# Cross-Document Non-Fiction Narrative Alignment

**Ben Miller♦, Jennifer Olive♠, Shakthidhar Gopavaram♣, and Ayush Shrestha♥**

♦Georgia State University, Departments of English and Communication
♠Georgia State University, Department of English
♣Georgia State University, Department of Computer Science
♥IEEE Member

miller@gsu.edu, jolive1@gsu.edu,
sgopavaram1@student.gsu.edu, ayush.shrestha@gmail.com

## Abstract

This paper describes a new method for narrative frame alignment that extends and supplements models reliant on graph theory from the domain of fiction to the domain of non-fiction news articles. Preliminary tests of this method against a corpus of 24 articles related to private security firms operating in Iraq and the Blackwater shooting of 2007 show that prior methods utilizing a graph similarity approach can work but require a narrower entity set than commonly occurs in non-fiction texts. They also show that alignment procedures sensitive to abstracted event sequences can accurately highlight similar narratological moments across documents despite syntactic and lexical differences. Evaluation against LDA for both the event sequence lists and source sentences is provided for performance comparison. Next steps include merging these semantic and graph analytic approaches and expanding the test corpus.

## 1 Introduction

Changing patterns of news consumption and circulation such as disconnecting individual articles from their bundled newspaper sources, sharing individual articles, and the increasing velocity of article generation all require techniques for building ad hoc collections of articles on emerging topics (Caswell, 2015). Identifying articles that describe similar events could help answer this challenge and show the narrative similarity of those sections. However, these moments of similarity can occur in small sections of those articles. An approach with a highly granular focus that identifies a coherent piece of narrative, generates a structured representation of that narrative unit, and compares it against a corpus would aid readers' efforts to find and follow stories across articles. A coherent narrative textual unit describes a section of text that can be segmented from its surroundings while still describing a possibility, an act, and a result, a definition consistent with (Bal, 1997). Research on aligning these sections, or narrative frames, has been pursued in various domains (Prud'hommeaux and Roark, 2012)(Miller et al., 2015)(Reiter, 2014); this paper describes preliminary

work extending that work to identify moments of narratological similarity but in the domain of non-fiction news articles.

To that end, we propose an expansion to a method for cross-document coreference of narrative units described in (Miller et al., 2015) that focused on the cross-document coreference of character and location entities. That method identified events in free text using EVITA (Saurí et al., 2005) then built adjacency matrices capturing entity-entity co-occurrence for each event. Similarity matrices were produced after combining the adjacency matrices and comparing the resulting story matrices using the Kronecker Product (Van Loan, 2000)(Weichsel, 1962) for sparse graph similarity measurements. Characters and locations were aligned by that method across stories based upon event-specific interaction patterns. This paper supplements that method with a process for better narrative segmentation and cross-document narrative correspondence identification. Frequently, these identifications lie four or more standard deviations from mean correspondence levels. These correspondences were found despite the narrative units crossing sentential boundaries, despite a high degree of semantic similarity across the corpus, and despite significant lexical and focal differences between the event descriptions. This work differs from other work in the domain of narrative/frame learning such as (Chambers and Jurafsky, 2009) in that it is sequence independent, does not connect entities and objects to roles, and focuses on discovering narrative situations for comparison rather than semantic role labeling. Like that example, the hypernym sequencing method described below does not rely on supervised techniques, hand-built knowledge, or pre-defined classes of events.

The test corpus is a set of articles related to Blackwater Worldwide. Blackwater (now Academi) is a private security company that has been contracted since 2003 by various American agencies to operate in Iraq. On September 16, 2007, Blackwater operatives killed 17 civilians and injured 20 more during an operation that went through Baghdad's Nisour Square. Articles on Blackwater approach their story from many angles. Some focus on the appearance of key Blackwater executives before congress. Others look to relate witnesses' perspectives on the massacre and contain translated quotes. Yet others summarize the trial that con-

victed four of the firm's private security officers for crimes committed during that event. The heterogeneity of the articles' foci on that event prevented the cross-document linking of specific event descriptions based on lexical features or with topic modeling algorithms. That challenge and the articles' connection to human rights violations, a persistent interest of the authors, drove the choice of corpus.

