

Steps Towards a System to Extract Formal Narratives from Text

Daniele Metilli
University of Pisa &
ISTI-CNR, Pisa, Italy
daniele.metilli@isti.cnr.it

Valentina Bartalesi
ISTI-CNR, Pisa, Italy
valentina.bartalesi@isti.cnr.it

Carlo Meghini
ISTI-CNR, Pisa, Italy
carlo.meghini@isti.cnr.it

Abstract

In this paper we present a first step towards a system to extract formal narratives from text. This work is part of a wider research on the introduction of narratives in Digital Libraries. We represent narratives as networks of events, each set in space and time, endowed with factual components, and linked to each other through semantic relations. In order to extract a narrative from text, the first step is to automatically detect and classify the events in the text. We present a software we developed that uses neural networks for event detection and classification. It was trained on a dataset of annotated biographies of writers and artists from the English Wikipedia and on the ACE 2005 training corpus. We tested the software on the biography of Florentine poet Dante Alighieri. This software constitutes the first component of a broader system for narrative extraction from natural language text.

1 Introduction

In the last few years, we have been working on introducing narratives into Digital Libraries (DLs).¹ At present, DLs offer search functionalities that respond to a user's query with a list of disparate digital objects based on metadata descriptors. We believe that narratives could act as a semantic glue linking together the digital objects, thereby satisfying the user's information needs in a better way. In this context, we developed an initial Ontology for Narratives, i.e. a formal model to represent narratives [BMM17] based on Semantic Web technologies and built as an extension of the CIDOC CRM ontology [Doe03]. The ontology developed so far allows the formal representation of a narrative. We represent narratives as networks of events, each set in space and time, endowed with factual components (space-time region of occurrence, involved people or objects, etc.), and linked to each other through semantic relations.

In the general definition we adopted as a basis for our ontology, a narrative is composed of a *fabula*, i.e. a network of events in chronological order, and a *plot*, i.e. a network of events selected from the fabula and partially ordered by the narrator [Pro10, Her11]. A narrative also includes a non-formal expression of the story, called *narration*, which can be a natural language text or other medium, e.g. audio or video [Ba97].

Copyright © 2019 for the individual papers by the paper's authors. Copying permitted for private and academic purposes. This volume is published and copyrighted by its editors.

In: A. Jorge, R. Campos, A. Jatowt, S. Bhatia (eds.): Proceedings of the Text2StoryIR'19 Workshop, Cologne, Germany, 14-April-2019, published at <http://ceur-ws.org>

¹<https://dlnarratives.eu>

On top of the ontology, we developed the Narrative Building and Visualising Tool (NBVT), a semi-automatic software that allows users to construct and visualize narratives through a Web interface [Met16]. The tool is able to import knowledge (e.g. locations, people, organisations) from the Wikidata knowledge base [VK14], thereby facilitating the user in the construction of the narrative. The tool has been used to create four narratives² about different subjects: the life of Florentine poet Dante Alighieri, the life of Austrian painter Gustav Klimt, the history of the giant squid, and the history of climate change.

In order to improve the narrative building process, it would be particularly useful to endow the tool with a functionality to automatically extract elements of narrative from natural language text. For instance, the tool could identify the main events in the text, their factual components, and the relations existing among them. In this paper we present the results of an initial experiment towards this goal, focusing on event detection and classification, and we discuss how we plan to proceed further.

Section 2 describes the requirements we identified for the task of narrative extraction from text. Section 3 reports some previous studies that we have found useful when designing our event detection and classification system. Section 4 describes the implementation of our experimental system, including its implementation and evaluation. Section 5 describes our initial work on a user interface for building narratives based on the extracted data. Finally, Section 6 reports conclusions and future works.

