Towards Multilingual Event Extraction Evaluation: A Case Study for the Czech Language

Josef Steinberger
DCSE, NTIS Centre, FAV
University of West Bohemia
Univerzini 8, 306 14 Plzen
Czech Republic
jstein@kiv.zcu.cz

Hristo Tanev
Joint Research Centre
European Commission
via Fermi 2749, Ispra
Italy
hristo.tanev@jrc.ec.europa.eu

Abstract

This paper presents a multilingual corpus of news, annotated with event metadata information. The events in our corpus are from the domain of violence, natural and man made disasters. The main goal of the corpus is automatic evaluation of event detection and extraction systems in different languages. As a use case, we take a rule-based event extraction system, extend it to cover a new language, Czech in our case, and evaluate it on the corpus. We explain what needs to be done to cover a new language, especially learning domain-specific dictionaries and event extraction patterns. The evaluation of the Czech system can be viewed as a starting point for further research into the evaluation of multilingual event extraction systems, which is an important stage during the development of such systems. The comparison of the performance for the Czech and English systems indicates the importance for multilingual event extraction evaluation.

1 Introduction

The quantity of information on Internet has reached a critical point. Simple keyword indexing cannot satisfy any more the need for fast and accurate access to this information ocean. In this light, the development of effective methods for information extraction are of particular importance. In this paper we will discuss issues related to automatic event metadata extraction. Mainstream media and part of the social media are event-oriented, therefore development of methods for accurate identification, classification and extraction of metadata about events is of particular importance. Noteworthy, crisis events, such as natural, man-made disasters, crime and armed conflicts are the most frequent types of events, described in online news and often referred to in social media.

Due to the complexity of the event extraction task, preparing a gold standard and evaluation of event extraction systems is not straightforward. Event annotation can be done in many different ways. Different taxonomies of event types can be used, as well as different event properties may be annotated. Moreover, one cannot give a single accuracy number, which characterizes an event extraction system performance. Rather than that, the accuracy for the extraction of each event property is measured separately. Even measuring the overlap between the gold standard and the output, produced by a system, can be done in different ways. Similarly, evaluating the similarity between event types, such as *bombing* and *terrorist attack* requires investigating into the nature of the events and the goals of the evaluation schema.

In this paper we make a small step into the infinite field of problems and solutions which the evaluation of event extraction system poses in front of the researchers in the field of information extraction.

We propose an event annotation model which consists of a taxonomy for classification of crisis events, as well as a template model with their most important slots. Then, we present a multilingual corpus annotated according to this model. Finally, we describe semi-automatic acquisition of lingistic resources for event extraction in Czech language. We plug these resources into a state-of-the-art event metadata extraction system and then we evaluate the performance of the system, using the annotated event corpus. Clearly, our solution is just an island in the sea of possible annotation and evaluation schemas.

The rest of the paper is structured as follows: Section 2 reports about related work. Section 3 describes the event annotation model. Section 4 is about the creation of the corpus. Section 5 de-

Proceedings of Recent Advances in Natural Language Processing, pages 627–635,
Hissar, Bulgaria, Sep 7–9 2015.

scribes the creation of event extraction resources for the Czech language. Finally, we discuss the results of our case study evaluation.

2 Related work

Recently, there is a significant amount of work, regarding automatic event detection from traditional and social media. However, few systems extract event metadata. Similarly, there are not many corpora, annotated with such metadata. In (Kim et al., 2008) annotation of event corpus from the biomedical domain is presented. The annotation is carried out according to event ontology, which partially overlaps with the GENIA ontology. A similar corpus is presented also in (Vincze et al., 2008).

FactBank (Saurí and Pustejovsky, 2009) is a corpus annotated with factuality information about news events. The GDELT database (Leetaru and Schrodt, 2013) contains automatically extracted metadata for politically-motivated events.

Most of the existing corpora are in English. The only multilingual corpus annotated with event metadata was created in the framework of the News Reader project (NewsReader et al., 2014). However, the corpus was annotated automatically in this project. Most of other event corpora are in the biomedical domain and few represent the domain of generic news discussed in the media.

Regarding automatic acquisition of event extraction resources, one of the first system for learning of event extraction lexicon and patterns is AutoSlog (Riloff and others, 1993). Other systems are presented in (Yangarber et al., 2000) and (Du and Yangarber, 2015). The problem with these and the other learning systems is that they rely on language-specific resources and consequently work only for the English language.