## 2 Methodology

Comparison of narrative frames requires the production of structured representations. The graph similarity method from the prior work, and the hypernym sequence comparison methods operate in parallel to produce structured representations of entities on a per-event basis, and event similarity on a sliding window basis. Both processes begin with a set of $n$ articles to be segmented by event. This segmentation is done using EVITA as documented in (Miller et al., 2015). The result is a document segmented into a highly granular event sequence.

### 2.1 Event Segmentation and Direct Hypernym Sequences

EVITA uses statistical and linguistic approaches to identify and classify the language denoting orderable dynamic and stative situations (Llorens et al., 2010) and outperforms or is competitive with other event recognizers. EVITA's overall accuracy in event recognition was found by (Llorens et al., 2010) to be $80.12\% F_{-}\{\beta\} = 1$ over TimeBank with 74.03% precision and 87.31% recall.

Following granular segmentation, the key event word recognized by EVITA is lemmatized and a lookup is performed to WordNet for the word's direct hypernym. The word sense was chosen using the Simplified Lesk method (Vasilescu et al., 2004). Each event is automatically typified with a keyword from the source text, but not every keyword has an identified direct hypernym. If no hypernym match was returned, the event word is used; that substitution occurred for 16.3% of the $5,422$ events. Sequences of hypernyms were built to encompass enough events to be commensurate with narratological theory of possibility, event, and aftermath (Bal, 1997). After experimenting with different length sequences, it was found that sequences of hypernyms that contained a number of events 3 times the average number of events per sentence, or approximately 3 sentences long, captured a span long enough to exemplify the theory but short enough to be distinct. In the case of this corpus, a preliminary random sample of 9 articles contained $2,112$ events in $464$ sentences yielding an average of $4.55$ events per sentence, which when multiplied by 3 to match the narrative theory of possibility, event, and aftermath, $13.65$ events. Rounded up, our method yielded 14 events per sequence. Each sequence is offset by one event from the prior sequence, thereby producing a sliding, overlapping narrative unit

window that goes across sentential boundaries. Two examples of generated sequences are provided in Table 1.

| Sequence Number | Hypernym Sequence |
|---|---|
| 209 | talk, blast, disappoint, prevent, veto, surprise, blast, act, injure, veto, label, cease, blast, injure |
| 210 | blast, disappoint, prevent, veto, surprise, blast, act, injure, veto, label, cease, blast, injure, inform |

Table 1: Example of two consecutive hypernym sequences from article 1.

### 2.2 Corpus

Our non-fiction corpus consisted of 24 news articles related to the September 16, 2007, shooting of Iraqi civilians by Blackwater security officers in Nisour Square, the investigation of the company's activities in Iraq prior to this incident, the outcome of those investigations, and the context of private security firms in Iraq. The subset are from 11 distinct international sources and were published between October 2007 and January 2011. Those articles were a random subset of the 616 articles returned by Lexis-Nexis for the following search: "Blackwater" and "shooting" with a length of $1,000 - 1,750$ words. That sample was selected as it contained a key focal event. All 24 articles were processed for the graph similarity method, and a smaller sample of 9 articles were used for testing the hypernym sequence matching method. Processing a larger sample is feasible as the hypernym sequencing method is entirely automatic but would require implementing $k$-means or $k$-nearest neighbors to help identify the correspondences.

### 2.3 Construction of Adjacency Matrices

Named-entity recognition (NER) and anaphora resolution was performed to establish entities in each event. Four raters performed overlapping manual entity extraction and resolution as current NER tools such as Stanford CoreNLP were not precise enough with multi-word entities. NER and anaphora resolution lie outside the focus of this paper. Manual tagging was done according to an index of significant entities with corresponding unique reference codes. Significance was determined in the context of the corpus as entities mentioned multiple times across the corpus.

Using the entities listed in the index, individual event adjacency matrices were generated. These matrices record the presence or absence of entities in an event frame to show entity co-occurrence for every event. An example of a section of an adjacency matrix for article 1 is in Table 2. Each matrix is symmetrical with respect

|        | BWSO | BWEX | IrVi | IrCi | NiSq | BaGZ | Bagh | USDOS | BWCO |
|--------|------|------|------|------|------|------|------|-------|------|
| BWSO   | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0     | 0    |
| BWEX   | 0    | 1    | 1    | 0    | 0    | 0    | 0    | 1     | 1    |
| IrVi   | 0    | 1    | 1    | 0    | 0    | 0    | 0    | 1     | 1    |
| IrCi   | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0     | 0    |
| NiSq   | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0     | 0    |
| BaGZ   | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0     | 0    |
| Bagh   | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0     | 0    |
| USDOS  | 0    | 1    | 1    | 0    | 0    | 0    | 0    | 1     | 1    |
| BWCO   | 0    | 1    | 1    | 0    | 0    | 0    | 0    | 1     | 1    |

