Recognizing Identical Events with Graph Kernels

Goran Glavaš and **Jan Šnajder**
University of Zagreb
Faculty of Electrical Engineering and Computing
Text Analysis and Knowledge Engineering Lab
Unska 3, 10000 Zagreb, Croatia
{goran.glavas,jan.snajder}@fer.hr

Abstract

Identifying news stories that discuss the same real-world events is important for news tracking and retrieval. Most existing approaches rely on the traditional vector space model. We propose an approach for recognizing identical real-world events based on a structured, event-oriented document representation. We structure documents as graphs of event mentions and use graph kernels to measure the similarity between document pairs. Our experiments indicate that the proposed graph-based approach can outperform the traditional vector space model, and is especially suitable for distinguishing between topically similar, yet non-identical events.

1 Introduction

News stories typically describe real-world events. Topic detection and tracking (TDT) aims to detect stories that discuss identical or directly related events, and track these stories as they evolve over time (Allan, 2002). Being able to identify the stories that describe the same real-world event is essential for TDT, and event-based information retrieval in general.

In TDT, an event is defined as something happening in a certain place at a certain time (Yang et al., 1999), while a topic is defined as a set of news stories related by some seminal real-world event (Allan, 2002). To identify news stories on the same topic, most TDT approaches rely on traditional vector space models (Salton et al., 1975), as more sophisticated natural language processing techniques have not yet proven to be useful for this task. On the other hand, significant advances in sentence-level event extraction have been made over the last decade, in particular as the result of standardization efforts such as TimeML (Pustejovsky et al., 2003a) and TimeBank (Pustejovsky et al., 2003b), as well as dedicated evaluation tasks (ACE, 2005; Verhagen et al., 2007; Verhagen et al., 2010). However, these two lines of research have largely remained isolated from one another.

In this paper we bridge this gap and address the task of recognizing stories discussing identical events by considering structured representations from sentence-level events. More concretely, we structure news stories into *event graphs* built from individual event mentions extracted from text. To measure event-based similarity of news stories, we compare their event graphs using graph kernels (Borgwardt, 2007). We conduct preliminary experiments on two event-oriented tasks and show that the proposed approach can outperform traditional vector space model in recognizing identical real-world events. Moreover, we demonstrate that our approach is especially suitable for distinguishing between topically similar, yet non-identical real-world events.

2 Related Work

The traditional vector space model (VSM) (Salton et al., 1975) computes the cosine between bag-of-words representations of documents. The VSM is at the core of most approaches that identify same-topic news stories (Hatzivassiloglou et al., 2000; Brants et al., 2003; Kumaran and Allan, 2005; Atkinson and Van der Goot, 2009). However, it has been observed that some word classes (e.g., named entities, noun phrases, collocations) have more significance than the others. Among them, named entities have been considered as particularly important, as they often identify the participants of an event. In view of this, Hatzivassiloglou et al. (2000) restrict the set of words to be used for document representation to words constituting noun phrases and named entities. Makkonen et

Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, pages 797–803,
Sofia, Bulgaria, August 4-9 2013. ©2013 Association for Computational Linguistics

al. (2004) divide document terms into four semantic categories (locations, temporal expressions, proper names, and general terms) and construct separate vector for each of them. Kumaran and Allan (2004) represent news stories with three different vectors, modeling all words, named-entity words, and all non-named-entity words occurring in documents. When available, recognition of identical events can rely on meta-information associated with news stories, such as document creation time (DCT). Atkinson and Van der Goot (2009) combine DCT with VSM, assuming that temporally distant news stories are unlikely to describe the same event.

In research on event extraction, the task of recognizing identical events is known as *event coreference resolution* (Bejan and Harabagiu, 2010; Lee et al., 2012). There, however, the aim is to identify sentence-level event mentions referring to the same real-world events, and not stories that discuss identical events.

3 Kernels on Event Graphs

To identify the news describing the same real-world event, we (1) structure event-oriented information from text into event graphs and (2) use graph kernels to measure the similarity between a pair of event graphs.

