"Making the News": Identifying Noteworthy Events in News Articles

Shyam Upadhyay and **Christos Christodoulopoulos** and **Dan Roth**
Department of Computer Science
University of Illinois at Urbana-Champaign
{upadhya3,christod,danr}@illinois.edu

Abstract

Most events described in a news article are background events – only a small number are noteworthy, and a even smaller number serve as the trigger for writing of that article. Although these events are difficult to identify, they are crucial to NLP tasks such as first story detection, document summarization and event coreference, and to many applications of event analysis that depend on event counting and identifying trends. In this work, we introduce the notion of *news-peg*, a concept borrowed from the political science literature, in an attempt to remedy this problem. A news-peg is an event which prompted the author to write the article, and it serves as a more fine-grained measure of noteworthiness than a summary. We describe a new task of news-peg identification and release an annotated dataset for its evaluation. We formalize an operational definition of a news-peg, on which human annotators achieve high inter-annotator agreement (over 80%), and present a rule-based system for this task, which exploits syntactic features derived from established journalistic devices.

1 Introduction

The narratives in news articles often follow certain established styles, such as *inverted pyramid* reporting, to emphasize certain parts more than others. Such narratives nicely illustrate discourse level texture – parts which supply the main points and are crucial to conveying information constitute the *foreground*, while the parts which assist in providing supporting facts or setting the scene are referred to as *background* (Hopper and Thompson, 1980).

> Grigory Pasko, crusading Russian journalist who <u>documented</u> Russian Navy's <u>mishandling</u> of nuclear waste, is **<u>released</u>** on parole after <u>serving</u> two-thirds of his four-year prison sentence.

Figure 1: A human-generated summary from the New York Times Corpus with event predicates and the news-peg shown <u>underlined</u> and in **bold** respectively.

When reasoning about events in news articles, such document level texture necessitates the ability to distinguish foreground objects from background. Event recognition alone is not sufficient to make such distinctions, and the roles these events play in framing of a given story. Consider the summary shown in Figure 1 – it is easy to infer that the "release" event is the reason why the news article was written, and other events are present merely to qualify the entities present. Distinguishing the "release" event from others can help understanding what the document is reporting.

This paper introduces *news-peg identification*: a new task aimed at finding events which triggered the creation of the news article. Our notion of news-peg serves (formally defined in §4) as a measure of noteworthiness, assessing how much was the event responsible in prompting the author to write the article. Such events are also called dominant news elements in the social science literature. Note that news-pegs determine a stronger measure of noteworthiness than summaries, as a summary can contain events which are not news-pegs. We discuss how the ability to identify news-pegs can aid progress in several NLP tasks.

The contributions of our work are as follows:

- We define a new task, news-peg identification, and annotate data for its evaluation.

- We experimentally demonstrate the feasibility of the task, by showing a high inter-annotator agreement of 81.3% on a manually annotated evaluation dataset of 100 documents.

- We also evaluate several baseline approaches which exploit syntax to identify news-pegs and propose a rule-based approach which attains a F1 score of 54.7 points.

2 Motivation

The knowledge of whether an event is a news-peg in a news article can prove useful for several tasks. We briefly describe a few application below:

Cross-Document Event Coreference The notion of a news-peg is closely related to that of a non-anaphoric entities in the task of *entity coreference*; instead of looking for intra-document event mentions however, we look across documents. During the lifetime of an event, it is a news-peg for a short duration near its origin, as it is likely that several news sources deem it newsworthy at that time. Therefore, if an event is a news-peg, it is unlikely that it will refer to an earlier event instance (in a different article). It has been shown that detecting whether an entity is non-anaphoric benefits entity-level coreference resolution (Peng et al., 2015; de Marneffe et al., 2015; Wiseman et al., 2015; Ng, 2004; Ng and Cardie, 2002) – we expect cross-document event coreference to benefit in a similar way from news-peg identification.