2 Technical Requirements for Narrative Extraction

In this section we present some technical requirements for narrative extraction that we have identified. We want to be able to extract as much knowledge as possible from natural language text, in order to aid the user in the narrative building process. The requirements are:

1. *Event Detection.* First of all, events need to be detected in the text. We plan to do so using deep learning [LBH15] techniques allowing us to identify event triggers, i.e. words that express the occurrence of an event.
2. *Event Classification.* Events that are detected in the text need to be classified, i.e. categorized into a set of event classes (for instance, World War II could be classified in a class called *Conflict*).
3. *Named Entity Recognition.* We need to extract from the text the named entities that identify the factual components of the events.
4. *Event Component Extraction.* After we have found all entities and events in the text, we need to identify which entities act as arguments of the events.
5. *Temporal Entity Extraction.* Temporal entities such as dates, years, and other indications of time also need to be extracted. These entities are needed to correctly place the event in the narrative’s timeline.
6. *Relation Extraction.* We want to recognize the relations existing between events and also between events and their components (e.g. the role that a person plays in the event).
7. *Entity and Event Linking.* It is useful to perform entity linking and event linking to connect them to an external knowledge base.
8. *Narrative Construction.* Once we have identified all entities and all events, and linked them to each other, we have to construct the fabula and plot of the narrative.

In this paper we focus on the first two requirements, i.e. event detection and classification, and describe the system we developed to achieve this goal. The satisfaction of the remaining requirements will be tackled in a future publication.

3 Related Works

In this section we report the most relevant related works for our present study about event extraction and narrative extraction from natural language text.

Since the 1990s, machine learning has been applied to the task of event extraction from text, first through rule-based systems [Chi98, LCF⁺92] and later through supervised classifiers [Fre98, CN02]. Starting in 2004,

²<https://dlnarratives.eu/narratives.html>

the ACE program has focused on recognition of entities, values, temporal expressions, relations, and events from natural language text [DMP⁺04]. In particular, the ACE 2005 evaluation involved the extraction of several classes of events from an extensive annotated corpus [WM05].

In the 2010s, research in event extraction made significant progress, with the development of new feature-based and pattern-based approaches [Cao17] and a gradual shift from traditional dependency parsing [LJH13] to deep learning [LBH15]. In particular, convolutional neural networks are often able to identify the structural features of a sentence in a deeper way than feature-based approaches [NG15], but they also have shortcomings, e.g. they find difficulty with sentences containing multiple events [CXL⁺15]. Recurrent neural networks, and in particular bidirectional LSTMs [SP97], are better able to capture long-term dependencies in natural language text [NCG16, FQL18].

The application of event extraction techniques to narrative representation has been limited. These works are particularly relevant to our research topic since they share our goal of extracting not just events, but narratives from text. Chambers & Jurafsky were the first to apply event extraction algorithms to narrative extraction from text [CJ08, CJ09]. Another relevant approach is that of Elson [EDM10], who extracted social networks from a set of 19th Century literary texts. More recently, narratives were applied to information extraction from news streams [VCK15] in the context of the NewsReader project [VRS⁺14].

At the same time, research has been very active on the extraction of events from social media streams [SOM10, BNG11, WL11] and the subsequent development of a knowledge graph based on the detected events [SST⁺18]. Event extraction has also been researched extensively in the biomedical field [YTMT00, BS11].

4 An Event Extraction System for Narratives

As an initial experiment, we implemented a system to perform the first two tasks that we identified as requirements in section 2, i.e. event extraction and classification. After studying the main event extraction systems reported in Section 3, we decided to adopt a recurrent neural network architecture based on bidirectional Long Short-Term Memory (LSTM) [SP97, HS97], since the results obtained through this approach are promising [FQL18]. In this phase of our research, the goal is not to develop a system with better performance on event extraction from text in comparison with the previous ones, but only to investigate if this approach can be satisfactory in order to be integrated in our architecture to build and visualize narratives.

In order to evaluate if the LSTM could be successfully applied for our aims, we selected a case study to test this approach. We chose the narrative of the life of Dante Alighieri as case study, focusing on 12 classes of event that were especially relevant in this case study. The classes are reported in Table 1.

Table 1: The 12 event classes considered in our study.