Table 2: Populated section of the co-occurrence adjacency matrix for article 1, event 53 (helping), from the sentence, "Prince disputed that, but said, 'If the government doesn't want us to do this, we'll go do something else.' Waxman also charged that the State Department acted as an 'enabler' for the company by **helping** it to cover up shootings and compensate victims"(Facts on File World News Digest, 2007)

to the number of entities identified in the articles. 12 events were extracted from the sentence and populated by 6 entities from the complete entity list and coding instructions as shown in Table 3.

| Code  | Important Entities               |
|-------|----------------------------------|
| BWSO  | Blackwater Security Operatives   |
| BWEX  | Blackwater Executives            |
| IrVi  | Iraqi Victims                    |
| IrCi  | Iraqi Civilians                  |
| NiSq  | Nisour Square / The Traffic Circle |
| BaGZ  | Baghdad's Green Zone             |
| Bagh  | Baghdad                          |
| USDOS | U.S. Department of State         |
| BWCO  | Blackwater                       |
| Witn  | Witnesses                        |
| IrOf  | Iraqi Officials                  |
| IrAg  | Iraqi Agency                     |
| AmOf  | American Officials               |
| AmAg  | American Agency                  |
| PrvSec | Private Security Firm           |
| IrSf  | Iraqi Security Forces            |
| Iraq  | Iraq                             |
| USA   | United States                    |
| USMi  | U.S. Military                    |

Table 3: Named entity list with codings

## 2.4 Creation of Similarity Matrices

With event hypernym sequences and event-specific adjacency matrices, we proceeded to determine similarity between narrative frames within our corpus. The adjacency matrix similarity measurement method used is as per (Miller et al., 2015), which was inspired by Blondel et al.'s HITS (Hyperlink-Induced Topic Search) algorithm (Blondel et al., 2004).

Hypernym sequence similarity of narrative units proceeded by pairwise comparison of all sequences across all articles. This process resulted in $2,188,573$ total comparisons that were scaled from $0$, indicating no overlap between sequences, to $1$, indicating identical

sequences. This comparison was order independent (i.e. the sequence "a, b, c" is equivalent to "c, b, a") and is simply a measure of the number of overlapping terms.

Entity similarity measurement proceeded according to the methodology detailed in (Miller et al., 2015). That methodology builds a $3D$ matrix of the adjacency matrices where the axes from these individual matrices compose the first two dimensions and the event number composes the third dimension. Events are sequentially numbered $1$ to $n$ on a per document basis. Those similarity graphs are then cross-factored using the Kronecker Product to assess possible cross-document entity-to-entity alignment. Our extension of that method to non-fiction intended to use that measure as a weighting factor for narrative unit alignment, but that procedure yielded a negative result as described below.

## 2.5 Evaluation

Comparison of the hypernym sequence matching method was done against LDA using Gibbs sampling for parameter estimation and inference. Sentences lemmatized with Stanford CoreNLP from the full corpus and the hypernym sequences from articles 1 to 9 were tested with both a 20 topic model and a 50 topic model using an alpha of $40/k$, a beta of $0.2$, and $2,000$ sample iterations. As this work is preliminary, no gold standard training data was produced for the comparison; topic model allocations were manually reviewed by three raters for coherence.

## 3 Preliminary Results and Discussion

Preliminary results revealed strong correspondences of narrative units across the corpus and suggests the viability of this method for cross-document narrative frame alignment. Negative results noted above in relation to the entity similarity measures suggest that it requires further development before application to non-fiction generally and news articles in particular.

## 3.1 Event Similarity

Comparing the degree of overlap of these sequences in a pairwise manner yielded a set of correspondence scores that were visualized with dissimilarity matrices as seen in Figure 1. High correspondence sequences were identified as those more than 3 standard deviations from the mean correspondence for each matrix. Discourse order of the hypernyms in the sequence is not considered by this process, as the system needs to be agnostic relative to aspects of focalization such as flashbacks or first-, second-, or third-person narration. Sentence groups encapsulating those sequences were returned via a lookup and manually verified.

