%) of
verbs judged as
inducing real events
are correct

Run 2: "Find observable and non-observable predicates"
187 sentences, 9 test questions, 12¢ per task,
5 users per judgment

Our gold standard:
205 action verbs

354 observable
predicates

185 non-observable
predicates

133/205 (64.9%) of
predicates judged as
observable are corect

285/334 (85.3%) of predicates
judged as non-observable
are correct

205 verbs classified
by crowd as observable

334 predicates classified by crowd
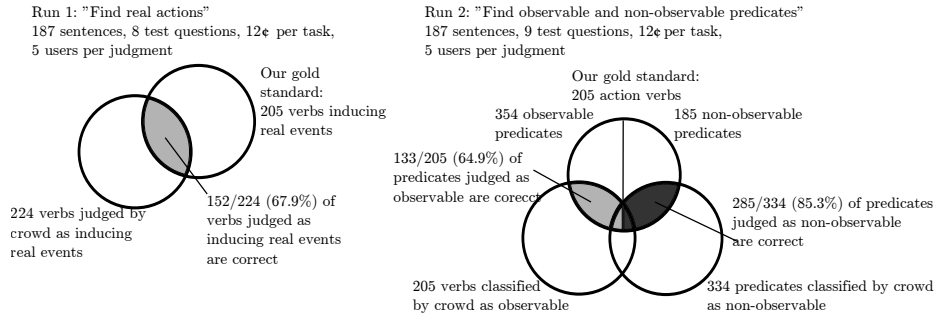as non-observable

Fig. 2: Results of two crowdsourcing runs where the predicates of real events were annotated in English sentences. In both runs, the confidence value of the answers had to be above 0.5 in order to be considered.

Run 2 For this second run, the crowd was asked to read each given sentence, look for all verbs, and categorize them into either *observable* or *not-observable*.

Observable events/facts were defined as follows:[2] An observable fact can be an occurrence (e.g., "arrive", "destroy"), a reporting (e.g., "report"), or an immediate action (e.g., "approve"). Observable facts are characterized by the fact that they could be observed or confirmed by third persons directly (e.g., in case of "say") or indirectly (e.g., in case of "confirm"). Non-observable facts describe states which characterize persons or objects, but which are not observable by other persons than the persons involved. Such non-observable facts are states which last for an indefinite/unspecified period of time (e.g., "be happy"), immediate states (e.g., "believe", "worried"), aspects (e.g., "start", "continue"), or perceptions (e.g., "feel"). The categorization into observable vs. non-observable facts is here done independently of the fact whether the event has happened (or the state is) for sure or not. The categorization into the past/presence or future is performed in a separate crowdsourcing task.

As dataset we used all first sentences of news articles which were published on one day (2014/05/28) by the news agency Bloomberg and where the news articles contained some information about Apple Inc. In total we manually annotated 187 sentences to assess the performance of our crowdsourcing tasks. Crowd sourcing was performed on the platform Crowdflower.[3] In Run 1 (Run 2), users had to answer 8 (9) quiz test questions before entering the actual task. In both runs, users got 12 cent per task consisting of 4 questions each. For each question we gained results from 5 users and took the answers where there was an inter-rater agreement of at least 50%.

The results of our crowdsourcing annotation experiments are summarized in Fig. 2. It became apparent that completing the crowdsourcing tasks requires high cognitive efforts in comparison to other crowdsourcing tasks. A considerable amount of users did not pass the test questions at the beginning. Even if we

---

[2] The definition is based on the TimeBank annotation guidelines.

[3] http://crowdflower.com

9

admit only users who worked on our job in the past sufficiently well, creating a big annotated corpus is tricky. As Run 2 shows, already the distinction between observable events, i.e. events showing up in the real world, and not-observable events is hard to perform. Although we put much effort in refining the task descriptions the question arises whether a better approach to annotating the factuality of events is achievable.

5   Conclusions

If events are extracted from text in a fine-grained manner, huge amounts of events are gathered, but only a fraction of them represent real events and, hence, are worthwhile to process further on. In this paper, we gave an overview of existing linguistic work about the detection of *real events*. In order to evaluate a proposed system which extracts semantically-structured, real events from text, we defined requirements and proposed a methodology to create a gold standard dataset. Preliminary experiments with crowdsourcing showed that the annotation of text with factual information is non-trivial. Still, we believe that the creation of such a dataset is necessary for many event detection systems in the future.