There are different event-extraction systems, presented in the literature: the KEDS/TABARI project (Schrodt, 2001), whose purpose is automatic detection and extraction of event metadata for political events, the Proteus system (Yangarber and Grishman, 1998) and others. There are two main classification schemas for political events: CAMEO (Gerner et al., 2002), developed inside the KEDS project and IDEA (Bond et al., 2003).

3 Event annotation model

The model we use for annotating events consists of two parts: a taxonomy of event classes and a template, whose slots represent the properties of the events. As a matter of fact, both parts of this model can be united into an ontology, where the taxonomy represents the is-a relations and the template slots are represented as ontological properties.

3.1 Event taxonomy

The event taxonomy is inspired by the one used in the NEXUS event extraction system (Tanev et al., 2008). We tried to create classes which correspond to the main crisis event types, mentioned in the news and social media. Definitely, more detailed event classification can be done. On the other hand, going for a very fine grained classification, would result in annotations which are difficult to be matched by event extraction systems. The crisis events in our taxonomy fall mainly in one of the two big top classes: *Disaster* and *Violence-related event*. The third group of events modeled in our taxonomy is related to the violent events: the class *Juridical event*. Juridical event is the smallest cluster, it contains arrests, trials, detentions, executions and raids of security forces. The category *Violence-related event* encompasses mainly events in which there is violence or attempt for violence against people, such as armed conflicts, crime, terrorism etc., as well as events, which can turn violent, such as demonstrations and strikes. We consider also the class *Sabotage* to be under *Violence-related event*, even if it does not include violence against people, it implies intentional damage of infrastructure and machines. Similarly, *Asylum/Fleeing a country for political reason* is considered to be *Violence-related*, since when people flee a country for political reason, their life and liberties are most likely threatened.

The category *Disaster* has two main sub classes - *Natural disaster* and *Man made disaster*. Natural disasters are storms, quakes, floodings, forest fires and others. Man made disasters are divided in extraordinary, like industrial accidents and explosions, as well as ordinary ones, which include traffic and aircraft accidents.

The category *Violence* is divided in three subcategories: *Politcally-motivated violence*, which includes differen types of armed conflicts and terrorist attacks, *Crime*, and *Socio-political event*, which includes different forms of protest actions: demonstrations, riots, sabotages, etc.

The event classes in our taxonomy reflect the nature of the event - its dynamics and the means,

Violence-related event (upper level subclasses)		
	Politically motivated	
		Political execution
		Armed conflict
		Terrorist attack
		Anti-terrorist operation
		Assassination
		Kidnapping/Hostage taking (political)
		Hostage release (political)
		Military movements
		Asylum/Fleeing a country for political reason
	Criminal	
		Robbery
		Kidnapping/Hostage taking (criminal)
		Hostage release (criminal)
		Shooting (criminal)
		Stabbing
		Abusing/offending people
		Physical attack
		Drug trade
		Vandalism
		Arson/Firebombing
		Piracy
		Cyber attack
		Prison break
	Socio-political	
		Boycott/Strike
		Public demonstration
		Riot
		Sabotage
		Mutiny
	Juridical	
		Arrest
		Charging
		Trial
		Execution
		Raid

Table 1: A part of the event taxonomy - violence-related events.

which were used, but also the motivation behind it and its context. While some event types may look similar, like *Shooting* as a subtype of *Armed conflict* and *Shooting* as a criminal event, in our taxonomy they are two different event classes, since the context and the motivation behind these actions are different. In the armed conflict shooting, the action is carried out by troops which serve their country, while in the criminal shooting, the main actors are criminals, whose motivation is to rob, to defend themselves from the police, etc. In the same way, we make difference between politically-motivated executions, executions by terrorists, and normal executions ordered by the court, without political motivations. Consideration of the motivation and the context is important, since they can give birth to different participants, means in use and consequences from the events. On the other hand, it is difficult for an event extraction system to draw the line between similar event classes. In order to overcome this last issue, during our experiments, we allowed for mapping of one class of the event extraction output to several classes from our model. For example, the event extraction system type *Execution* is considered a correct match for any of the execution classes used in our model.

Clearly, a taxonomy is not a complete knowledge representation model, since it does not represent relations other than *is-a* relation between event classes. In order to have more comprehensive knowledge-representation schema, the event taxonomy should be transformed into ontology. The structure of a crisis event is usually complicated: One event encompasses many subevents, which are related via causal relations. For example, event of type *Piracy* may include as subevents *Shooting* and *Kidnapping/Hostage taking*, which on its own may trigger event *Raid* by security forces to free the hijacked ship which can trigger event of type *HostageRelease*. In order to model this type of relations, the event ontology should encompass different types of relations, such as *causes* and *subevent-of*. The upper level violence-related classes of our taxonomy are shown in table 1.

