Event Linking: Grounding Event Reference in a News Archive

Joel Nothman[⊕] and **Matthew Honnibal**[+] and **Ben Hachey**[#] and **James R. Curran**[⊕]

[⊕]ə-lab, School of IT
University of Sydney
NSW, Australia

[⊕]Capital Markets CRC
55 Harrington St
Sydney
NSW, Australia

[+]Department of Computing
Macquarie University
NSW, Australia

[#]R&D, Thomson Reuters Corporation
St. Paul
MN, USA

{joel,james}@it.usyd.edu.au

{honnibal,ben.hachey}@gmail.com

Abstract

Interpreting news requires identifying its constituent events. Events are complex linguistically and ontologically, so disambiguating their reference is challenging. We introduce *event linking*, which canonically labels an event reference with the article where it was first reported. This implicitly relaxes coreference to *co-reporting*, and will practically enable augmenting news archives with semantic hyperlinks. We annotate and analyse a corpus of 150 documents, extracting 501 links to a news archive with reasonable inter-annotator agreement.

1 Introduction

Interpreting news requires identifying its constituent events. Information extraction (IE) makes this feasible by considering only events of a specified type, such as *personnel succession* or *arrest* (Grishman and Sundheim, 1996; LDC, 2005), an approach not extensible to novel events, or the same event types in sub-domains, e.g. sport. On the other hand, topic detection and tracking (TDT; Allan, 2002) disregards individual event mentions, clustering together articles that share a topic.

Between these fine and coarse-grained approaches, event identification requires grouping references to the same event. However, strict coreference is hampered by the complexity of event semantics: poison, murder and die may indicate the same effective event. The solution is to tag mentions with a canonical identifier for each news-triggering event.

This paper introduces *event linking*: given a past event reference in context, find the article in a news archive that first reports that the event happened. The task has an immediate practical application: some online newspapers link past event mentions to relevant news stories, but currently do so with low coverage and consistency; an event linker can add referentially-precise hyperlinks to news.

The event linking task parallels entity linking (NEL; Ji and Grishman, 2011), considering a news archive as a knowledge base (KB) of events, where each article exclusively represents the zero or more events that it first reports. Coupled with an appropriate event extractor, event linking may be performed for all events mentioned in a document, like the named entity disambiguation task (Bunescu and Paşca, 2006; Cucerzan, 2007).

We have annotated and analysed 150 news and opinion articles, marking references to past, newsworthy events, and linking where possible to canonical articles in a 13-year news archive.

2 The events in a news story

Approaches to news event processing are subsumed within broader notions of topics, scenario templates, or temporal entities, among others. We illustrate key challenges in processing news events and motivate event linking through the example story in Figure 1.

Salience Our story highlights carjackings and a police warning as newsworthy, alongside events like feeding, drove and told which carry less individual weight. Orthogonally, parts of the story are new events, while others are previously reported events that the reader may be aware of (illustrated in Figure 1). Online, the two background carjackings and the police warning are hyperlinked to other SMH articles where they were reported. Event schemas tend not to directly address salience: MUC-style IE (Gr-

Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, pages 228–232,
Jeju, Republic of Korea, 8-14 July 2012.

N	**Sydney man carjacked at knifepoint**
B	There has been another carjacking in Sydney, two weeks after two people were stabbed in their cars in separate incidents.
N	A 32-year-old driver was walking to his station wagon on Hickson Road, Millers Point, after feeding his parking meter about 4.30pm yesterday when a man armed with a knife grabbed him and told him to hand over his car keys and mobile phone, police said. The carjacker then drove the black 2008 Holden Commodore... He was described as a 175-centimetre-tall Caucasian...
B	Police warned Sydney drivers to keep their car doors locked after two stabbings this month. On September 4, a 40-year-old man was stabbed when three men tried to steal his car on Rawson Street, Auburn, about 1.20am. The next day, a 25-year-old woman was stabbed in her lower back as she got into her car on Liverpool Road...

Figure 1: Possible event mentions marked in an article from SMH, segmented into news (N) and background (B) event portions.

ishman and Sundheim, 1996) selects an event type of which all instances are salient; TDT (Allan, 2002) operates at the document level, which avoids differentiating event mentions; and TimeML (Pustejovsky et al., 2003) marks the main event in each sentence. Critiquing ACE05 event detection for not addressing salience, Ji et al. (2009) harness cross-document frequencies for event ranking. Similarly, reference to a previously-reported event implies it is newsworthy.