Comparison of event sequences throughout the sample of the 9 articles within the corpus resulted in a comparison score mean of $0.212$ with a standard deviation of $0.123$ for $2, 188, 573$ total comparisons across $72$ unique article comparisons. Values more than 3 standard deviations from the mean were found to correctly indicate similarity of narrative units. In part, this occurred because using the hypernyms of the event words tagged by EVITA generalized each event's description and allowed for more meaningful cross-document event alignment. Analysis of these significant similarity scores showed sequence matches in multiple articles. One example was found in articles $1$, $6$, and $7$ within our corpus; the matching sequences are shown in Table 4.

Comparison of 6 and 7, as shown by the dissimilarity graph in Figure 1, found sequences $184$ and $185$ in article 6 and sequence $48$ in article 7 as $0.857$ similar. That graph is the pairwise comparison of each of the $227$ sequences from article 6 (columns) against each of the $231$ sequences from article 7 (rows). Values are color coded on as red to yellow to green along a 0-to-1 scale. Areas of similarity, such as the one just described that appears in the bottom left corner of figure 1, fade in and out of the background dissimilarity as the sequences move into increasing then decreasing alignment. Comparison of articles 1 and 7 found sequences $209$ and $210$ in article 1 and sequences $43$, $44$, $45$, $46$, and $47$ as $0.786$ similar. Rather than drop sharply, this high rate of similarity continues into sequence $48$ of article 7 with a $0.714$ similarity. The connection of these three similarity scores using article 7 as a vector for comparison indicates that the corresponding events are similar within each of the articles.

The original passages support this finding as each describes a car rolling forward, Blackwater security officers opening fire on the car, and subsequent fire on Iraqi civilians. The sentences from which these hypernym sequences were extracted are included in Table 4 with their associated article numbers and hypernym sequences.

## 3.2 Entity Similarity

Entity-to-entity graph similarity tests produced lower than expected similarity rates. These negative results,
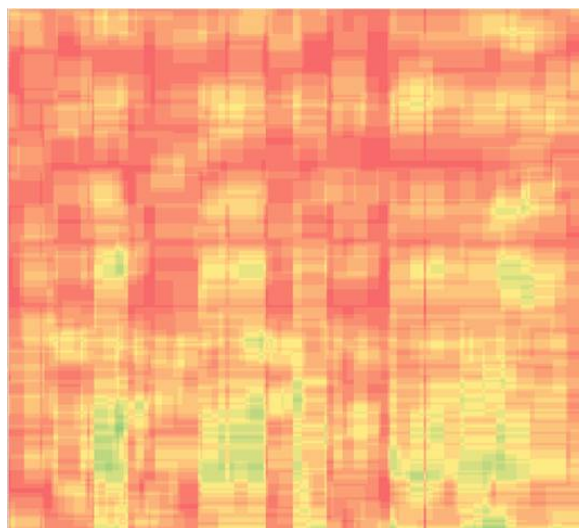


Figure 1: Dissimilarity graph showing the hypergram comparison across articles 6 and 7 using a color gradient scale where red indicates $< 50\%$, yellow indicates $50\%$, green indicates $> 50\%$ and up to $100\%$.

we theorize, occurred because non-fiction generally and news stories in particular feature more entities than fiction. That higher number of key entities led to more diverse entity co-occurrences and, therefore, more unique adjacency matrices. For our corpus, there were 27 unique entity sets with a mean of 6.6 occurrences per set and a standard deviation of 6.39. Without more significant overlap amongst the entity sets, the similarity analysis procedure yields sparsely populated graphs. The entity co-occurrences are too unique to compare with a large set of entities.

## 3.3 Findings

Despite the negative results in the entity similarity assessment portion, the core hypernym-based portion of this method correctly indicated cross-document similarity of narratives frames in a non-fiction corpus.