3.2 Event properties

The properties of the event types in our model are represented through a unified template, which features the union of the properties of all event types. This is a simplification, since in reality the three big event classes: *Violent event*, *Natural disaster* and *Juridical event* have different properties. Properties related to the participants of the events: dead, wounded, kidnapped, arrested, etc. are actually pairs - specification of the participants, e.g. *five people* and their number, e.g. *5*. The properties template is shown on table 2.

Property
Time
Location
Dead count and specification
Missing count and specification
Wounded count and specification
Perpetrator count and specification
Kidnapped count and specification
Arrested count and specification
Weapons used

Table 2: The event properties template.

In addition, the model includes quantifiers where it is applicable. Examples:

- at least 20 people died (or not more than 20) = 20-
- over 20 dead = 20+
- hudreds of injured = 100x
- around 100 people = 100˜

4 Creating multilingual corpus with annotated events

Annotating articles about same events in multiple languages gives us a possibility to evaluate a multilingual event extraction system and the results are then directly comparable among languages. By comparing the results among languages, one could analyse how different language properties affect the quality of template extraction. As we want to make our corpus available to the community, we selected Wikinews as the source of event-related articles, since its licence allows us to share the news.

As a first application of the multilingual corpus, we wanted to evaluate our system in a newly supported language, namely Czech. Because of that, our starting page was the Czech Wikinews site. We manually selected event-related articles. We selected only articles which were available for more languages (visible in the left bar of the Wikinews site).

As the coverage of Czech Wikinews is not that high we included articles from the Multiling'13 corpus[1]. For now, we included only Czech, English and Spanish variants from the Multiling corpus.

An example of an event topic with English and Czech data and annotation can be found in table 3.

There are 109 topics in the corpus. Altogether, it includes 344 articles in 14 langauges. Distribution of between languages is given in table 4.

Language	Articles
cs	109
en	96
es	39
fr	34
de	18
it	13
ru	11
pt	6
pl	6
bg	3
ar	3
fi	2
no	1
gr	1

Table 4: Topics per language counts.

Most of the annotated news articles were available both in Czech and English languages.

4.1 Statistics about event roles

Regarding the event slots, which are represented in the corpus, the predominant event-specific role

[1] A corpus created by the summarization community: http://multiling.iit.demokritos.gr/pages/view/662/multiling-2013

topic metadata	event type	violence - criminal - shooting
	date	September 23, 2008
en	article title	School shooting in Kauhajoki, Finland kills eleven
	article perex	At approximately 11:00 a.m. Central European Summer Time, a man in his twenties entered the Kauhajoen vocational school in Kauhajoki, Finland with a gun and began to open fire, killing 11 people.
	perpetrator count	1
	perpetrator specification	a man in his twenties
	victim count	11
	victim specification	eleven; 11 people
cs	article title	Střelba ve finské škole
	article perex	Ve finském městě Kauhajoki na severozápadu země došlo ke střelbě. Na zdejší ekonomické škole jeden ze studentů vypálil po svých spolužácích, policie se obává, že incident si vyžádal několik obětí. Útočník se nachází ještě stále v budově školy.
	perpetrator count	1
	perpetrator specification	jeden ze studentů
	victim count	—
	victim specification	svých spolužácích

Table 3: An example of an annotated event topic.

for all the languages was found to be: *victim*, which includes dead, injured and kidnapped people. There are around 600 victims mentions (usually they are mentioned in both title and the first paragraph). For Czech only we have found more than 110 victim mentions. 15 weapons mentioned - too little to provide a proper basis for evaluation; 73 perpetrator mentions; 64 arrested people mentions; 37 sentenced people mentions.

We plan to extend the corpus with articles from English Wikinews and translate them to other languages. [2]

5 Event extraction system and semi automatic acquisition of dictionaries for it

We created Czech dictionaries and a cascaded grammar for analysis of crisis events, as well as boolean combination of keywords for recognition of event types, which was then used in the multilingual event extraction system – NEXUS (Tanev et al., 2008).