First Story Detection (FSD) Current approaches to FSD (Petrović et al., 2010; Petrović et al., 2012) use similarity metrics like Latent Semantic Hashing (Salakhutdinov and Hinton, 2009) and use an inverted index to compare it with $O(1)$ (≈ 1000) of the most recent documents. A shortcoming of this problem formulation is that they treat the entire document as a event and do not account for multiple events in the same document. This formulation works well with tweets (assuming most tweets describe a single event) but is unsatisfactory when working with news articles. If we allow for n events in a document (on average), the number of comparisons will be $O(n^2)$.

A news-peg, on the other hand, identifies the reported event and thus allows us to focus on the most noteworthy event in the article, bringing the number of comparisons back to $O(1)$. Using news-pegs we can also perform a heuristic pruning of this search space, by allowing the system to ignore documents which do not have a similar event as their news-peg. For instance, if a newly arrived article describes a bombing event, we can prune out all articles from the 1000 most recent articles whose news-peg was not a bombing event.

Event Linking Nothman et al. (2012) introduced event linking as the task of grounding a event mention (*referent*) to a article in a news archive that first reports it (*anchor article*). They noted that annotating a large corpus was impractical because of the under-specification of "newsworthiness" and because the same article can be the anchor for some events and not for others. Annotation effort can be significantly reduced by knowing what is being reported for the first time in a document as it narrows the set of possible events the document can be an anchor for. News-pegs can help in segmenting a document into events which are being reported for the first time, and events which convey background information.

Document Summarization Even though we propose to use summarization as a sub-routine in our system (§5), a good news-peg classifier can prove useful to a summarization system. Any good summary for a document must contain a reference to its news-peg, as it is the most noteworthy among all events in the document. Drawing on this intuition, we can use the presence/absence of a news-peg as a alternative measure of quality of a summarization. Moreover, extractive summarization (Carbonell and Goldstein, 1998) approaches can prune out sentences which do not contain (or refer to) the news-peg, thereby reducing the search space of sentences from which the summary is constructed.

NLP applications such as Event Timeline Construction (Do et al., 2012) and headline generation (Woodsend and Lapata, 2012; Alfonseca et al., 2013) can benefit similarly from news-pegs.

3 Related Work

Attempts to distinguish foreground and background regions in text date back to the 1980s. Decker (1985) generated summaries from newspaper reports, where they used deterministic syntactic rules to label foreground events. These rules were based on predictable reporting styles in journalism such as the *inverted pyramid* and *block paragraph*[1], and drew heavily on the syntactic correlation between grounding and information content. We analyze the performance of these rules for news-peg identification in our experiments (§6).

The study of dominant elements of discourse has been formally studied in linguistics as a part of *centering theory* (Grosz et al., 1995), a broader theory of attention and coherence in discourse, both of which were analyzed on a document-level basis (i.e. local discourse). The authors suggested the use of centering constructs to keep track of the key entities, which change with discourse.

Document-level importance of entities (which include events) was explored by Gamon et al. (2013). The authors use the term *salience* to denote entity importance and graded entities into 3 categories – most salient, less salient, not salient. They extracted supervision from web-search logs to semi-automatically obtain noisy salience judgments for a large web corpus. Salient entities in a web document were then identified using graph centrality measures.

Our event extraction approach (§4.1) closely resembles the Open-IE event extraction approach (Fader et al., 2011; Hu et al., 2013; Do et al., 2011) which views events as sentence-level relations. Events are extracted via syntactic and lexical constraints, which are imposed on sentence level structure, such as dependency parse. For example, Sun et al. (2015) use the *nsubj* and *dobj* relations to identify relation pairs, which are then merged if they share the same predicate to form a `(Subj,Pred,Obj)` tuple expressing an event. Unlike traditional event paradigms like ACE (NIST, 2004) and ERE (ERE, 2013), the Open-IE event paradigm enjoys portability and domain-independence.

---

[1]Also known as *nut-paragraph*.

4 News-Peg Definition

In this section we define what constitutes an event and then formally describe the criteria to determine if an event is a news-peg. We use the example sentences marked with events and news-pegs in Table 1 to help us illustrate our definition.