Most significantly, from a narratological perspective, the hypernym sequence model improved upon existing methodologies for cross-document narrative comparison in a manner consistent with narrative theory. This method operates at the clausal level, identifying the possibility, event, and outcome stages in a manner agnostic to sentential boundaries. This phenomenon can be seen in the similarity score between article 1 sequence 209 and article 7 sequences 43-47. As noted earlier, there is a slight drop in the similarity score as the narrative unit moves to sequence 48, which begins with the last events depicted at the end of a sentence: "Not one witness heard or saw // any gunfire coming from Iraqis around the square." In this example, the break between sequence 47 and 48 occurs at the "//", which was added for the purposes of this explanation. This slight decrease in similarity score and corresponding division of a sentence suggests that the events nar-

| Src. | Sq. | Hypernym Sequence | Source Sentence |
|---|---|---|---|
| 1 | 209 | talk, blast, disappoint, prevent, veto, surprise, blast, act, injure, veto, label, cease, blast, injure | "The shooting began at 12:08p.m., when at least one contractor began to fire on a car that failed to stop. The driver was killed and the car caught fire, but the contractors continued to shoot, killing the passengers and other Iraqis. At least one contractor reportedly called out to cease fire during the shooting, and another pointed his gun at a colleague" (Facts on File World News Digest, 2007). |
| 1 | 210 | blast, disappoint, prevent, veto, surprise, blast, act, injure, veto, label, cease, blast, injure, inform | |
| 6 | 184 | gunfire, express, perceive, perceive, blast, injure, express, blast, express, cut, affect, inspect, express, act | "All he saw, Sabah said, was that 'the white sedan moved a little bit and they started shooting.' As events unfolded and the Blackwater guards unleashed a storm of gunfire into the crowded square, Mr. Waso and Mr. Ali both said, they could neither hear nor see any return fire. 'It was one-sided shooting from one direction,' Mr. Waso said. 'There wasn't any return fire.' Mr. Waso said that what he saw was not only disturbing, but also in some cases incomprehensible. He said that the guards kept firing long after it was clear that there was no resistance" (Glanz, 2007). |
| 6 | 185 | express, perceive, perceive, blast, injure, express, blast, express, cut, affect, inspect, express, act, blast | |
| 7 | 43 | act, scat, injure, prevent, act, change_state, blast, injure, express, challenge, appear, injure, veto, talk | "The car continued to roll toward the convoy, which responded with an intense barrage of gunfire in several directions, striking Iraqis who were desperately trying to flee. Minutes after that shooting stopped, a Blackwater convoy – possibly the same one – moved north from the square and opened fire on another line of traffic a few hundred yards away, in a previously unreported separate shooting, investigators and several witnesses say. But questions emerge from accounts of the earliest moments of the shooting in Nisour Square. The car in which the first people were killed did not begin to closely approach the Blackwater convoy until the Iraqi driver had been shot in the head and lost control of his vehicle. Not one witness heard or saw any gunfire coming from Iraqis around the square" (Glanz and Rubin, 2007). |
| 7 | 44 | scat, injure, prevent, act, change_state, blast, injure, express, challenge, appear, injure, veto, talk, come | |
| 7 | 45 | injure, prevent, act, change_state, blast, injure, express, challenge, appear, injure, veto, talk, come, injure | |
| 7 | 46 | prevent, act, change_state, blast, injure, express, challenge, appear, injure, veto, talk, come, injure, suffer | |
| 7 | 47 | act, change_state, blast, injure, express, challenge, appear, injure, veto, talk, come, injure, suffer, perceive | |
| 7 | 48 | change_state, blast, injure, express, challenge, appear, injure, veto, talk, come, injure, suffer, perceive, cut | |

Table 4: Correspondences from articles 1, 6, and 7 with Hypernym Sequences and Source Sentences

rated in the first part of the sentence have a higher degree of similarity with sequence 109 in article 1. While the still significant score shows a relation between these two sets of sequences, it also shows the granularity at which the similarity assessments are made.

### 3.4 Future Work

While the automatic nature of the hypernym sequence comparison method will allow for it to scale, more sophisticated clustering techniques such as $k$-nearest neighbor will be needed to facilitate sequence similarity identification. Adapating the semantic role labling method from (Chambers and Jurafsky, 2009) might address the reliance of the graph simliarity method on insufficiently granular NER.

### 3.5 Evaluation

Evaluation of the hypernym sequence method against LDA proceeded as follows with the parameters as described above. The goal of this evaluation was to see whether the sequence method yielded more coherent clusters of meaningful narrative units. Each sentence was considered as one document. Using a java implementation of a Gibbs Sampling LDA method (Phan and Nguyen, 2006) on sentences that were lemmatized using Stanford CoreNLP, the corpus' $1,208$ sentences clustered into 20 topics with a mean of 78 sentences per topic and a standard deviation of 18.