Event Class	Example
Birth	<i>Dante was born in Florence</i>
Conflict	<i>Dante fought in the Battle of Campaldino</i>
Creation	<i>Dante created the Divine Comedy</i>
Death	<i>Dante died in Ravenna</i>
Education	<i>Dante studied under Brunetto Latini</i>
Election	<i>Dante was elected prior</i>
Marriage	<i>Dante married Gemma Donati</i>
Meeting	<i>Dante met Beatrice Portinari</i>
Membership	<i>Dante joined a Florentine guild</i>
Residence	<i>Dante settled in Verona</i>
Sentence	<i>Dante was condemned to exile</i>
Travel	<i>Dante traveled to Bologna</i>

4.1 Annotation of Training and Test Sets

We analyzed several corpora containing annotated events, such as the ACE 2005 corpus,³ ASTRE [NTFB16], and MEANTIME [MSU⁺16]. Among these, we decided to adopt the ACE 2005 corpus since it contains several event classes that would be useful for our purposes. Unfortunately, some important classes that we need to extract are not present in the corpus (e.g. *Creation*, *Education*, *Residence*, *Membership*).⁴ Furthermore, some of the classes that are present in the corpus are under-represented (e.g. *Birth*). For these reasons, we decided to supplement the ACE 2005 dataset with an additional training dataset of our own development.

Given that our case study about the life of Dante Alighieri is based on the English Wikipedia page about the poet, we selected 10 biographies of writers and artists from Wikipedia as additional training set. Our aim was to collect texts containing events that describe the life of a person. We included biographies of people who lived in a wide range of time, from the Middle Ages to the 20th century. We developed a simple annotation interface that we used to tag all event triggers found in the text. The interface, shown in Figure 4.1, allows the annotation of any number of classes defined by the user, and is able to export the annotated dataset in JSON⁵ format. The annotation of the training set was performed by two annotators. The inter-annotator agreement, measured using Cohen’s kappa statistic [Coh60], is reported in Table 2.

Table 2: Inter-annotator agreement in the Wikipedia training set.

Event Class	Cohen’s Kappa
Birth	0.92
Conflict	0.75
Creation	0.82
Death	0.95
Education	0.91
Election	0.86
Marriage	0.86
Meeting	0.90
Membership	0.80
Residence	0.85
Sentence	0.78
Travel	0.89
Event (overall)	0.86

The total number of annotated events in this training set is 604, with an average of about 50 events per class. Given the fact that the subjects of the Wikipedia pages are all writers and artists, the most represented event class is *Creation* (183 events). The least represented classes are *Membership* (8) and *Sentence* (9). All other classes appear at least 15 times in the set.

As test set, we annotated the English Wikipedia page about Dante Alighieri using the same annotation interface that we developed for the training corpus and following the same annotation methodology. The resulting annotation contains 88 events, with a minimum of 3 instances, and an average of about 7 instances, for each event class. The most represented event class is *Conflict* (14), which is understandable given the fact that Dante was not just a poet but also a politician, and he actively took part in the political and military conflicts of his era. The second most represented class is *Creation* (11), reflecting Dante’s work as a poet and writer.

³<https://projects.ldc.upenn.edu/ace/>

⁴The ACE 2005 event classes are listed at: <https://www.ldc.upenn.edu/files/english-events-guidelines-v5.4.3.pdf>