3.1 Event graphs

An event graph is a vertex- and edge-labeled mixed graph in which vertices represent individual event mentions and edges represent temporal relations between event mentions. We adopt a generic representation of event mentions, as proposed by Glavaš and Šnajder (2013): each mention consists of an *anchor* (a word that conveys the core meaning) and four types of *arguments* (agent, target, time, location). Furthermore, we consider four types of temporal relations between event mentions: *before*, *after*, *overlap*, and *equal* (Allen, 1983). As relations *overlap* and *equal* are symmetric, whereas *before* and *after* are not, an event graph may contain both directed and undirected edges.

Formally, an event graph G is represented as a tuple $G = (V, E, A, m, r)$, where V is the set of vertices, E is the set of undirected edges, A is the set of directed edges (arcs), $m : V \to M$ is a bijection mapping the vertices to event mentions, and $r : E \to R$ is the edge-labeling function, assigning temporal relations to edges (cf. Fig. 1).

The construction of an event graph from a news story involves the extraction of event mentions (anchors and arguments) and the extraction of temporal relations between mentions. We use a supervised model (with 80% F1 extraction performance) based on a rich set of features similar to those proposed by Bethard (2008) to extract event anchors. We then employ a robust, rule-based approach proposed by Glavaš and Šnajder (2013) to extract generic event arguments. Finally, we employ a supervised model (60% micro-averaged F1 classification performance) with a rich set of features, similar to those proposed by Bethard (2008), to extract temporal relations between event mentions. A detailed description of the graph construction steps is outside the scope of this paper.

To compute event graph kernels (cf. Section 3.2), we need to determine whether two event mentions co-refer. To resolve cross-document event coreference, we use the model proposed by Glavaš and Šnajder (2013). The model determines coreference by comparing factual event anchors and arguments of four coarse-grained semantic types (*agent*, *target*, *location*, and *time*), and achieves an F-score of 67% (79% precision and 57% recall) on the cross-document mention pairs from the EventCorefBank dataset (Bejan and Harabagiu, 2008). In what follows, $cf(m_1, m_2)$ denotes whether event mentions m_1 and m_2 co-refer (equals 1 if mentions co-refer, 0 otherwise).

3.2 Graph kernels

Graph kernels are fast polynomial alternatives to traditional graph comparison techniques (e.g., subgraph isomorphism), which provide an expressive measure of similarity between graphs (Borgwardt, 2007). We employ two different graph kernels: *product graph kernel* and *weighted decomposition kernel*. We chose these kernels because their general forms have intuitive interpretations for event matching. These particular kernels have shown to perform well on a number of tasks from chemoinformatics (Mahé et al., 2005; Menchetti et al., 2005).

Product graph kernel. A product graph kernel (PGK) counts the common walks between two input graphs (Gärtner et al., 2003). The graph product of two labeled graphs, G and $G^{'}$, denoted $G_P = G \times G'$, is a graph with the vertex set

$$V_P = \big\{(v, v') \mid v \in V_G, v' \in V_{G'}, \delta(v, v')\big\}$$

where $\delta(v, v')$ is a predicate that holds when vertices v and v' are identically labeled (Hammack et al., 2011). Given event graphs $G = (V, E, A, m, r)$ and $G' = (V', E', A', m', r')$, we consider the vertices to be identically labeled if the corresponding event mentions co-refer, i.e., $\delta(v, v') \doteq cf(m(v), m'(v'))$. The edge set of the graph product depends on the type of the product. We experiment with two different products: *tensor product* and *conormal product*. In the tensor product, an edge is introduced iff the corresponding edges exist in both input graphs and the labels of those edges match (i.e., both edges represent the same temporal relation). In the conormal product, an edge is introduced iff the corresponding edge exists in at least one input graph. Thus, a conormal product may compensate for omitted temporal relations in the input graphs.

Let A_P be the adjacency matrix of the graph product G_P built from input graphs G and G'. The product graph kernel that counts common walks in G and G' can be computed efficiently as:

$$K_{PG}(G, G') = \sum_{i,j=1}^{|V_P|} [(I - \lambda A_P)^{-1}]_{ij} \quad (1)$$

when $\lambda < 1/t$, where t is the maximum degree of a vertex in the graph product G_P. In our experiments, we set λ to $1/(t+1)$.