4.1 Event Extraction

Definition 1 *An event is a predicate-argument structure, whose predicate (verbal or nominal) describes a single occurrence (eg. died, married) or an aggregate of occurrences (eg. elections, shootings etc.).*

We adopt a event extraction approach based on FrameNet (Baker et al., 1998). We automatically generate a set of *acceptable frames* from FrameNet which are associated with events of our interest. This list dictates the frames of the occurrences that we will consider events. For each predicate-argument structure, we identify the frame evoked by the predicate, and accept it as an event if the triggered frame belongs to the set of acceptable frames.

4.2 News-Peg Definition

Team of French archaeologists **work** at piece-by-piece reconstruction of ancient Baphuon temple in Siem Reap, Cambodia.
Kuwait's Interior Ministry says young Kuwaiti man who fled to Saudi Arabia after terrorist shooting in Kuwait that killed one American and wounded another has **confessed** to the attack.
Three Israeli soldiers are **shot** and **killed** in what Israeli officials describe as **ambush** by Palestinian gunmen near West Bank city of Hebron.
Bush administration officials have **concluded** that international inspectors are unlikely to find tangible and irrefutable evidence that Iraq is hiding weapons of mass destruction so administration is **preparing** its own assessment that will rely heavily on evidence from Iraqi defectors.
United Nations Security Council took grueling nine weeks to **negotiate** Nov. 8 resolution to make Iraq give up its illegal weapons, and now United States may have to go courting again to secure votes of five countries that have just become non-permanent members.

Table 1: Example sentences where event predicates identified by our extraction approach are underlined, and news-peg predicates appear in bold.
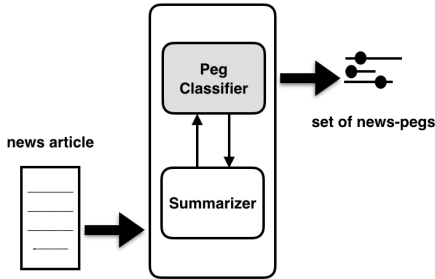
Figure 2: Schematic diagram of an end-to-end system. We evaluate the news-peg classifier (gray box), controlling for other sources of error by using a human generated summary. It is possible that the summarizer and news-peg classifier operate jointly.

Definition 2 *An event is a news-peg for an article if it was responsible in prompting the author to write the article.*

Two (or more) event predicates can be news-pegs if they are being co-reported (both are stated as if reported for the first time). For example, in Table 1, the third and fourth examples have events being co-reported (*shoot*, *kill* and *ambush* are co-reported, similarly concluded and preparing are co-reported). We allow a document to have multiple news-pegs, provided neither one of them heads a predicate which takes the other as an argument. For example, in Table 1, in the second example, the *fled*, *shooting* and *killed* events all are mentioned only to qualify the entity, and serve as elaboration for the news-peg, the *confessed* event. This ensures that events which appear under elaborative or subordinating clauses are not marked as news-pegs.

Reporting of events in news articles is often made indirectly by using journalistic devices such as *hedging* (example 2 in Table 1). Other complications arise from light verb constructions, such as the phrase "took grueling nine weeks to negotiate" in example 5 of Table 1. To mitigate this, we demand that the marked news-peg is not indirectly reporting another potentially more noteworthy news-peg. In case of light verb constructs, we consider the noun in the construct as the news-peg. So the news-peg in "took grueling nine weeks to negotiate" is "negotiate" instead of "took".

Note that our definition operates under the assumption that whether a event is an news-peg can be determined using document local discourse alone, without appealing to cross-document dis-

course. This assumption resonates with the local scoping assumption made by Gamon et al. (2013).

5 News-Peg Identification

We describe the task of identifying news-pegs as follows,

Input: A piece of text (news article/summary of a news article).

Output: A set of events in that piece of text that are the news-pegs.

Note that we intentionally do not specify if the piece of text is a document or a summary, because a document could have multiple news-pegs (albeit referents of the same event), which introduces the issue of event coreference (NIST, 2004). Ideally, the output would contain a cluster of events all of which refer to a news-peg. A possible strategy to circumvent this problem would be to first generate a summary (or the k-best summaries) and then obtain news-peg judgements per summary. Figure 2 shows a diagram for such an end-to-end system, where a news document is first summarized before identifying the news-peg.

