

Towards the Automatic Analysis of the Structure of News Stories

Iqra Zahid

School of Arts, Languages and Cultures
University of Manchester, UK
iqra.zahid@student.manchester.ac.uk

Hao Zhang

School of Computer Science
University of Manchester, UK
hao.zhang-17@postgrad.manchester.ac.uk

Frank Boons

Alliance Manchester Business School
University of Manchester, UK
frank.boons@manchester.ac.uk

Riza Batista-Navarro

School of Computer Science
University of Manchester, UK
riza.batista@manchester.ac.uk

Abstract

News stories are distinct from other types of narratives in that they typically follow a complex and non-chronological time structure. This poses challenges to the narrative analysis of news, specifically with respect to the construction of event sequences. In this paper, we propose to segment news story text according to news schema categories, which allow for identifying sentences describing a news story’s main action and other actions that happened beforehand or subsequently. To automate this task, we made observations on the linguistic devices that are used by news writers, based on a manually annotated corpus of news articles that we have constructed. Heuristics capturing these linguistic devices were then developed, underpinned by natural language processing tools as well as carefully curated look-up lists of cues. While encouraging preliminary results were obtained, the work can be further expanded by observing and capturing more linguistic devices, which can be facilitated by further annotation of news stories based on news schema categories.

1 Introduction

In analysing narratives, understanding the sequence in which events occur is key [Ell05]. Most types of narratives, e.g., novels, personal accounts of experiences, present events in chronological order. However, *news stories*, narratives that are written or recorded to “inform the public about current events, concerns or ideas” [Whi], deviate from other types of narratives in that they follow a complex time structure. News writers are expected to prioritise certain news values, i.e., criteria for judging “newsworthiness” (e.g., negativity, unexpectedness, superlativeness) [Bel91]. In producing news stories that adhere to such news values, news writers adopt the

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>

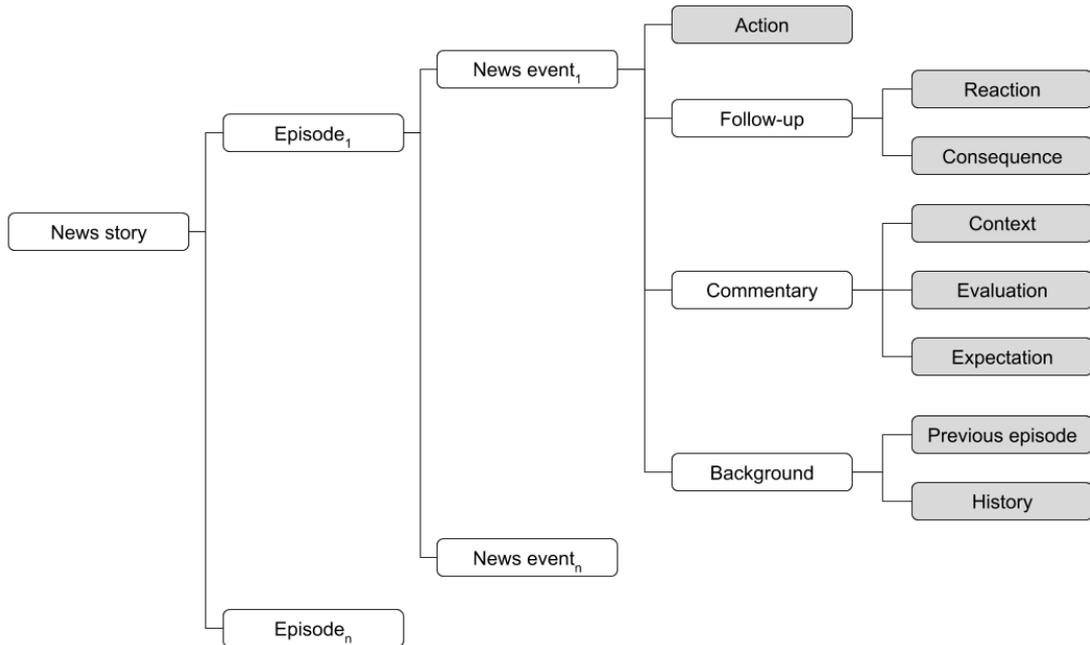


Figure 1: News schema proposed by Allan Bell [Bel91]. Shown in grey are the most specific categories in the schema.

instalment method, whereby an event that was introduced in the earlier parts of a story, may be described in detail only later on in the story, possibly in multiple, separate instances. Consequently, events are usually presented in news stories in a non-chronological order.