Weighted decomposition kernel. A weighted decomposition kernel (WDK) compares small graph parts, called *selectors*, being matched according to an equality predicate. The importance of the match is weighted by the similarity of the contexts in which the matched selectors occur. For a description of a general form of WDK, see Menchetti et al. (2005).

Let $S(G)$ be the set of all pairs (s, z), where s is the selector (subgraph of interest) and z is the context of s. We decompose event graphs into individual vertices, i.e., we define selectors to be the individual vertices. In this case, similarly as above, the equality predicate $\delta(v, v')$ for two vertices $v \in G$ and $v' \in G'$ holds if and only if the corresponding event mentions $m(v)$ and $m'(v')$ co-refer. Using selectors that consist of more than one vertex would require a more complex and perhaps a less intuitive definition of the equality predicate δ. The selector context Z_v of vertex v is a subgraph of G that contains v and all its immediate neighbors. In other words, we consider as context all event mentions that are in a direct temporal relation with the selected mention. WDK between event graphs G and G' is computed as:

$$K_{WD}(G, G') = \sum_{v \in V_G, v' \in V_{G'}} cf(m(v), m'(v'))\, \kappa(Z_v, Z'_{v'}) \quad (2)$$

where $\kappa(Z_v, Z'_{v'})$ is the *context kernel* measuring the similarity between the context Z_v of selector $v \in G$ and the context $Z'_{v'}$ of selector $v' \in G'$. We compute the context kernel κ as the number of coreferent mention pairs found between the contexts, normalized by the context size:

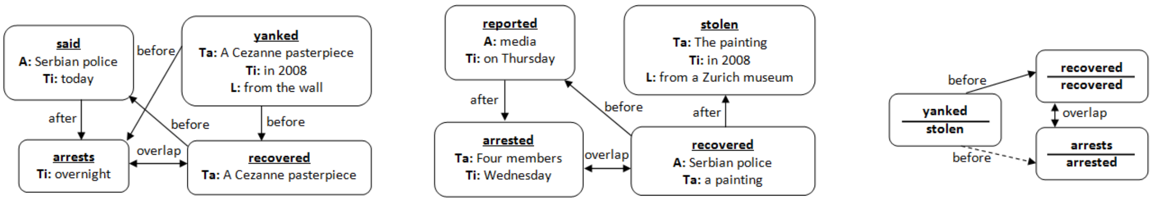
$$\kappa(Z_v, Z'_{v'}) = \frac{\sum_{w \in V_{Z_v}, w' \in V_{Z'_{v'}}} cf(m(w), m'(w'))}{\max(|V_{Z_v}|, |V_{Z'_{v'}}|)}$$

The intuition behind this is that a pair of coreferent mentions $m(v)$ and $m'(v')$ should contribute to the overall event similarity according to the number of pairs of coreferent mentions, $m(w)$ and $m'(w')$, that are in temporal relation with v and v', respectively.

Graph kernels example. As an example, consider the following two story snippets describing the same sets of real-world events:

> **Story 1:** *A Cezanne masterpiece worth at least $131 million that was the yanked from the wall of a Zurich art gallery in 2008 has been recovered, Serbian police said today. Four arrests were made overnight in connection with the theft, which was one of the biggest art heists in recent history.*
>
> **Story 2:** *Serbian police have recovered a painting by French impressionist Paul Cezanne worth an estimated 100 million euros (131.7 million U.S. dollars), media reported on Thursday. The painting "A boy in a red vest" was stolen in 2008 from a Zurich museum by masked perpetrators. Four members of an international crime ring were arrested Wednesday.*

The corresponding event graphs G and G' are shown in Fig. 1a and 1b, respectively, while their product is shown in Fig. 1c. There are three pairs of coreferent event mentions between G and G': (*yanked*, *stolen*), (*recovered*, *recovered*), and (*arrests*, *arrested*). Accordingly, the product graph P has three nodes. The dashed edge between vertices (*yanked*, *stolen*) and (*arrests*, *arrested*) exists only in the conormal product graph. By substituting into (1) the adjacency matrix and maximum vertex degree of tensor product graph P, we obtain