5.1 NEXUS

NEXUS is a multilingual rule-based event extraction system, developed at the Joint Research Centre, EC, which extracts event metainformation from online news in several languages. NEXUS essentially performs two types of tasks: first, using semantic grammar rules, backed up by domain specific dictionaries, it identifies in the text a set of noun phrases, which are assigned certain semantic roles. For example, in the text *The prime minister was kidnapped by masked gunmen*, the system will extract *the prime minister* as kidnapped victim and *masked gunmen* as perpetrators. Moreover, the system classifies the events, based on combinations of keywords. In the previously mentioned text *gunmen* and *kidnapped* will trigger the event type *kidnapping*. In order to plug in a new language in our event extraction system, we implement new domain specific dictionaries, as well as keyword combinations for event classifications. The grammars in use are also changed, although between similar languages, the change is small. This is due to the fact that the linguistic knowledge is mostly encoded in the domain-specific dictionaries: for example, for English we have all the possible patterns for *kill*: *was killed*, *have been killed*, *murdered*, *murdered by*, etc. This solution puts a stress on the domain-specific dictionaries, which are usually large and therefore we use semi-automatic methods, in order to learn them.

5.2 Learning dictionaries and linguistic patterns

The dictionaries used by NEXUS are developed following a semi-automatic procedure described in (Tanev et al., 2009). For each dictionary, the following steps are performed:

1. The user provides manually a seed set of entries

[2] The corpus will be available for download at http://nlp.kiv.zcu.cz.

2. It runs the LexiClass multilingual dictionary expansion tool which suggests more words and multiwords, distributionally similar to the seed set, ordered by their l similarity

3. The expanded dictionary is cleaned manually by looking at its top elements (which are most similar to the seed set)

For example, if the seed set are the English words: *soldiers*, *policemen*, *security forces*, the top elements from the expanded dictionary are *troops*, *civilians*, *officers*, *personnel*, *militants*, *peacekeepers*. A better description of the algorithm is provided in (Tanev et al., 2009) , where the precision of the algorithm for Portuguese was found to be 51% and for Spanish 71%. The algorithm is described also in (Tanev and Zavarella, 2013). Following the above-mentioned algorithm. we created the following Czech dictionaries, which are used by NEXUS: dictionary of noun phrases, referring to people and a dictionary of modifiers of these noun phrases. Moreover, we manually created a list of Czech numerals. These three resources were used in the first layer of the event-slot extraction grammar, which is responsible for detection of references to people.

The second layer of the grammar detects patterns, which co-occur with the person references, found on the first level. These patterns express different semantic roles which people take in the crisis event contexts: *dead* or *wounded victim*, *perpetrator*, etc. In order to discover the patterns, first we used the previously-described procedure to learn verbs and nouns, which introduce the considered semantic roles. Then, we searched automatically in a corpus co-occurrence patterns between these role-expressing words and references to people. A detailed description of the algorithm is provided in (Tanev et al., 2009). As an example, the output of the algorithm for English language for the semantic role *dead victim* will be patterns like *killed [PERSON]*, *[PERSON] was murdered*, etc.

Using these algorithms, we acquired 270 person-referring nouns, 600 person modifiers and 250 patterns for dead, wounded, arrested, kidnapped and perpetrators.

5.3 Providing keyword combinations for event type detection

In our event extraction system, the event class, e.g. *armed conflict*, *robbery*, etc., are detected through boolean combinations of keywords. We created these keyword combination mostly manually, using in some cases the LexiCLass system.

6 Evaluation

6.1 Methodology

We have run the NEXUS event extraction system on the news from our annotated corpus and evaluated the results. As NEXUS can currently detect only part of the event types in the corpus, we run the system only on the events, whose types are detectable by the system.

It is an important issue in the event extraction evaluation that the annotated event types and the detected by the system can differ in their specificity. For example, if the annotation is *suicide bombing* and the system says *terrorist attack*, is that a correct match? Probably, it is appropriate to consider this as a correct hit. However, if the system says that the type is *terrorist executing hostages* and the annotation is *suicide bombing*, then the match should not be considered correct.

In our experiments, we adopted a simple solution, which even if not perfect, provides a basis for evaluation of the event class detection. We simply mapped both annotated event types and the detected ones to event types, which were found to be specific enough, but not too specific, i.e. their taxonomy depth is somewhere in the middle. For example, all the daughter nodes of *socio-political* were mapped to this event type, the same for *terrorist attack.* Apart from the event type, we evaluated the following event participant properties:

- dead victim specification
- dead victim count
- wounded victim specification
- wounded victim count
- perpetrator spectification
- arrested spectification

Another problem in evaluating the performance of an event extraction system is the difference in the span of the annotated and detected slot fillers. For example, an event extraction system may detect as victims *Chinese*, while the annotation may be seven Chinese businessmen from Beijing. Our solution to this problem was that we used partial matching, i.e. if the system finds a part of the specification, it counts as a correct match. The obvious disadvantage is that we do not evaluate the completeness of the phrase detection, however from a practical point of view, event a partial match is useful.