In order to understand the flow of events in news stories, it is necessary to analyse their schema, i.e., the overall form of news discourse, by which topics are organised. A *news schema* defines the syntax of news stories, providing a set of formal categories that form the basis of the hierarchical organisation and ordering of textual units [Dij85]. In early work by Labov and Waletzky on discourse analysis, categories such as Abstract, Orientation, Complicating Action, Evaluation, Resolution and Coda were proposed in order to organise narratives of personal experiences [LW67]. Building upon that work, van Dijk [Dij85] developed a schema specifically for analysing news discourse. Each category in the schema, e.g., Main Event, Background and History, corresponds to a piece of text, i.e., a sequence of sentences. According to case studies carried out on hundreds of news reports published in more than 260 newspapers from 100 countries, this news schema is applicable at an international scale [vD98]. A few years later, building upon van Dijk’s work, Bell proposed a finer-grained news schema [Bel91]. We reproduce a tree-like depiction of this schema in Figure 1, in which the most specific (or lowest-level) categories are shown in grey.

Since the chronological order of events is not maintained in news stories (as discussed above), narrative analysis of news is more challenging, compared to that of other types of narratives (e.g., novels). As human readers, we are accustomed to the style of reporting employed in news stories, and thus we might find the task of determining the correct sequence of events a simple and straightforward task. However, to an automated system designed to support narrative analysis, the non-chronological order in which events are presented in news stories would pose a barrier in the reconstruction of event sequences.

In this paper, we aim to facilitate machine understanding of news stories by automatically decomposing them according to news schema categories. To this end, we firstly developed a corpus of written news stories in which spans of text corresponding to news schema categories have been manually annotated and labelled following the work of Bell [Bel91]. We then identified the various linguistic devices that are usually employed by news writers, that can help in the task of mapping news story text, to respective news schema categories. On the basis of these, a heuristics-based approach was developed in order to automate the said task.

The remainder of this paper is organised as follows. Section 2 presents a review of previously reported related work. In Section 3, an analysis of linguistic devices used in the different news schema categories is presented,

supported by an annotated corpus that we have recently developed. We also provide a discussion of the heuristics that were developed to detect the use of such linguistic devices, in order to identify parts of news story text that correspond to schema categories. Our preliminary results are then discussed in Section 4. Finally, we conclude and present our next steps in Section 5.

2 Related Work

Most of the efforts that have been carried out in the way of analysing news were focussed on identifying boundaries between news stories, rather than on analysing their structure individually. Early work employed cue words as well as named entities (e.g., names of people, places, organisations) in closed caption text, in order to detect transitions from one news story to another [MMM97]. The Broadcast News Navigator (BNN) system similarly used cue words and named entities in segmenting closed caption text according to individual news stories [May98]. Additionally, by selecting the sentence with the highest frequency of named entities, the system was able to generate a gist for each news story—slightly similar to the Action category in Bell’s framework which pertains to the central or main action in a news story.

TextTiling, a text segmentation approach proposed by Hearst, was applied to the detection of boundaries between consecutive news articles in the Wall Street Journal [Hea97]. Specifically, the approach segmented text according to subtopics, which were identified through the measurement of lexical cohesion. Other work sought to improve the definition and measurement of lexical cohesion, by incorporating richer features (e.g., number of pronouns, similarities obtained by Latent Dirichlet Allocation), and were applied to the detection of news story boundaries in transcripts of broadcast news [Sto03, RH06, PMD09].

More similar to our own work are efforts aimed at segmenting individual narratives. The work of Kauchak and Chen [KC05] was aimed at segmenting an individual narrative according to the topics it contains, casting the problem as a text classification task in which features such as word groups (identified with the aid of WordNet) and entity groups were learned using machine learning-based methods, i.e., support vector machines and decision stumps (one-level decision trees). This approach was applied on books autobiography-style books and encyclopaedia articles. Our approach is distinct from this in that we seek to segment individual news stories, which as discussed in the previous section, follow a different structure relative to other types of narratives. While the work of Cardoso et al. [CTP13] was targeted towards the analysis of news text (written in Brazilian Portuguese), they also used topics as the basis of segmentation.