(a) Event graph G (Story 1) (b) Event graph G' (Story 2) (c) Product graph P

Figure 1: Example event graphs and their product

the tensor PGK score as:

$$K_{PG} = \sum_{i,j=1}^{3} \left[\left(I - \frac{1}{3} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \right)^{-1} \right]_{i,j} \approx 5.6$$

Similarly, for the conormal product graph P we obtain the conormal PGK score of $K_{PG} = 9$. By substituting G and G' into (2), we obtain the WDK score as:

$$K_{WD} = \sum_{(v,v') \in V_P} \kappa(Z_v, Z'_{v'}) = \frac{2}{3} + \frac{3}{4} + \frac{2}{4} \approx 1.9$$

where V_P contains pairs of coreferent event mentions: (*yanked*, *stolen*), (*recovered*, *recovered*), and (*arrests*, *arrested*).

4 Experiments

We conducted two preliminary experiments to investigate whether kernels on event graphs can be used to recognize identical events.

4.1 Task 1: Recognizing identical events

Dataset. In the first experiment, we classify pairs of news stories as either describing identical real-world events or not. For this we need a collection of stories in which pairs of stories on identical events have been annotated as such. TDT corpora (Wayne, 2000) is not directly usable because it has no such annotations. We therefore decided to build a small annotated dataset.[1] To this end, we use the news clusters of the EMM NewsBrief service (Steinberger et al., 2009). EMM clusters news stories from different sources using a document similarity score. We acquired 10 randomly chosen news clusters, manually inspected each of them, and retained in each cluster only the documents that describe the same real-world events. Additionally, we ensured that no documents from different clusters discuss the same event. To obtain the gold standard dataset, we build all pairs of documents. The final dataset consists of 64 documents in 10 clusters, with 195 news pairs from the same clusters (positive pairs) and 1821 news pairs from different clusters (negative pairs). We divide the dataset into a train and a test set (7:3 split ratio). Note that, although our dataset has ground-truth annotations, it is incomplete in the sense that some pairs of documents describing the same events, which were not recognized as such by the EMM, are not included. Furthermore, because EMM similarity score uses VSM cosine similarity as one of the features, VSM cosine similarity constitutes a competitive baseline on this dataset.

Model	P	R	F
Tensor PGK	89.7	82.3	85.8
Conormal PGK	89.3	77.8	83.2
WDK	88.6	73.7	80.5
SVM Graph	91.1	87.6	89.3
SVM Graph + VSM	**93.8**	**96.2**	**95.0**
VSM baseline	90.9	82.9	86.7

Table 1: Results for recognition of identical events

Results. For each graph kernel and the VSM baseline, we determine the optimal threshold on the train set and evaluate the classification performance on the test set. The results are given in Table 1. The precision is consistently higher than recall for all kernels and the baseline. High precision is expected, as clusters represent topically dissimilar events. PGK models (both tensor and conormal) outperform the WDK model, indicating that common walks correlate better to event-based document similarity than common subgraphs. Individually, none of the graph kernels outperforms the baseline. To investigate whether the two kernels complement each other, we fed the

[1]Datasets for both experiments are available at: `http://takelab.fer.hr/evkernels`

Original
"Taliban militants have attacked a prison in north-west Pakistan, freeing at least 380 prisoners... "
Event-preserving paraphrase
"Taliban militants in northwest Pakistan attacked the prison, liberated at least 380 prisoners ... "
Event-shifting paraphrase
"Taliban militants have been arrested in north-west Pakistan. At least 380 militants have been arrested... "

Table 2: Event paraphrasing example

individual kernel scores to an SVM model (with RBF kernel), along with additional graph-based features such as the number of nodes and the number of edges (*SVM graph* model). Finally, we combined the graph-based features with the VSM cosine similarity (*SVM graph + VSM* model). *SVM graph* model significantly (at $p < 0.05$, student's 2-tailed t-test) outperforms the individual kernel models and the baseline. The combined model (*SVM graph + VSM*) significantly (at $p < 0.01$) outperforms the baseline and all kernel models.