Another problem in front of evaluation of event extraction systems is matching the numbers of victims. In some cases, the system may detect a number which is close to the annotated number. For example, if in the text there is the phrase *more than 100 died*, the event extraction system may suggest *100* as number of dead. This is not correct, but again, from a practical point of view, it is better to have a rough estimation of the death toll, rather than having no estimation. In such cases, we consider the system output as correct.

6.2 Event type detection

75% of the events in the corpus could be mapped to NEXUS event types. The system classifies the event type with .38 precision and .60 recall (F is .46).

The easiest type is Shooting, the system correctly classified all events. On the other side is Suicide bombing (a terrorist attack), which was most of the times wrongly classified as Explosion (a man-made disaster). The solution will be to make more complex patterns which would distinguish these lexically similar event types.

A large corpus and a trainable classifier would be a good solution for event type detection, although distinguishing close event types would require a very large number of countersamples.

6.3 Event roles detection

The system predicts an event property with .49 recall and .85 precision (F is .63). It performs the best on predicting dead victim specifications (F is .80), the most difficult is perpetrator specification (F is .42). Counts of dead and wouded victims are predicted with F=.57 and F=.62. The complete results are given in table 5.

6.4 Discussion

In 56% of the wrong predictions, the problem was in the grammar. An example:

CZ: Ozbrojenci se dostali do nigerijské věznice tím, že odpálili nálože a zabili při přestřelce jednoho strážce.

EN: Gunmen entered a Nigerian prison by bombing their way inside and killing a guard during a shootout.

Property	R	P	F
dead victim specification	.67	1	.80
dead victim count	.48	.71	.57
wounded victim specification	.63	1	.77
wounded victim count	.50	.80	.62
perpetrator specification	.29	.80	.42
arrested specification	.33	.75	.46
all	.49	.85	.63

Table 5: Results of event roles detection for Czech.

The lexical resources contain both *ozbrojenci = gunmen* as a possible actor, *zabili = killed* as a pattern and *jednoho strážce = a guard* as another possible actor. The perpetrator patterns contain '[perpetrator-group] zabili [dead-group]', however, the word spans between the pattern items does not allow to catch the pattern. A solution could be to allow larger gaps between the pattern items, but this can result in a lower precision.

In 44% of the wrong predictions, the lexical resources were missing the specification. Examples of missing complex person groups:

CZ-1: militantní skupina al-Šabab spojená s al Káidou

EN-1: the militant group al-Shabab associated with al Qaeda

CZ-2: programátor otevřeného software

EN-2: programmer of open software

The are several challenges connected to a rule-based approach and dealing with the Czech language. First, Czech has a free word order. The grammar patterns would need to capture all the following statements. In the following example, all the four sentences could be found in news:

CZ-1: Bombový útok zabil v lednu na moskevském letišti Domodědovo 36 lidí.

CZ-2: Bombový útok zabil na moskevském letišti Domodědovo v lednu 36 lidí.

CZ-3: Bombový útok zabil 36 lidí na moskevském letišti Domodědovo v lednu.

CZ-4: 36 lidí zabil bombový útok v lednu na moskevském letišti Domodědovo.

EN: The suicide bombing killed 36 people at the Moscow's Domodedovo airport in January.

Then, an object can preseed a subject and a lexical form of the nouns cannot distinguish them. The system can thus wrongly exchange a victim and a perpetrator. In the following example, the following sentences are equal and the roles can be distiguished only by their case, not by the position.

CZ-1: Sebevražedný atentátník zabil osm desítek Pákistánců

CZ-2: osm desítek Pákistánců zabil Sebevražedný atentátník
EN: a suicide bomber killed eighty Pakistanians.

As the corpus includes only violent event texts, we cannot see to what extent the system detects false positives (wrongly detects a violent event in a non-event article). We ran the system on 944 general news articles and found only 3 cases of non-violent events captured (0.3%). As an example, the following was classified as an armed conflict, which is not correct as the conflict not happened yet.