Our work aims to segment a narrative with the end-goal of following the flow or sequence of events, rather than identifying the different topics or themes it contains. While this bears similarities with the narrative segment annotation task proposed by Reiter [Rei15] which was manually applied to short stories, our approach is specifically aimed at automatically analysing news stories. The topic of a news story may consist of multiple interconnected events, and thus can be segmented according to news schema categories delineating the main event from events leading to it and following it. To the best of our knowledge, our proposed approach is the first to attempt to automatically analyse the structure of news stories in this way.

3 Methodology

Our proposed approach to the automatic segmentation of news text according to news schema categories is based on the analysis of the various linguistic devices used by writers.

3.1 Corpus development

In order to support our analysis of linguistic devices, we developed a small corpus of written news articles, retrieved using the LexisNexis library¹. Containing a total of 22 articles from various news agencies (listed in Appendix A), the corpus is partitioned into two: (1) a pre-2005 set, mostly containing news stories published in the 1980s and 1990s, that were used by van Dijk [Dij85, vD98] and Bell [Bel91] in designing their respective frameworks; and (2) a post-2005 set, containing eight news stories published more recently.

An annotation scheme was designed requiring annotators to map pieces of news text to the most specific categories of the news schema tree (shown in grey in Figure 1). Guidelines were then established, to promote consistency between annotators. Specifically, annotators were asked to provide annotations at the sentence level. That is, category labels were assigned to individual sentences, not to whole paragraphs (sequence of sentences) nor to clauses or phrases (parts of sentences). If a sequence of sentences corresponds to only one category, then

¹<https://www.lexisnexis.com/uk/legal/>

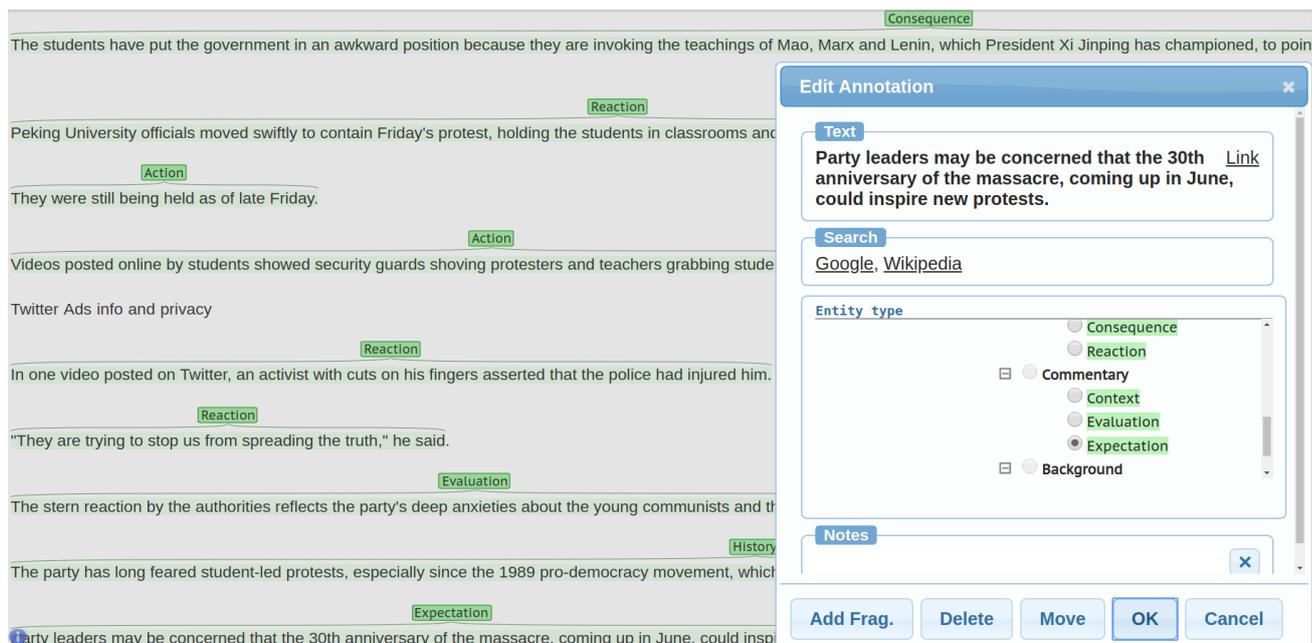


Figure 2: News story “Students Defiant as Chinese University Cracks Down on Young Communists” (Javier C.Hernandez, The New York Times, 28 December 2018) annotated and visualised using the brat annotation tool.

each sentence in that sequence was labelled as that one category. On each sentence, only one of eight of the most specific news schema categories (Action, Reaction, Consequence, Context, Evaluation, Expectation, Previous episode, History) was applied. We refer the reader to Appendix B for the definitions of these categories, as well as corresponding example sentences coming from a news story. Annotators were encouraged to firstly identify sentences in a news story that pertain to Action, as the other seven news schema categories are defined in relation to it. In cases where a sentence seemed to map to multiple categories, the annotator was asked to choose only one based on his or her best judgement.