4.2 Task 2: Event-based similarity ranking

Dataset. In the second experiment we focus on the task of distinguishing between news stories that describe topically very similar, yet distinct events. For this purpose, we use a small set of event paraphrases, constructed as follows. We manually selected 10 news stories from EMM NewsBrief and altered each of them to obtain two meaning-preserving (event-preserving) and two meaning-changing (event-shifting) paraphrases. To obtain the meaning-preserving paraphrases, we use Google translate and round-trip translation via two pairs of arbitrarily chosen languages (Danish/Finnish and Croatian/Hungarian). Annotators manually corrected lexical and syntactic errors introduced by the round-trip translation. To obtain meaning-changing paraphrases, we asked human annotators to alter each story so that it topically resembles the original, but describes a different real-world event. The extent of the alteration was left to the annotators, i.e., no specific transformations were proposed. Paraphrase examples are given in Table 2. The final dataset consists of 60 news pairs: 30 positive and 30 negative.

Results. For each method we ranked the pairs based on the assigned similarity scores. An ideal method would rank all positive pairs above all negative pairs. We evaluated the performance using two different rank evaluation metrics: R-precision (precision at rank 30, as there are 30 positive pairs) and average precision. The performance of graph kernel models and the VSM baseline is given in Table 3. We tested the significance of differences using stratified shuffling (Yeh, 2000). When considering average precision, all kernel models significantly (at $p < 0.01$) outperform the baseline. However, when considering R-precision, only the conormal PGK model significantly (at $p < 0.05$) outperforms the baseline. There is no statistical significance in performance differences between the considered kernel methods. Inspection of the rankings reveals that graph kernels assign very low scores to negative pairs, i.e., they distinguish well between textual representations of topically similar, but different real-world events.

Model	R-prec.	Avg. prec.
Tensor PGK	86.7	96.8
Conormal PGK	**93.3**	**97.5**
WDK	86.7	95.7
VSM baseline	80.0	77.1

Table 3: Results for event-based similarity ranking

5 Conclusion

We proposed a novel approach for recognizing identical events that relies on structured, graph-based representations of events described in a document. We use graph kernels as an expressive framework for modeling the similarity between structured events. Preliminary results on two event-similarity tasks are encouraging, indicating that our approach can outperform traditional vector-space model, and is suitable for distinguishing between topically very similar events. Further improvements could be obtained by increasing the accuracy of event coreference resolution, which has a direct influence on graph kernels.

The research opens up many interesting directions for further research. Besides a systematic evaluation on larger datasets, we intend to investigate the applications in event tracking and event-oriented information retrieval.

Acknowledgments

This work has been supported by the Ministry of Science, Education and Sports, Republic of Croatia under the Grant 036-1300646-1986. We thank the reviewers for their constructive comments.

References

ACE. 2005. Evaluation of the detection and recognition of ACE: Entities, values, temporal expressions, relations, and events.

James Allan. 2002. *Topic Detection and Tracking: Event-based Information Organization*, volume 12. Kluwer Academic Pub.

James Allen. 1983. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843.

Martin Atkinson and Erik Van der Goot. 2009. Near real time information mining in multilingual news. In *Proceedings of the 18th International Conference on World Wide Web*, pages 1153–1154. ACM.

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 (LREC 2008)*.

Cosmin Adrian Bejan and Sanda Harabagiu. 2010. Unsupervised event coreference resolution with rich linguistic features. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1412–1422. Association for Computational Linguistics.

Steven Bethard. 2008. *Finding Event, Temporal and Causal Structure in Text: A Machine Learning Approach*. Ph.D. thesis, University of Colorado at Boulder.

Karsten Michael Borgwardt. 2007. *Graph Kernels*. Ph.D. thesis, Ludwig-Maximilians-Universität München.

Thorsten Brants, Francine Chen, and Ayman Farahat. 2003. A system for new event detection. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 330–337. ACM.