Corresponding event sequences from the hypernym matching method did not perfectly align with the clustering of sentences proposed by LDA. In the three event-frame match across articles 1, 6, and 7, the hypernym method found a multi-sentence match across all three articles. LDA placed one of those sentences from article 1 and one sentence from article 6 in the same topic. Only one contributing sentence from each event frame was categorized into that topic. The surrounding sentences, though describing part of the same event, were identified as belonging to other topics. Briefly, narrative frames were not preserved – only semantic correspondences between individual sentences. LDA, by working at the document level, or in this case, at the sentence level, incorrectly preserves sentential boundaries in cases where narratives do not and does not allow for context to influence clustering. A narrative unit can begin in any clause of a sentence; tools for cross-document narrative coreference needs to work across sentential boundaries at the clausal level while still returning full sentence source texts to provide context. In our preliminary evaluations, LDA did not function as well as our hypernym sequence comparison.

## 4 Conclusion

Cross-document narrative unit similarity measurement is a promising area of research for the alignment of news articles. This successful preliminary work on abstracted event-keyword comparison based on event segmentation worked well in finding multi-sentence, statistically significant narrative unit correspondences across a small corpus of related articles. Extensions of an existing method for narrative alignment using graph similarity measures were not successful. We theorize this result because of the greater number of entities and intra-event entity sets that occur in non-fiction news reporting than in fiction. Future work looks to use the hypernym sequence comparison method to cluster events into narrative units, and then apply the entity co-occurrence method as a weighting factor for similarity measurement. While automatic NER would facilitate the integration of these two methods, a manual approach that focuses on the high similarity sections might curtail the task sufficiently to allow for it to remain feasible as the corpus size increases. We also plan to integrate $k$-means clustering into the analytic pipeline to facilitate identification of corresponding narrative units.

### Acknowledgments

# References

Mieke Bal. 1997. *Narratology: Introduction to the theory of narrative*. University of Toronto Press.

Vincent D. Blondel, Anahí Gajardo, Maureen Heymans, Pierre Senellart, and Paul Van Dooren. 2004. A measure of similarity between graph vertices: Applications to synonym extraction and web searching. *SIAM review*, 46(4):647–666.

DA Caswell. 2015. Structured narratives as a framework for journalism: A work in progress. In *Proceedings of the 6th Workshop on Computational Models of Narrative (CMN 2015)*.

Nathanael Chambers and Dan Jurafsky. 2009. Unsupervised learning of narrative schemas and their participants. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2*, pages 602–610. Association for Computational Linguistics.

Hector Llorens, Estela Saquete, and Borja Navarro-Colorado. 2010. Timeml events recognition and classification: learning crf models with semantic roles. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 725–733. Association for Computational Linguistics.

Ben Miller, Ayush Shrestha, Jennifer Olive, and Shakthidhar Gopavaram. 2015. Cross-document narrative frame alignment. In *2015 Workshop on Computational Models of Narrative*, volume 42, page forthcoming.

Xuan-Hieu Phan and Cam-Tu Nguyen. 2006. Jgibblda: A java implementation of latent dirichlet allocation (lda) using gibbs sampling for parameter estimation and inference.

Emily T Prud'hommeaux and Brian Roark. 2012. Graph-based alignment of narratives for automated neurological assessment. In *Proceedings of the 2012 Workshop on Biomedical Natural Language Processing*, pages 1–10. Association for Computational Linguistics.

Nils Reiter. 2014. Discovering structural similarities in narrative texts using event alignment algorithms.

Roser Saurí, Robert Knippen, Marc Verhagen, and James Pustejovsky. 2005. Evita: a robust event recognizer for qa systems. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 700–707. Association for Computational Linguistics.

Charles F Van Loan. 2000. The ubiquitous kronecker product. *Journal of computational and applied mathematics*, 123(1):85–100.

Florentina Vasilescu, Philippe Langlais, and Guy Lapalme. 2004. Evaluating variants of the lesk approach for disambiguating words.

Paul M Weichsel. 1962. The kronecker product of graphs. *Proceedings of the American Mathematical Society*, 13(1):47–52.