References

1. Gabrilovich, E., Dumais, S., Horvitz, E.: Newsjunkie: providing personalized newsfeeds via analysis of information novelty. WWW '04, New York, NY, USA, ACM (2004) 482–490
2. Karkali, M., Rousseau, F., Ntoulas, A., Vazirgiannis, M.: Efficient Online Novelty Detection in News Streams. In Lin, X., et al., eds.: Web Information Systems Engineering – WISE 2013. Springer Berlin Heidelberg (2013) 57–71
3. Zhang, Y., Callan, J., Minka, T.: Novelty and Redundancy Detection in Adaptive Filtering. SIGIR '02, New York, NY, USA, ACM (2002) 81–88
4. Zhang, K., Zi, J., Wu, L.G.: New Event Detection Based on Indexing-tree and Named Entity. SIGIR '07, New York, NY, USA, ACM (2007) 215–222
5. Li, X., Croft, W.B.: Novelty Detection Based on Sentence Level Patterns. CIKM '05, New York, NY, USA, ACM (2005) 744–751
6. Kosmerlj, A., Belyaeva, J., Leban, G., Fortuna, B., Grobelnik, M.: Crowdsourcing Event Extraction. In: NewsKDD – Workshop on Data Science for News Publishing at KDD 2014. (2014)
7. Xie, B., Passonneau, R.J., Wu, L., Creamer, G.G.: Semantic Frames to Predict Stock Price Movement. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics. (2013) 873–883
8. Wang, D., Li, T., Zhu, S., Ding, C.: Multi-document Summarization via Sentence-level Semantic Analysis and Symmetric Matrix Factorization. SIGIR '08, New York, NY, USA, ACM (2008) 307–314
9. Yeh, P.Z., Puri, C.A., Kass, A.: A Knowledge Based Approach for Capturing Rich Semantic Representations from Text for Intelligent Systems. Int. J. Adv. Intell. Paradigms **2**(1) (November 2010) 33–48
10. Bejan, C.A.: Learning event structures from text. PhD thesis, The University of Texas at Dallas (2009)

11. Bach, E.: The Algebra of Events. Linguistics and Philosophy (1986) 5–16
12. Dowty, D.R.: Word Meaning and Montague Grammar: the semantics of verbs and times in generative semantics and in Montague's PTQ. Reidel (1979)
13. Vendler, Z.: Linguistics in Philosophy. Cornell University Press (1967)
14. Moens, M., Steedman, M.: Temporal Ontology and Temporal Reference. Computational Linguistics **28**(3) (1988) 15–28
15. Pustejovsky, J.: The syntax of event structure. Cognition **41** (1991) 47–81
16. Dorr, B.J., Olsen, M.B.: Deriving Verbal and Compositonal Lexical Aspect for NLP Applications. Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics (ACL) (1997) 151–158
17. Palmer, F.: Mood an Modality. Cambridge University Press (1986)
18. Horn, L.: A Natural History of Negation. University of Chicago Press (1989)
19. Sauri, R., Pustejovsky, J.: From structure to interpretation: A double-layered annotation for event factuality. Proceedings of the Second Linguistic Annotation Workshop (2008)
20. Wiebe, J., Wilson, T., Cardie, C.: Annotating expressions of opinions and emotions in language. Language Resources and Evaluation **39**(2) (2005) 165–210
21. Miltsakaki, E., Prasad, R., Joshi, A., Webber, B.: The Penn Discourse Treebank. Proceedings of LREC 2004 (2004)
22. Pustejovsky, J., et al.: The TIMEBANK Corpus. Proceedings of Corpus Linguistics 2003 (2003) 647–656
23. Pustejovsky, J., Knippen, R., Littman, J., Saurí, R.: Temporal and event information in natural language text. Language Resources and Evaluation **39**(2) (2005) 123–164
24. Walker, C., Strassel, S., Medero, J., Maeda, K.: ACE 2005 Multilingual Training Corpus LDC2006T06 (2006)