Using the brat rapid annotation tool [SPT⁺12], two annotators carried out the annotation task. One annotator, a final-year linguistics student (the first author of this paper), marked up all of the 22 articles. The other annotator, a researcher with expertise in natural language processing and text mining (the last author), annotated only the post-2005 set. Shown in Figure 2 is a sample news story annotated and visualised in brat.

The resulting corpus contains a total of 570 sentences. The average number of sentences is 26, with the shortest and longest news stories containing 11 and 53 sentences, respectively. Shown in Figure 3 is the number of annotated sentences in the corpus for each news schema category.

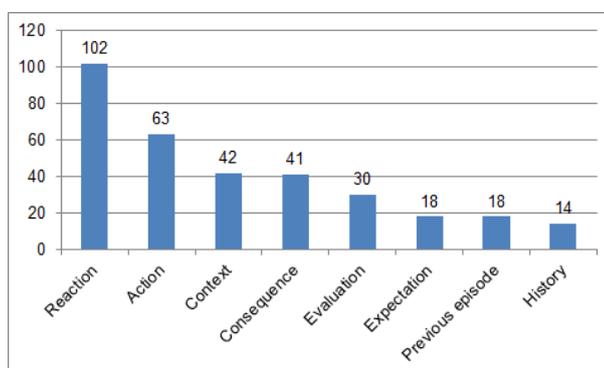


Figure 3: Number of sentences in the annotated corpus, for each news schema category.

3.2 Linguistic Devices

Utilising our manually annotated corpus, we made observations on the various linguistic devices used by news story writers, as we posit that these are helpful in discriminating between different news schema categories. These observations are discussed for each of the eight most specific news schema categories, together with our proposed heuristics for automatically capturing them.

3.2.1 Action.

We observed that sentences pertaining to Action, defined as a central or main action in a news story, share lexical similarities with the title of the news story. For example, in the news article published by the BBC News on the 21st December 2018 entitled, “Cheshire BMW driver jailed over speeding ticket lie”, the following sentences pertain to Action: (1) “A man who claimed his BMW had been cloned as part of an elaborate scam to avoid a speeding fine has been jailed.”; and (2) “Robson was jailed for nine months at Chester Crown Court on Thursday.” Words in these sentences such as “jailed” and “speeding” are shared (verbatim) with the title of the news story. Based on this observation, we automatically identified text pertaining to Action by checking for exact matches between the lemmatised words of a sentence and those of the news story title.

3.2.2 Reaction.

A Reaction is a verbal response to an Action given by an actor. A commonly used linguistic device used by writers to describe Reactions is *attribution*, which indicates “who expressed what”, where *what* pertains to a quotation or perception, and “who” denotes its original source, i.e., the actor. The information that is commonly attributed is direct speech (e.g., He said, “She will deny it.”) or indirect speech (e.g., He said that she will deny it.) in which reporting verbs (e.g., “said”, “announce”, “comment”, “mention”) are often used. However, verbs that are less neutral and bear either positive or negative connotations may also be used, such as “applaud”, “praise” and “complain”.

To detect whether a sentence contains a Reaction, we leveraged previously reported work on attribution extraction, which is underpinned by a lexicon of attribution verbs [ZBBNss]. As Reactions are responses to Actions, they often contain mentions that co-refer to either the Action itself or actors participating in it. Hence, a check for the use of definite noun phrases was also implemented in order to detect whether an attributed quotation contains any co-referring mentions.

3.2.3 Consequence.

A Consequence is an occurrence that transpires as a result of an Action, with the exception of verbal responses (which are classified as Reactions). As such, Consequences often contain mentions that co-refer to either the Action itself or actors participating in it. Furthermore, discourse connectives signifying causation (e.g., “as a result”, “because”, “thereby”) tend to be used in sentences pertaining to Consequence. In order to detect the use of such linguistic devices, we checked for the existence of definite noun phrases in sentences, as well as for the use of any of the discourse connectives annotated in the Penn Discourse Treebank (PDTB) [PDL⁺08] that denote the Contingency relation (with a minimum frequency = 4).