When evaluating a news-peg classifier, using system-generated summaries would be unfair as we cannot judge whether the classifier or the summarizer is the source of error. To prevent this, we annotated the human generated summaries of the sampled documents, and ran the news-peg classifier on these summaries. While this is not the appropriate evaluation for an end-to-end system which would accept documents as input, we want to analyze the performance of the news-peg identification subroutine, for which summary level analysis suffices. Using the summary also precludes the need for event coreference.

5.1 Baselines

We use deterministic syntactic rules as baselines. To generate these features, we use the pipeline described in §6. We evaluate the following baseline approaches.

- **Active Voice**: If the event predicate appears in the active voice, mark it as the news-peg.

- **Main Clause**: If the event predicate appears in a main clause, mark it as the news-peg.

Algorithm 1 Rule-Based Classifier
___
Input: Event Predicate P
Output: Label L whether the predicate is a news-peg
 1: **if** Voice(P) = Active OR Clause(P) = Main **then**
 2:     $L \leftarrow 1$
 3: **end if**
 4: **if** Clause(P) = Relative **then**
 5:     $L \leftarrow 0$
 6: **end if**
 7: **if** Type(P) = Verb AND !Appositive(P) AND
    !FirstNonAppositiveVerbal(P) **then**
 8:     $L \leftarrow 0$
 9: **end if**
10: **return** L
___

- **First Predicate**: If the event predicate is the first predicate of the first sentence, mark it as the news-peg.

5.2 Rule-Based Classifier

We also developed a rule based system, shown in Algorithm 1, which uses a combination of these syntactic constructs used in the baselines to identify news-pegs. The classifier can be viewed as a decision list which deterministically predicts the label for a predicate. The primitives Appositive(P) detects if the predicate is embedded inside an appositive, while the primitive FirstNonAppositiveVerbal(P) returns true if the predicate is the first such verbal predicate which is not inside an appositive.

6 Experimental Analysis

We use the New York Times Annotated Corpus (Sandhaus, 2008) for all our experiments. The corpus contains around 650k documents annotated with human-generated summaries. We only work with the "World News" section of the corpus from 2003 to 2007. We randomly sampled 100 articles to be manually annotated and used for evaluation.

We generated the list of acceptable frames by identifying the frame evoked by each event trigger in ACE (NIST, 2004) and ERE (ERE, 2013). We use this list of frames to determine whether a predicate-argument structure qualifies as an event. For identifying the frame evoked by a given predicate, we use the state-of-the-art frame-identifier packaged in SEMAFOR (Das and Smith, 2011). For our event

extraction, all the documents were annotated using Illinois-SRL (Punyakanok et al., 2008), a state-of-the-art SRL system, to identify nominal and verbal predicates. Note that we do not take into account semantic role assignments, as we believe this does not have any consequence on event extraction.

To extract features, we use the constituency and dependency parser from Stanford CoreNLP (Manning et al., 2014) to identify the clause structure of sentences. Any clause labeled "SBAR" is considered subordinate, and a subordinate clause starting with a *Wh-* word is considered relative[2]. All remaining clauses are considered main. To identify appositions, we used the Illinois-Comma-SRL (Arivazhagan et al., 2016) package.

6.1 Annotation Setup

We obtained news-peg judgments using the Brat annotation tool (Stenetorp et al., 2012) from two annotators[3]. Each annotator was shown the human generated summary from New York Times Corpus, where event predicates detected by our event extraction procedure (§4.1) were highlighted. They were prompted with the definition of a news-peg from §4.2 and were instructed to mark the predicate(s) which they believed to indicate the news-peg(s). In case the news-peg was not one of the extracted events, the annotators were instructed to mark it.

Inter-annotator agreement was computed on how often the annotators mark the same predicate as the news-peg. The annotators achieved a high agreement of 81.3% on the 100 documents[4]. The judgments were then adjudicated by the first author. Out of 699 event predicates identified in the sampled 100 document summaries, only 103 were identified as news-pegs, accounting for only 14% of all verbal and nominal predicates.