Diversity IE traditionally targets a selected event type (Grishman and Sundheim, 1996). ACE05 considers a broader event typology, dividing eight thematic event types (*business*, *justice*, etc.) into 33 subtypes such as *attack*, *die* and *declare bankruptcy* (LDC, 2005). Most subtypes suffer from few annotated instances, while others are impractically broad: sexual abuse, gunfire and the Holocaust each constitute *attack* instances (is told considered an *attack* in Figure 1?). Inter-annotator agreement is low for most types.[1] While ACE05 would mark the various *attack* events in our story, police warned would be unrecognised. Despite template adaptation (Yangarber et al., 2000; Filatova et al., 2006; Li et al., 2010; Chambers and Jurafsky, 2011), event types are brittle to particular tasks and domains, such as bio-text mining (e.g. Kim et al., 2009); they cannot reasonably handle novel events.

[1]For binary sentence classification, we calculate an interquartile range of $\kappa \in [0.46, 0.64]$ over the 33 sub-types. Coarse event type classification ranges from $\kappa = 0.47$ for *business* to $\kappa = 0.69$ for *conflict*.

Identity Event coreference is complicated by partitive (sub-event) and logical (e.g. causation) relationships between events, in addition to lexical-semantic and syntactic issues. When considering the relationship between another carjacking and grabbed, drove or stabbed, ACE05 would apply the policy: "When in doubt, do not mark any coreference" (LDC, 2005). Bejan and Harabagiu (2008) consider event coreference across documents, marking the "most important events" (Bejan, 2010), albeit within Google News clusters, where multiple articles reporting the same event are likely to use similar language. Similar challenges apply to identifying event causality and other relations: Bejan and Harabagiu (2008) suggest arcs such as feeding $\xrightarrow{\text{precedes}}$ walking $\xrightarrow{\text{enables}}$ grabbed – akin to instantiations of FrameNet's frame relations (Fillmore et al., 2003). However, these too are semantically subtle.

Explicit reference By considering events through topical document clusters, TDT avoids some challenges of precise identity. It prescribes *rules of interpretation* for which stories pertain to a seminal event. However, the carjackings in our story are neither preconditions nor consequences of a seminal event and so would not constitute a TDT cluster. TDT fails to account for these explicit event references. Though Feng and Allan (2009) and Yang et al. (2009) consider event dependency as directed arcs between documents or paragraphs, they generally retain a broad sense of topic with little attention to explicit reference.

3 The event linking task

Given an explicit reference to a past event, event linking grounds it in a given news archive. This applies to all events worthy of having been reported, and harnesses explicit reference rather than more general notions of relevance. Though analogous to NEL, our task differs in the types of expressions that may be linked, and the manner of determining the correct KB node to link to, if any.

3.1 Event-referring expressions

We consider a subset of newsworthy events – *things that happen and directly trigger news* – as candidate referents. In TimeML's event classification (Pustejovsky et al., 2003), newsworthy events would gen-

erally be *occurrence* (e.g. die, build, sell) or *aspectual* (e.g. begin, discontinue), as opposed to *perception* (e.g. hear), *intentional state* (e.g. believe), etc. Still, we are not confined to these types when other classes of event are newsworthy. All references must be explicit, reporting the event as factual and completed or ongoing.

Not all event references meeting these criteria are reasonably LINKABLE to a single article:

MULTIPLE many distinct events, or an event type, e.g. world wars, demand;

AGGREGATE emerges from other events over time, e.g. grew 15%, scored 100 goals;

COMPLEX an event reported over multiple articles in terms of its sub-events, e.g. 2012 election, World Cup, scandal.

3.2 A news archive as a KB

We define a canonical link target for each event: *the earliest article in the archive that reports the given event happened or is happening*. Each archival article implicitly represents zero or more related events, just as Wikipedia entries represent zero or one entity in NEL. Links target the story as a whole: closely related, *co-reported* events link to the same article, avoiding a problematically strict approach to event identity. An archive reports only selected events, so a valid target may not exist (NEL's NIL).

4 An annotated corpus

We link to a digital archive of the Sydney Morning Herald: Australian and international news from 1986 to 2009, published daily, Monday to Saturday.[2] We annotate a randomly sampled corpus of 150 articles from its 2009 *News and Features* and *Business* sections including news reports, op-eds and letters.

For this whole-document annotation, a single word of each past/ongoing, newsworthy event mention is marked.[3] If LINKABLE, the annotator searches the archive by keyword and date, selecting a target, *reported here* (a self-referential link) or NIL. An annotation of our example story (Figure 1) would produce five groups of event references (Table 1).