3.2.4 Evaluation.

Evaluation consists of observations on an Action provided by the news writer (i.e., journalist) or an actor, that assesses its impact or significance. As in Reaction, attribution is often used in Evaluation, specifically in cases where the observations are coming from actors. However, sentences conveying Evaluation can be identified by checking for the presence of graded adjectives, as these often indicate assessment, i.e., the degree to which a quality holds (e.g., “deep”, “strongest”, “biggest”). In support of this step, more than 260 graded adjectives were collected from the Collins Cobuild Grammar Patterns reference book [Sin98] and compiled into a look-up list.

3.2.5 Expectation.

Like Evaluation, Expectation is comprised of observations provided by the news writer or an actor (in which case attribution is also used), but pertains to their views on what could happen in the future. As such, sentences

corresponding to Expectation make use of speculative language. To facilitate the detection of speculative language, we checked for the presence of modal verbs (e.g., “could”, “may”), as well as for presence of modifiers that indicate uncertainty. A list of such modifiers was drawn from uncertainty cues in the WikiWeasel 2.0 corpus, that were manually annotated by Vincze [Vin14].

3.2.6 Context.

Similar to Evaluation and Expectation, Context refers to observations given by either the news writer or an actor in order to provide additional information that help explain or clarify details surrounding an Action. Based on our observations, sentences that fall under this category do not have any defining linguistic features (unlike Evaluation and Expectation as described above), except for the prevalent use of co-referring mentions. We detected this by checking if definite noun phrases appear as either the subject or object of sentences.

3.2.7 Previous episode.

Sentences pertain to a Previous episode if they describe any event that happened prior to an Action, in the not-so-distant (or near) past. The main verbs of such sentences are often in either the past or past perfect tense. Additionally, relative temporal expressions pertaining to recent points in time (e.g., “last week”, “previously”, “on Friday”) also tend to be used in specifying the time of occurrence of events falling under Previous episode.

3.2.8 History.

Similar to Previous episode, History describes events that happened prior to an Action, but before the near past. Sentences that belong to this category typically have main verbs in either the past or past perfect tense. They also describe events whose time of occurrence are mentioned in the form of absolute temporal expressions (e.g., “in 1989”). However, relative temporal expressions may also be used, although these would pertain to a point in time from the distant past (e.g., “three decades ago”).

4 Preliminary Results

In implementing the heuristics for capturing linguistic devices that were discussed in the previous section, a pipeline for preprocessing was developed, based on three tools. Firstly, we made use of the LingPipe sentence splitter², to automatically segment news text into individual sentences. Each sentence is then decomposed into tokens by the GENIA Tagger³. The tokenised sentence is then given as input to the Enju Parser⁴, through which we obtained not only the part-of-speech (POS) tag and lemma for each token but also predicate-argument structures identifying the sentence’s main verb and its arguments (i.e., subject and object).

We then developed (in Python) rules for analysing the preprocessing results, for each news schema category (as described in the previous section). These include: (1) checking for specific values of POS tags, e.g., to check for verb tense and for modal verbs; (2) matching lemmatised tokens in look-up lists, e.g., of uncertainty cues, graded adjectives; (3) checking for definite noun phrases and whether they act as the subject or object of a main verb; and (4) matching against regular expressions designed to capture absolute and relative temporal expressions.

Upon application on the post-2005 set of our annotated corpus (containing eight news stories), our heuristics for identifying sentences obtained an overall performance of 64% (over all eight news schema categories) in terms of F-score (precision = 70%, recall = 59%).

5 Future Work and Conclusion

In this paper, we presented our work on automatically analysing the structure of news stories according to news schema categories. While our preliminary results are encouraging, there is significant room for improvement. Recognising that the current version of our annotated corpus is limited in size, we shall a dedicate a large part of our immediate next step on expanding it. This will allow us to observe any further linguistic devices used in each news schema category, and in turn, to eventually extend our heuristics. Our annotated corpus will be made publicly available upon completion of this planned expansion. We then intend to investigate how

²<http://alias-i.com/lingpipe/index.html>

³<http://www.nactem.ac.uk/GENIA/tagger/>

⁴<http://www.nactem.ac.uk/enju/>

our automatically assigned news schema categories can be used as features to inform event temporal relation extraction, in the way of automatically constructing event sequences.