⁵<https://www.json.org>

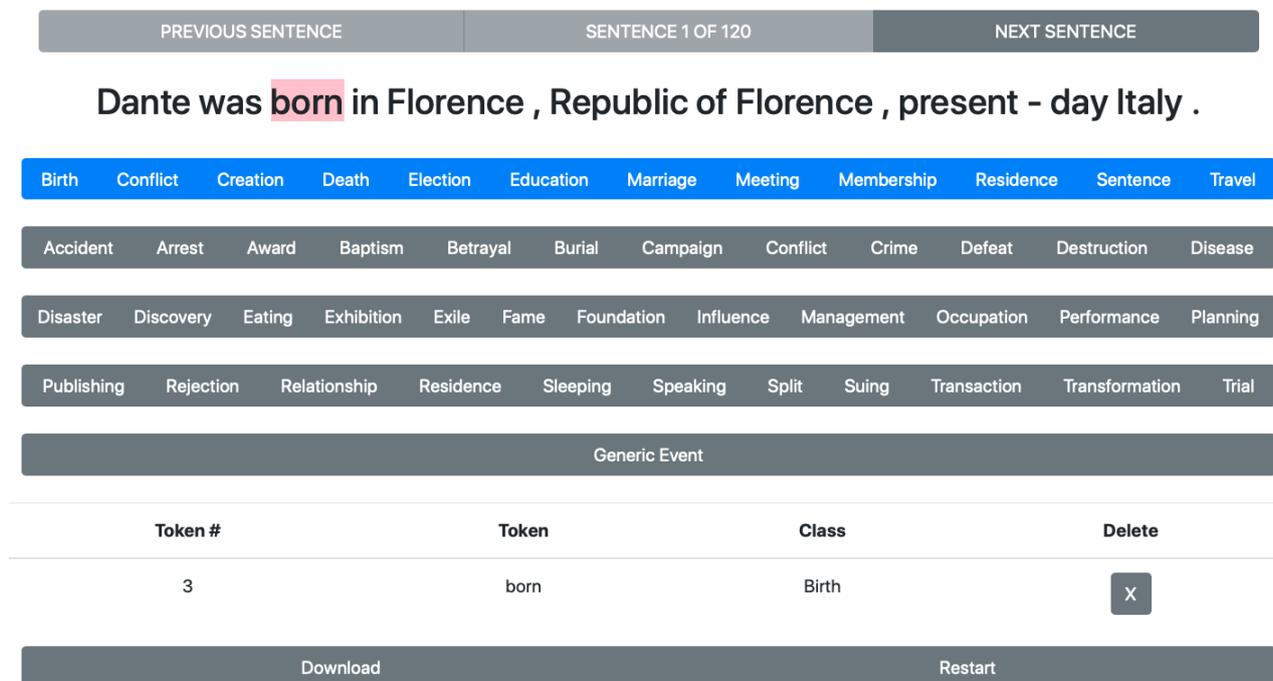


Figure 1: The annotation interface that we developed to build our training corpus.

4.2 Implementation of Event Extraction

We implemented a software for event extraction based on LSTM, a type of recurrent neural network. We started from an open source implementation,⁶ [RG17] to which we applied several modifications (in particular, we added layer normalization [BKH16] and K-max pooling [KGB14]). These modifications were shown experimentally to achieve better performance. An ablation analysis showed that removing layer normalization results in a 2.9% decrease in the F_1 -score. Removing K-max pooling reduces the F_1 -score by 0.5%.

The software takes as input the annotated sentences, tokenizes them, and applies the Komninos dependency-based word embeddings [KM16]. The embeddings are given as input to a convolutional neural network (CNN), after which pooling is applied. The results are fed to two LSTM layers of different sizes, in sequence. Dropout is applied both to the CNN and to each LSTM layer. Finally, a hidden dense layer computes the final results by applying a softmax function [Bri90].

4.3 Evaluation

Given the heterogeneity of the two training sets, we decided to run the neural network model twice, first on the ACE 2005 dataset and then on the set of 10 Wikipedia biographies we annotated. In both cases, we performed validation on 10% of the data, which was held out from the two training sets. After running the model on the two datasets, we combined the results using a weighting function. We performed some tests assigning different weights to the two datasets in order to identify the optimal weights to maximize the F_1 -score. Based on the results of our tests, we decided to assign a 0.33 weight to the ACE dataset and a 0.66 weight to the Wikipedia dataset. Finally, we computed precision, recall, and F_1 -score using as test set our annotation of the Wikipedia page about Dante Alighieri. The results of the evaluation are reported in Table 3.

Table 3: Results of the evaluation.

Task	Precision	Recall	F_1 -Score
Event Detection	73.7	72.3	73.0
Event Classification (overall) ⁷	73.3	67.5	70.3

⁶<https://github.com/UKPLab/emnlp2017-bilstm-cnn-crf>

⁷Computed as the unweighted macro-average among the 12 event classes.