[2]The archive may be searched at `http://newsstore.smh.com.au/apps/newsSearch.ac`

[3]We couple marking and linking since annotators must learn to judge newsworthiness relative to the target archive.

Mentions	Annotation category / link
carjacking; grabbed [him]	LINKABLE, reported here
[were] stabbed; incidents; stabbings	MULTIPLE
[Police] warned	LINKABLE, linked: *Sydney drivers told: lock your doors*
[man] stabbed	LINKABLE, linked: *Driver stabbed after Sydney carjacking*
[woman] stabbed	LINKABLE, linked: *Car attack: Driver stabbed in the back*

Table 1: Event linking annotations for Figure 1

Agreement unit	AB	AC	JA	JB	JC
Token has a link	27	21	61	42	34
Link target on agreed token	48	73	84	83	74
Set of link targets per document	31	40	69	51	45
Link date on agreed token	61	80	87	93	89
Set of link dates per document	36	44	71	54	56

Table 2: Inter-annotator and adjudicator F_1 scores

All documents were annotated by external annotator A; external annotators B and C annotated 72 and 24 respectively; and all were adjudicated by the first author (J). Pairwise inter-annotator agreement in Table 2 shows that annotators infrequently select the same words to link, but that reasonable agreement on the link target can be achieved for agreed tokens.[4] Adjudicator-annotator agreements are generally much higher than inter-annotator agreements: in many cases, an annotator fails to find a target or selects one that does not first report the event; J accepts most annotations as valid. In other cases, there may be multiple articles published on the same day that describe the event in question from different angles; agreement increases substantially when relaxed to accept date agreement. Our adjudicated corpus of 150 documents is summarised in Table 3.

Where a definitive link target is not available, an annotator may erroneously select another candidate: an opinion article describing the event, an article where the event is mentioned as background, or an article anticipating the event.

The task is complicated by changed perspective between an event's first report and its later reference.

[4]$\kappa \approx F_1$ for the binary token task (F_1 accounts for the majority class) and for the sparse link targets/date selection.

Category	Mentions	Types	Docs
Any markable	2136	655	149
LINKABLE	1399	417	144
linked	501	229	99
reported here	667	111	111
nil	231	77	77
COMPLEX	220	79	79
MULTIPLE	328	102	102
AGGREGATE	189	57	57

Table 3: Annotation frequencies: no. of mentions, distinct per document, and document frequency

Can overpayed link to what had been acquired? Can 10 died be linked to an article where only nine are confirmed dead? For the application of adding hyperlinks to news, such a link might be beneficial, but it may be better considered an AGGREGATE.

The schema underspecifies definitions of 'event' and 'newsworthiness', accounting for much of the token-level disagreement, but not directly affecting the task of linking a specified mention to the archive. Adjectival mentions such as Apple's *new* CEO are easy to miss and questionably explicit. Events are also confused with facts and abstract entities, such as bans, plans, reports and laws. Unlike many other facts, events can be grounded to a particular time of occurrence, often stated in text.

5 Analysis and discussion

To assess task feasibility, we present bag-of-words (BoW) and oracle results (Figure 2). Using the whole document as a query[5] retrieves 30% of gold targets at rank 10, but only 60% by rank 150. Term windows around each event mention perform close to our oracle consisting of successful search keywords collected during annotation, with over 80% recall at 150. No system recalls over 30% of targets at 1-best, suggesting a reranking approach may be required.

Constraining search result dates is essential; annotators' constraints improve recall by 20% at rank 50. These constraints may draw on temporal expressions in the source article or external knowledge. Successful automated linking will therefore require extensive use of semantic and temporal information.

Our corpus also highlights distinctions between

[5]Using Apache Solr defaults: TFIDF-weighted cosine similarity over stemmed and stopped tokens.

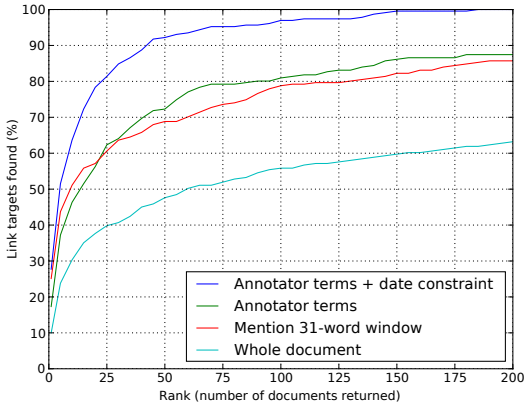

Figure 2: Recall for BoW and oracle systems

explicit event reference and broader relationships. Yang et al. (2009) makes the reasonable assumption that news events generally build on others that recently precede them. We find that the likelihood a linked article occurred fewer than d days ago reduces exponentially with respect to d, yet the rate of decay is surprisingly slow: half of all link targets precede their source by over 3 months.