CZ: Turci před pár týdny poslali k hranici s Irákem sto tisíc vojáků.
EN: Turks sent to the border with Iraq hundred thousand soldiers a few weeks ago.

We compared the performance of the Czech system to English, which is already well covered in the corpus. The event types in English were recognized better by .16 in F-score and event roles by .17. This can roughly quantify the difference in difficulty between the event extraction task done in these languages.

7 Conclusion

We describe our work towards multilingual evaluation of event extraction systems. Namely, creation of a multilingual event metadata corpus and evaluation of event extraction for the Czech language.

There are many opened issues. First, we plan to extend the evaluation resources. This would make possible training and testing of supervised algorithms for event extraction. As the language coverage of in the corpus differs, the next task is to translate each topic to all the languages. In this way cross-language performance will be more comparable. When working on the event extraction itself, one research direction is machine learning. In the case of event type classification, we need a very large traning corpus to be able to distinguish lexically close event types. For learning of event-role detection features and their frequency by supervised approaches, a large corpus is necessary as well, especially in the case of free-word order languages like Czech. When using a rule-based approach and automatic resource acquision, there are difficulties to cover all the necessary patterns and rules. The current grammars can be further improved by adding some language-specific elements in the rules. The partial coverage of the Czech resources leads to a lower recall. We can improve further the dictionaries by adding the different morphological forms for the words.

Acknowledgments

This work was supported by project MediaGist, EU's FP7 People Programme (Marie Curie Actions), no. 630786.

References

Doug Bond, Joe Bond, Churl Oh, J Craig Jenkins, and Charles Lewis Taylor. 2003. Integrated data for events analysis (idea): An event typology for automated events data development. *Journal of Peace Research*, 40(6):733–745.

Mian Du and Roman Yangarber. 2015. Acquisition of domain-specific patterns for single document summarization and information extraction. In *The Second International Conference on Artificial Intelligence and Pattern Recognition (AIPR2015)*, page 30.

Deborah J Gerner, Philip A Schrodt, Omur Yilmaz, and Rajaa Abu-Jabr. 2002. The creation of cameo (conflict and mediation event observations): An event data framework for a post cold war world. In *annual meeting of the American Political Science Association*, volume 29.

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

Kalev Leetaru and Philip A Schrodt. 2013. Gdelt: Global data on events, location, and tone, 1979–2012. In *ISA Annual Convention*, volume 2, page 4.

Proyecto NewsReader, Rodrigo Agerri, Eneko Agirre, Itziar Aldabe, Begona Altuna, Zuhaitz Beloki, Egoitz Laparra, Maddalen López de Lacalle, German Rigau, Aitor Soroa, et al. 2014. Newsreader project. *Procesamiento del Lenguaje Natural*, 53:155–158.

Ellen Riloff et al. 1993. Automatically constructing a dictionary for information extraction tasks. In *AAAI*, pages 811–816.

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

Philip A Schrodt. 2001. Automated coding of international event data using sparse parsing techniques. In *annual meeting of the International Studies Association, Chicago*.

H Tanev and V Zavarella. 2013. Multilingual learning and population of event ontologies. a case study for social media. *Towards Multilingual Semantic Web (Springer, Berlin & New York*.

Hristo Tanev, Jakub Piskorski, and Martin Atkinson. 2008. Real-time news event extraction for global crisis monitoring. In *Natural Language and Information Systems*, pages 207–218. Springer.

Hristo Tanev, Vanni Zavarella, Jens Linge, Mijail Kabadjov, Jakub Piskorski, Martin Atkinson, and Ralf Steinberger. 2009. Exploiting machine learning techniques to build an event extraction system for portuguese and spanish. *Linguamática*, 1(2):55–66.

Veronika Vincze, György Szarvas, Richárd Farkas, György Móra, and János Csirik. 2008. The bioscope corpus: biomedical texts annotated for uncertainty, negation and their scopes. *BMC bioinformatics*, 9(Suppl 11):S9.

Roman Yangarber and Ralph Grishman. 1998. Nyu: Description of the proteus/pet system as used for muc-7. In *In Proceedings of the Seventh Message Understanding Conference (MUC-7*. Citeseer.

Roman Yangarber, Ralph Grishman, Pasi Tapanainen, and Silja Huttunen. 2000. Automatic acquisition of domain knowledge for information extraction. In *Proceedings of the 18th conference on Computational linguistics-Volume 2*, pages 940–946. Association for Computational Linguistics.