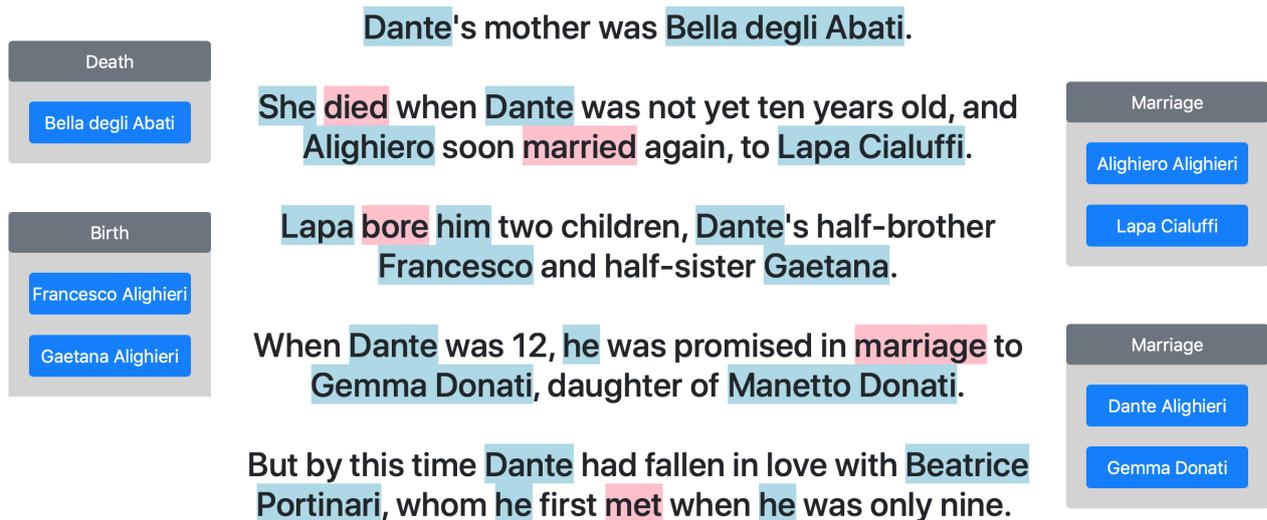


Figure 2: The user interface allows the user to automatically detect events in the text.

Our results cannot be compared directly to the baselines for the ACE 2005 dataset reported in the publications cited in the Related Works section, since we introduced a different corpus with a different set of classes. As baseline for the detection task, we chose the results of the event extraction component of the NewsReader software [VRS⁺14]. NewsReader is provided as a service by the NLP Hub of the DataMiner cloud computing system, based on the D4Science e-Infrastructure [CPSP17]. On our test set, NewsReader was able to detect events with 71.9% precision and 61.4% recall, achieving an F_1 -Score of 66.2%. Our software constitutes an improvement over this baseline (see Table 3). We plan to conduct a more detailed evaluation in the near future.

5 An Initial User Interface for Narrative Extraction

After completing our experiment on event extraction and classification, we started developing a web interface that we plan to integrate into our Narrative Building and Visualising Tool. The interface allows the user to build events starting from the ones identified as candidate events by the event extraction system.

In addition to events, the current version of the interface also shows named entities (extracted using Stanford CoreNLP [MSB⁺14]). When possible, the named entities are linked to Wikipedia by using the WAT [PF14] entity linking service, and subsequently to Wikidata. Since at this stage the system does not perform relation extraction, the interface allows the user to manually link the entities that are factual components of the events to the identified events through drag-and-drop functionality. The interface is shown in Figure 5.

6 Conclusions and Future Works

In this paper we have presented an initial study on narrative extraction from text, focusing on event detection and classification. To detect and classify events, we have developed a software based on recurrent neural networks. We have built a training set based on English Wikipedia pages of writers and artists, and tested the system on the biography of Florentine poet Dante Alighieri. Based on the current training set, the system has achieved an F_1 -score of 73.0 on event detection and an F_1 -score of 70.3 on event classification. We have also started development of a full-fledged narrative extraction interface. Currently, the interface allows the user to automatically detect events and named entities in the text, and manually link them to each other.

As future works, we plan to build a complete system for narrative extraction from text, structured in several components. In addition to the initial event detection and classification component we presented in this paper, the system will be endowed with components for further automating the narrative extraction process, e.g. by extracting relations between events and their related entities, and linking both entities and events to an existing knowledge base.