5.0.1 Acknowledgements

The research on which this article is based was partially funded by the Alliance Manchester Business School Strategic Investment Fund.

References

- [Bel91] Allan Bell. *The Language of News Media*, chapter 8, pages 147–174. Blackwell, Oxford, UK, 1991.
- [CTP13] Paula C. F. Cardoso, Maite Taboada, and Thiago A. S. Pardo. Subtopic annotation in a corpus of news texts: Steps towards automatic subtopic segmentation. In *Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology*, 2013.
- [Dij85] Teun A. Van Dijk. Structures of news in the press. In *In Discourse and Communication: New Approaches to the Analysis of Mass Media Discourse and Communication*, 1985.
- [Ell05] Jane Elliott. *Using Narrative in Social Research*, chapter 1, pages 2–16. SAGE Publications Ltd, London, UK, 2005.
- [Hea97] Marti A. Hearst. Texttiling: Segmenting text into multi-paragraph subtopic passages. *Computational Linguistics*, 23(1), 1997.
- [KC05] David Kauchak and Francine Chen. Feature-based segmentation of narrative documents. In *Proceedings of the ACL Workshop on Feature Engineering for Machine Learning in Natural Language Processing*, FeatureEng '05, pages 32–39, Stroudsburg, PA, USA, 2005. Association for Computational Linguistics.
- [LW67] William Labov and Joshua Waletzky. Narrative Analysis. In *In Essays on the Verbal and Visual Arts: Proceedings of the 1966 Annual Spring Meeting of the American Ethnological Society*, pages 12–44, Seattle, USA, 1967. University of Washington Press.
- [May98] Mark T. Maybury. Discourse cues for broadcast news segmentation. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics - Volume 2*, ACL '98/COLING '98, pages 819–822, Stroudsburg, PA, USA, 1998. Association for Computational Linguistics.
- [MMM97] Andrew Merlino, Daryl Morey, and Mark Maybury. Broadcast news navigation using story segmentation. In *Proceedings of the Fifth ACM International Conference on Multimedia*, MULTIMEDIA '97, pages 381–391, New York, NY, USA, 1997. ACM.
- [PDL⁺08] Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind K. Joshi, and Bonnie L. Webber. The penn discourse treebank 2.0. In *LREC*, 2008.
- [PMD09] G. Poulisse, M. Moens, and T. Dekens. News story segmentation in multiple modalities. In *2009 Seventh International Workshop on Content-Based Multimedia Indexing*, pages 25–32, June 2009.
- [Rei15] Nils Reiter. Towards Annotating Narrative Segments. In *Proceedings of the 9th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, LaTeCH 2015, pages 34–38, Stroudsburg, PA, USA, 2015. Association for Computational Linguistics.
- [RH06] Andrew Rosenberg and Julia Hirschberg. Story segmentation of broadcast news in english, mandarin and arabic. In *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers*, NAACL-Short '06, pages 125–128, Stroudsburg, PA, USA, 2006. Association for Computational Linguistics.
- [Sin98] John Sinclair. *Collins Cobuild Grammar Patterns 2: Nouns and Adjectives*. Harper Collins Publishers, London, UK, 1998.

- [SPT⁺12] Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun’ichi Tsujii. BRAT: A Web-based Tool for NLP-assisted Text Annotation. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, EACL ’12, pages 102–107, Stroudsburg, PA, USA, 2012. Association for Computational Linguistics.
- [Sto03] Nicola Stokes. Spoken and written news story segmentation using lexical chains. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology: Proceedings of the HLT-NAACL 2003 Student Research Workshop - Volume 3*, NAACLstudent ’03, pages 49–54, Stroudsburg, PA, USA, 2003. Association for Computational Linguistics.
- [vD98] Teun van Dijk. *News Analysis: Case Studies of International and National News in the Press*. Lawrence Erlbaum Associates, Hillside, New Jersey, USA, 1998.
- [Vin14] Veronika Vincze. *Uncertainty Detection in Natural Language Texts*. PhD thesis, Doctoral School in Computer Science, University of Szeged, Szeged, Hungary, 7 2014.
- [Whi] Aimee Whitman. The Community Toolbox: Creating News Stories the Media Wants. Online: <https://ctb.ku.edu/en/table-of-contents/advocacy/media-advocacy/news-stories-media-wants/main>. Accessed: 2019-01-30.
- [ZBBNss] Hao Zhang, Frank Boons, and Riza Batista-Navarro. Whose Story Is It Anyway? Automatic Extraction of Accounts from News Articles. *Information Processing and Management*, In press.