Thomas Gärtner, Peter Flach, and Stefan Wrobel. 2003. On graph kernels: Hardness results and efficient alternatives. In *Learning Theory and Kernel Machines*, pages 129–143. Springer.

Goran Glavaš and Jan Šnajder. 2013. Exploring coreference uncertainty of generically extracted event mentions. In *Proceedings of 14th International Conference on Intelligent Text Processing and Computational Linguistics*, pages 408–422. Springer.

Richard Hammack, Wilfried Imrich, and Sandi Klavžar. 2011. *Handbook of Product Graphs*. Discrete Mathematics and Its Applications. CRC Press.

Vasileios Hatzivassiloglou, Luis Gravano, and Ankineedu Maganti. 2000. An investigation of linguistic features and clustering algorithms for topical document clustering. In *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 224–231. ACM.

Giridhar Kumaran and James Allan. 2004. Text classification and named entities for new event detection. In *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 297–304. ACM.

Giridhar Kumaran and James Allan. 2005. Using names and topics for new event detection. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 121–128. Association for Computational Linguistics.

Heeyoung Lee, Marta Recasens, Angel Chang, Mihai Surdeanu, and Dan Jurafsky. 2012. Joint entity and event coreference resolution across documents. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 489–500. Association for Computational Linguistics.

Pierre Mahé, Nobuhisa Ueda, Tatsuya Akutsu, Jean-Luc Perret, and Jean-Philippe Vert. 2005. Graph kernels for molecular structure-activity relationship analysis with support vector machines. *Journal of Chemical Information and Modeling*, 45(4):939–951.

Juha Makkonen, Helena Ahonen-Myka, and Marko Salmenkivi. 2004. Simple semantics in topic detection and tracking. *Information Retrieval*, 7(3):347–368.

Sauro Menchetti, Fabrizio Costa, and Paolo Frasconi. 2005. Weighted decomposition kernels. In *Proceedings of the 22nd International Conference on Machine Learning*, pages 585–592. ACM.

James Pustejovsky, José Castano, Robert Ingria, Roser Sauri, Robert Gaizauskas, Andrea Setzer, Graham Katz, and Dragomir Radev. 2003a. Timeml: Robust specification of event and temporal expressions in text. *New Directions in Question Answering*, 3:28–34.

James Pustejovsky, Patrick Hanks, Roser Sauri, Andrew See, Robert Gaizauskas, Andrea Setzer, Dragomir Radev, Beth Sundheim, David Day, Lisa Ferro, et al. 2003b. The TimeBank corpus. In *Corpus Linguistics*, volume 2003, page 40.

Gerard Salton, Anita Wong, and Chung-Shu Yang. 1975. A vector space model for automatic indexing. *Communications of the ACM*, 18(11):613–620.

Ralf Steinberger, Bruno Pouliquen, and Erik Van Der Goot. 2009. An introduction to the european media monitor family of applications. In *Proceedings of the Information Access in a Multilingual World-Proceedings of the SIGIR 2009 Workshop*, pages 1–8.

Marc Verhagen, Robert Gaizauskas, Frank Schilder, Mark Hepple, Graham Katz, and James Pustejovsky. 2007. Semeval-2007 Task 15: TempEval temporal relation identification. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 75–80. Association for Computational Linguistics.

Marc Verhagen, Roser Sauri, Tommaso Caselli, and James Pustejovsky. 2010. Semeval-2010 Task 13: TempEval-2. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 57–62. Association for Computational Linguistics.

Charles Wayne. 2000. Multilingual topic detection and tracking: Successful research enabled by corpora and evaluation. In *Proceedings of the Second International Conference on Language Resources and Evaluation Conference (LREC 2000)*, volume 2000, pages 1487–1494.

Yiming Yang, Jaime G Carbonell, Ralf D Brown, Thomas Pierce, Brian T Archibald, and Xin Liu. 1999. Learning approaches for detecting and tracking news events. *Intelligent Systems and their Applications, IEEE*, 14(4):32–43.

Alexander Yeh. 2000. More accurate tests for the statistical significance of result differences. In *Proceedings of the 18th Conference on Computational linguistics*, pages 947–953. Association for Computational Linguistics.