References

- [Bal97] Mieke Bal. *Narratology: Introduction to the theory of narrative*. University of Toronto Press, 1997.
- [BKH16] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. Layer normalization. *arXiv preprint arXiv:1607.06450*, 2016.
- [BMM17] Valentina Bartalesi, Carlo Meghini, and Daniele Metilli. A conceptualisation of narratives and its expression in the CRM. *International Journal of Metadata, Semantics and Ontologies*, 12(1):35–46, 2017.
- [BNG11] Hila Becker, Mor Naaman, and Luis Gravano. Beyond trending topics: Real-world event identification on Twitter. *Proceedings of ICWSM*, 11(2011):438–441, 2011.
- [Bri90] John S Bridle. Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. In *Neurocomputing*, pages 227–236. Springer, 1990.
- [BS11] Jari Björne and Tapio Salakoski. Generalizing biomedical event extraction. In *Proceedings of the BioNLP Shared Task 2011 Workshop*, pages 183–191. Association for Computational Linguistics, 2011.
- [Cao17] Kai Cao. *Improving Event Extraction: Casting a Wider Net*. Phd thesis, University of New York, 2017. Available from https://cs.nyu.edu/media/publications/cao_kai.pdf.
- [Chi98] Nancy Chinchor. Overview of MUC-7. In *Proceedings of the Message Understanding Conference*, 1998.
- [CJ08] Nathanael Chambers and Dan Jurafsky. Unsupervised learning of narrative event chains. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 789–797, 2008.
- [CJ09] Nathanael Chambers and Dan Jurafsky. Unsupervised learning of narrative schemas and their participants. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 602–610. Association for Computational Linguistics, 2009.
- [CN02] Hai Leong Chieu and Hwee Tou Ng. A maximum entropy approach to information extraction from semi-structured and free text. *AAAI/IAAI*, 2002:786–791, 2002.
- [Coh60] Jacob Cohen. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1):37–46, 1960.
- [CPSP17] Gianpaolo Coro, Giancarlo Panichi, Paolo Scarponi, and Pasquale Pagano. Cloud computing in a distributed e-infrastructure using the web processing service standard. *Concurrency and Computation: Practice and Experience*, 29(18):e4219, 2017.
- [CXL⁺15] Yubo Chen, Liheng Xu, Kang Liu, Daojian Zeng, and Jun Zhao. Event extraction via dynamic multi-pooling convolutional neural networks. In *Proceedings of of the Annual Meeting of the Association for Computational Linguistics*, volume 1, pages 167–176, 2015.
- [DMP⁺04] George R Doddington, Alexis Mitchell, Mark A Przybocki, Lance A Ramshaw, Stephanie Strassel, and Ralph M Weischedel. The automatic content extraction (ace) program-tasks, data, and evaluation. In *Proceedings of the International Conference on Language Resources and Evaluation*, volume 2, page 1, 2004.
- [Doe03] Martin Doerr. The CIDOC Conceptual Reference Module: an ontological approach to semantic interoperability of metadata. *AI magazine*, 24(3):75, 2003.
- [EDM10] David K Elson, Nicholas Dames, and Kathleen R McKeown. Extracting social networks from literary fiction. In *Proceedings of of the Annual Meeting of the Association for Computational Linguistics*, pages 138–147. Association for Computational Linguistics, 2010.
- [FQL18] Xiaocheng Feng, Bing Qin, and Ting Liu. A language-independent neural network for event detection. *Science China Information Sciences*, 61(9):092106, 2018.