Appendix

A. News articles in the annotated corpus

Date	News agency	Title
27 Aug 1979	BBC News	<i>Soldiers die in Warrenpoint massacre</i>
01 Mar 1982	Newsweek	<i>GUATEMALA: NO CHOICES</i>
15 Sept 1982	Bangkok Post	<i>ISRAELIS RETURN TO WEST BEIRUT</i>
15 Sept 1982	The New York Times	<i>GEMAYEL OF LEBANON IS KILLED IN BOMB BLAST AT PARTY OFFICES</i>
12 July 1984	International Herald Tribune	<i>Lebanese Committee Named to Secure Release of Moslem, Christian Hostages</i>
12 July 1984	The Times	<i>SHULTZ JOINS CRITICS OF INDONESIAN RULE</i>
05 Dec 1984	International Herald Tribune	<i>U.S.-Backed Coalition Wins Grenada Election</i>
05 Dec 1984	The Guardian London	<i>Reagan favourite sweeps Grenada</i>
05 Dec 1984	The Guardian	<i>Blaize the American way</i>
01 Apr 1990	Dominion Sunday Times	<i>Troops take over Lithuanian office</i>
02 Apr 1990	The Dominion	<i>US Troops ambushed in Honduras</i>
20 Nov 1995	BBC News	<i>Diana admits adultery in TV interview</i>
06 Feb 1997	BBC News	<i>Widow allowed dead husband’s baby</i>
16 Oct 2006	Manchester Evening News	<i>BBC move to Salford may be delayed a year</i>
18 Oct 2006	Manchester Evening News	<i>Red faces over BBC’s Salford radio blunder</i>
10 Nov 2006	The Bolton News	<i>Councils urge BBC to move north</i>
05 Dec 2006	Manchester Evening News	<i>BBC boss confident of Salford Quays move</i>
21 Dec 2018	BBC News	<i>Cheshire BMW driver jailed over speeding ticket lie</i>
23 Dec 2018	BBC News	<i>Shrewsbury Christmas Festival cancelled after one day</i>
23 Dec 2018	BBC News	<i>Uppermill death: Men charged with Daniel Hogan murder</i>
28 Dec 2018	The New York Times	<i>Students Defiant as Chinese University Cracks Down on Young Communists</i>

B. Bell's most specific news schema categories, with examples drawn from "Students Defiant as Chinese University Cracks Down on Young Communists" by Javier C.Hernandez, The New York Times, 28 December 2018.

Category	Definition	Example
Action	Central or main action	<i>More than a dozen students from Peking University in Beijing, in a rare rebuke of authority, protested Friday on campus to draw attention to the university's attempts to punish students for taking part in the campaign.</i>
Reaction	A response to the Action that was expressed verbally, e.g., a direct/indirect quote, speech, interview	<i>Peking University officials moved swiftly to contain Friday's protest, holding the students in classrooms and keeping them through the night for questioning, activists said.</i>
Consequence	An action that transpired as a result of the Action, excluding verbal responses	<i>The students have put the government in an awkward position because they are invoking the teachings of Mao, Marx and Lenin, which President Xi Jinping has championed, to point to problems in Chinese society including inequality, corruption and greed.</i>
Context	Additional information that help explain or clarify details surrounding the Action	<i>The students are part of a small but tenacious group of young communists using leftist ideology to shine a light on labor abuses across China and to call for better protections for the working class.</i>
Evaluation	An assessment of the significance of the Action	<i>The stern reaction by the authorities reflects the party's deep anxieties about the young communists and their unusual campaign.</i>
Expectation	A view on what could happen after the Action	<i>Party leaders may be concerned that the 30th anniversary of the massacre, coming up in June, could inspire new protests.</i>
Previous episode	Events that happened more recently (in the near past)	<i>The protest on Friday came after Peking University officials tried to block a Marxist student group from organizing a celebration for Mao's 125th birthday.</i>
History	Set of events that happened before the near past	<i>The party has long feared student-led protests, especially since the 1989 pro-democracy movement, which had deep student involvement and was crushed in a bloody crackdown around Tiananmen Square.</i>