6.2 Results

Table 2 shows the performance of the baselines and the rule based classifier on the manually annotated test data. The best performing system is the rule-

___
[2]*Wh*-words are what, who, whom, that, which, where, whose, whether

[3]Not the authors.

[4]We counted agreement only if annotators agreed on all news-peg predicate(s) in a document.

Approach	P	R	F1
Active Voice	31.0	54.3	39.4
Main Clause	22.2	**62.1**	32.7
First Predicate	41.0	37.9	39.4
Rule-Based Classifier	**68.1**	45.6	**54.7**

Table 2: Results comparing baselines and the rule based classifier. The classifier outperforms the baselines, but still achieves moderate F1 scores.

Three recent defectors from North Korea living in South Korea <u>draw</u> on their experience to **give** their own proposals for how to <u>deal</u> with unpredictable government of their impoverished homeland.
United Nations weapons inspectors in Iraq <u>make</u> dramatic <u>use</u> of their authority, **closing** exits and entrances to any site where they are <u>working</u> and **confining** thousands of people at <u>sprawling</u> government research complex, <u>including</u> Iraqi ambassador to UN, for almost six hours.

Table 3: Some hard examples from our test dataset where none of the compared approaches identified the correct news-peg (shown in **bold**).

based classifier, which is still far from human performance in terms of accuracy. It is evident that there is a lot of room for improvement.

A few examples that the rule-based classifier got wrong are shown in Table 3. The examples show that use of elaboration (eg. "draw on their experience") and light verb construction (eg. "make [. . .] use of authority") complicates the task. In the first example, the action being reported is the transfer of information from the defectors and what the information is about is auxilary information (therefore "deal" is not the news-peg). We believe that more discourse-aware features will be needed to improve performance.

7 Conclusion and Future Directions

In this paper, we proposed the new task of news-peg identification, and formalized a definition of news-pegs and shown its feasibility. We listed a number of applications which can benefit from a news-peg classifier. We also developed a rule-based news-peg classifier, which can be good basis for future work.

We also proposed an end-to-end system which relies on the synergy of a summarizer and news-peg classifier for identifying news-pegs. In future work,

we plan to investigate if jointly training a summarizer and a news-peg can actually improve performance of both systems. Also, the availability of human summaries from the New York Times Corpus opens the possibilities of exploring semi-supervised training approaches for news-peg classification.

References

Enrique Alfonseca, Daniele Pighin, and Guillermo Garrido. 2013. Heady: News headline abstraction through event pattern clustering. In *Proceedings of the 51st Annual Meeting of the ACL*, pages 1243–1253. Association for Computational Linguistics.

Naveen Arivazhagan, Christos Christodoulopoulos, and Dan Roth. 2016. Labeling the semantic roles of commas. In *AAAI*.

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet project. In *Proceedings of the 36th Annual Meeting of the ACL and 17th International Conference on Computational Linguistics - Volume 1*, ACL '98, pages 86–90, Stroudsburg, PA, USA. Association for Computational Linguistics.

Jaime Carbonell and Jade Goldstein. 1998. The use of MMR, diversity-based reranking for reordering documents and producing summaries. SIGIR '98, pages 335–336, New York, NY, USA. ACM.

Dipanjan Das and Noah A Smith. 2011. Semi-supervised frame-semantic parsing for unknown predicates. In *Proceedings of the 49th Annual Meeting of the ACL-HLT*, pages 1435–1444. Association for Computational Linguistics.

Marie-Catherine de Marneffe, Marta Recasens, and Christopher Potts. 2015. Modeling the lifespan of discourse entities with application to coreference resolution. *JAIR*, 52:445–475.

Nan Decker. 1985. The use of syntactic clues in discourse processing. In *Proceedings of the 23rd annual meeting on ACL*, pages 315–323. Association for Computational Linguistics.

Quang Xuan Do, Yee Seng Chan, and Dan Roth. 2011. Minimally supervised event causality identification. In *Proceedings of the Conference on EMNLP*, pages 294–303. Association for Computational Linguistics.

Quang Xuan Do, Wei Lu, and Dan Roth. 2012. Joint inference for event timeline construction. In *EMNLP*.