- [Fre98] Dayne Freitag. Multistrategy learning for information extraction. In *Proceedings of the International Conference on Machine Learning*, pages 161–169, 1998.
- [Her11] David Herman. *Basic elements of narrative*. John Wiley & Sons, 2011.
- [HS97] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- [KGB14] Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. A convolutional neural network for modelling sentences. *arXiv preprint arXiv:1404.2188*, 2014.
- [KM16] Alexandros Komninos and Suresh Manandhar. Dependency based embeddings for sentence classification tasks. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1490–1500, 2016.
- [LBH15] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436, 2015.
- [LCF⁺92] Wendy Lehnert, Claire Cardie, David Fisher, John McCarthy, Ellen Riloff, and Stephen Soderland. MUC-4 test results and analysis. In *Proceedings of the Message Understanding Conference*, pages 151–158. Association for Computational Linguistics, 1992.
- [LJH13] Qi Li, Heng Ji, and Liang Huang. Joint event extraction via structured prediction with global features. In *Proceedings of of the Annual Meeting of the Association for Computational Linguistics*, volume 1, pages 73–82, 2013.
- [Met16] Daniele Metilli. A Wikidata-based tool for the creation of narratives. Master’s thesis, University of Pisa, 2016. Available from <https://etd.adm.unipi.it>.
- [MSB⁺14] Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of of the Annual Meeting of the Association for Computational Linguistics*, pages 55–60, 2014.
- [MSU⁺16] A-L Minard, Manuela Speranza, Ruben Urizar, Begona Altuna, MGJ van Erp, AM Schoen, CM van Son, et al. Meantime, the newsreader multilingual event and time corpus. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)*, page 4417, 2016.
- [NCG16] Thien Huu Nguyen, Kyunghyun Cho, and Ralph Grishman. Joint event extraction via recurrent neural networks. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, pages 300–309, 2016.
- [NG15] Thien Huu Nguyen and Ralph Grishman. Event detection and domain adaptation with convolutional neural networks. In *Proceedings of of the Annual Meeting of the Association for Computational Linguistics*, volume 2, pages 365–371, 2015.
- [NTFB16] Kiem-Hieu Nguyen, Xavier Tannier, Olivier Ferret, and Romaric Besançon. A dataset for open event extraction in english. In *LREC*, 2016.
- [PF14] Francesco Piccinno and Paolo Ferragina. From TagME to WAT: a new entity annotator. In *Proceedings of the First International Workshop on Entity Recognition & Disambiguation*, pages 55–62. ACM, 2014.
- [Pro10] Vladimir Propp. *Morphology of the Folktale*. University of Texas Press, 2010.
- [RG17] Nils Reimers and Iryna Gurevych. Reporting score distributions makes a difference: Performance study of lstm-networks for sequence tagging. *arXiv preprint arXiv:1707.09861*, 2017.
- [SOM10] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. Earthquake shakes Twitter users: real-time event detection by social sensors. In *Proceedings of the 19th International Conference on World Wide Web*, pages 851–860. ACM, 2010.
- [SP97] Mike Schuster and Kuldeep K Paliwal. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11):2673–2681, 1997.

- [SST⁺18] Saeedeh Shekarpour, Ankita Saxena, Krishnaprasad Thirunarayan, Valerie L Shalin, and Amit Sheth. Principles for developing a knowledge graph of interlinked events from news headlines on Twitter. *arXiv preprint arXiv:1808.02022*, 2018.
- [VCK15] Piek Vossen, Tommaso Caselli, and Yiota Kontzopoulou. Storylines for structuring massive streams of news. In *Proceedings of the First Workshop on Computing News Storylines*, pages 40–49, 2015.
- [VK14] Denny Vrandečić and Markus Krötzsch. Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 57(10):78–85, 2014.
- [VRS⁺14] Piek Vossen, German Rigau, Luciano Serafini, Pim Stouten, Francis Irving, and Willem Robert Van Hage. Newsreader: recording history from daily news streams. In *LREC*, pages 2000–2007. Citeseer, 2014.
- [WL11] Jianshu Weng and Bu-Sung Lee. Event detection in Twitter. *Proceedings of ICWSM*, 11:401–408, 2011.
- [WM05] Christopher Walker and Julie Medero. *ACE 2005 Multilingual Training Data V6.0 LDC2005E18*. Web download, Linguistic Data Consortium, Philadelphia, 2005.
- [YTMT00] Akane Yakushiji, Yuka Tateisi, Yusuke Miyao, and Jun-ichi Tsujii. Event extraction from biomedical papers using a full parser. In *Biocomputing 2001*, pages 408–419. World Scientific, 2000.