The effect of coreporting rather than coreference is also clear: like {carjacking, grabbed} in our example, mention chains include {return, decide, recontest}, {winner, Cup} as well as more familiar instances like {acquired, acquisition}.

6 Conclusion

We have introduced *event linking*, which takes a novel approach to news event reference, associating each newsworthy past event with a canonical article in a news archive. We demonstrate task's feasibility, with reasonable inter-annotator agreement over a 150 document corpus. The corpus highlights features of the retrieval task and its dependence on temporal knowledge. As well as using event linking to add referentially precise hyperlinks to a news archive, further characteristics of news will emerge by analysing the graph of event references.

7 Acknowledgements

We are grateful to the reviewers for their comments. The work was supported by Capital Markets CRC post-doctoral fellowships (BH; MH) and PhD Scholarship (JN); a University of Sydney VCRS (JN); and ARC Discovery Grant DP1097291 (JRC).

References

James Allan, editor. 2002. *Topic Detection and Tracking: Event-based Information Organization*. Kluwer Academic Publishers, Boston, MA.

Cosmin Adrian Bejan and Sanda Harabagiu. 2008. A linguistic resource for discovering event structures and resolving event coreference. In *Proceedings of the 6th International Conference on Language Resources and Evaluation*, Marrakech, Morocco.

Cosmin Adrian Bejan. 2010. Private correspondence, November.

Razvan Bunescu and Marius Paşca. 2006. Using encyclopedic knowledge for named entity disambiguation. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics*, pages 9–16.

Nathanael Chambers and Dan Jurafsky. 2011. Template-based information extraction without the templates. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 976–986, Portland, Oregon, USA, June.

Silviu Cucerzan. 2007. Large-scale named entity disambiguation based on Wikipedia data. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 708–716.

Ao Feng and James Allan. 2009. Incident threading for news passages. In *CIKM '09: Proceedings of the 18th ACM international conference on Information and knowledge management*, pages 1307–1316, Hong Kong, November.

Elena Filatova, Vasileios Hatzivassiloglou, and Kathleen McKeown. 2006. Automatic creation of domain templates. In *Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions*, pages 207–214, Sydney, Australia, July.

Charles J. Fillmore, Christopher R. Johnson, and Miriam R. L. Petruck. 2003. Background to FrameNet. *International Journal of Lexicography*, 16(3):235–250.

Ralph Grishman and Beth Sundheim. 1996. Message understanding conference – 6: A brief history. In *COLING 1996 Volume 1: The 16th International Conference on Computational Linguistics*.

Heng Ji and Ralph Grishman. 2011. Knowledge base population: Successful approaches and challenges. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 1148–1158, Portland, Oregon, June.

Heng Ji, Ralph Grishman, Zheng Chen, and Prashant Gupta. 2009. Cross-document event extraction and tracking: Task, evaluation, techniques and challenges. In *Proceedings of Recent Advances in Natural Language Processing*, September.

Jin-Dong Kim, Tomoko Ohta, Sampo Pyysalo, Yoshinobu Kano, and Jun'ichi Tsujii. 2009. Overview of BioNLP'09 shared task on event extraction. In *Proceedings of the BioNLP 2009 Workshop Companion Volume for Shared Task*, pages 1–9, Boulder, Colorado, June.

LDC. 2005. ACE (Automatic Content Extraction) English annotation guidelines for events. Linguistic Data Consortium, July. Version 5.4.3.

Hao Li, Xiang Li, Heng Ji, and Yuval Marton. 2010. Domain-independent novel event discovery and semi-automatic event annotation. In *Proceedings of the 24th Pacific Asia Conference on Language, Information and Computation*, Sendai, Japan, November.

James Pustejovsky, José Casta no, Robert Ingria, Roser Saurí, Robert Gaizauskas, Andrea Setzer, and Graham Katz. 2003. TimeML: Robust specification of event and temporal expressions in text. In *Proceedings of the Fifth International Workshop on Computational Semantics*.

Christopher C. Yang, Xiaodong Shi, and Chih-Ping Wei. 2009. Discovering event evolution graphs from news corpora. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 34(4):850–863, July.

Roman Yangarber, Ralph Grishman, and Pasi Tapanainen. 2000. Automatic acquisition of domain knowledge for information extraction. In *In Proceedings of the 18th International Conference on Computational Linguistics*, pages 940–946.