2013. Linguistic data consortium, DEFT ERE annotation guidelines: Events v1.1.

Anthony Fader, Stephen Soderland, and Oren Etzioni. 2011. Identifying relations for open information extraction. In *Proceedings of the Conference on EMNLP*, pages 1535–1545. Association for Computational Linguistics.

Michael Gamon, Tae Yano, Xinying Song, Johnson Apacible, and Patrick Pantel. 2013. Identifying salient entities in web pages. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, pages 2375–2380. ACM.

Barbara J. Grosz, Scott Weinstein, and Aravind K. Joshi. 1995. Centering: A framework for modeling the local coherence of discourse. *Comput. Linguist.*, 21(2):203–225, June.

Paul Hopper and Sandra A. Thompson. 1980. Transitivity in grammar and discourse. *Language*, 56:251–299.

Zhichao Hu, Elahe Rahimtoroghi, Larissa Munishkina, Reid Swanson, and A. Marilyn Walker. 2013. Unsupervised induction of contingent event pairs from film scenes. In *Proceedings of the 2013 Conference on EMNLP*, pages 369–379. Association for Computational Linguistics.

Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the ACL: System Demonstrations*, pages 55–60.

Vincent Ng and Claire Cardie. 2002. Identifying anaphoric and non-anaphoric noun phrases to improve coreference resolution. In *COLING 2002: The 19th International Conference on Computational Linguistics*.

Vincent Ng. 2004. Learning noun phrase anaphoricity to improve conference resolution: Issues in representation and optimization. In *Proceedings of the 42nd Annual Meeting of the ACL*.

NIST. 2004. The ACE evaluation plan.

Joel Nothman, Matthew Honnibal, Ben Hachey, and James R Curran. 2012. Event linking: Grounding event reference in a news archive. In *Proceedings of the 50th Annual Meeting of the ACL: Short Papers-Volume 2*, pages 228–232. Association for Computational Linguistics.

Haoruo Peng, Kai-Wei Chang, and Dan Roth. 2015. A joint framework for coreference resolution and mention head detection. In *CoNLL*, page 10, University of Illinois, Urbana-Champaign, Urbana, IL, 61801, 7. ACL.

Saša Petrović, Miles Osborne, and Victor Lavrenko. 2010. Streaming first story detection with application to twitter. In *The 2010 Annual Conference of the NAACL-HLT*, pages 181–189. Association for Computational Linguistics.

Saša Petrović, Miles Osborne, and Victor Lavrenko. 2012. Using paraphrases for improving first story detection in news and twitter. In *Proceedings of the 2012 Conference of the NAACL-HLT*, pages 338–346. Association for Computational Linguistics.

Vasin Punyakanok, Dan Roth, and Wen tau Yih. 2008. The importance of syntactic parsing and inference in semantic role labeling. *Computational Linguistics*, 34(2).

Ruslan Salakhutdinov and Geoffrey Hinton. 2009. Semantic hashing. *International Journal on Approximate Reasoning*, 50(7):969–978, July.

E. Sandhaus. 2008. The New York Times Annotated Corpus. *Linguistic Data Consortium, Philadelphia*, 6(12).

Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun'ichi Tsujii. 2012. Brat: A web-based tool for nlp-assisted text annotation. In *Proceedings of the Demonstrations at the 13th Conference of the EACL*, pages 102–107, Avignon, France, April. Association for Computational Linguistics.

Rui Sun, Yue Zhang, Meishan Zhang, and Dong-Hong Ji. 2015. Event-driven headline generation. In *Proceedings of the 53rd Annual Meeting of the ACL*, pages 462–472. Association for Computational Linguistics.

Sam Wiseman, M. Alexander Rush, Stuart Shieber, and Jason Weston. 2015. Learning anaphoricity and antecedent ranking features for coreference resolution. In *Proceedings of the 53rd Annual Meeting of the ACL*, pages 1416–1426. Association for Computational Linguistics.

Kristian Woodsend and Mirella Lapata. 2012. Multiple aspect summarization using integer linear programming. In *Proceedings of the 2012 Joint Conference on EMNLP*, pages 233–243, Jeju Island, Korea, July. Association for Computational Linguistics.