Procesamiento del Lenguaje Natural, Revista nº 58, marzo de 2017, pp. 61-68 recibido 29-11-2016 revisado 12-01-2017 aceptado 07-02-2017

Cross-Document Event Ordering through Temporal Relation Inference and Distributional Semantic Models*

Ordenación de eventos multidocumento usando inferencia de relaciones temporales y modelos semánticos distribucionales

Estela Saquete, Borja Navarro-Colorado
Departamento de Lenguajes y Sistemas Informáticos
Universidad de Alicante
Carretera San Vicente s/n 03690 Alicante España
{stela,borja}@dlsi.ua.es

Abstract: This paper focuses on the contribution of temporal relations inference and distributional semantic models to the event ordering task. Our system automatically builds ordered timelines of events from different written texts in English by performing first temporal clustering and then semantic clustering. In order to determine temporal compatibility, an inference from the temporal relationships between events –automatically extracted from a Temporal Information Processing system– is applied. Regarding semantic compatibility between events, we analyze two different distributional semantic models: LDA Topic modeling and Word2Vec word embeddings. Both semantic models together with the temporal inference have been evaluated within the framework of SemEval 2015 Task 4 Track B. Experiments show that, using both models, the current State of the Art is improved, showing significant advance in the Cross-Document Event Ordering task.
Keywords: Temporal information, event coreference, temporal inference, distributional semantics, event ordering

Resumen: Este artículo se centra en estudiar la contribución que la inferencia de relaciones temporales y los modelos semánticos distribucionales hacen a la tarea de ordenación de eventos. Nuestro sistema construye automáticamente líneas de tiempo con eventos extraídos de diferentes documentos escritos en inglés. Para ello realiza primero una agrupación temporal y posteriormente una agrupación semántica. Para determinar la compatibilidad temporal se realiza una inferencia sobre las relaciones temporales entre los eventos extraídos de una sistema automático de procesamiento de información temporal. Para la compatibilidad semántica entre eventos hemos analizado dos modelos semánticos distribucionales distintos: LDA Topic Modeling y Word2Vec Word Embeddings. Ambos modelos semánticos junto con la inferencia temporal han sido evaluados bajo el marco de evaluación de SemEval 2015 Task 4 Track B. Los experimentos muestran que, usando ambos modelos se mejora el estado del arte actual, implicando un avance importante en la tarea de ordenación de eventos multidocumento.
Palabras clave: Información temporal, correferencia de eventos, inferencia temporal, semántica distribucional, ordenación de eventos

1 *Introduction*

Cross-document event ordering was the topic of the latest SemEval-2015 Task4 (Minard et al., 2015), called "TimeLine: Cross-Document Event Ordering". It consists of, first, extracting events involving a particular target entity among different documents, and, then, ordering them chronologically in a timeline.

Considering one specific entity as the target entity, all the events related to the target entity are extracted from several documents and arranged in a timeline.

The approach to cross-document event ordering presented in this paper is based on the idea that two or more event mentions corefer if they have not only temporal compatibility but

* This paper has been partially supported by the Spanish government, project TIN2015-65100-R, project TIN2015-65136-C2-2-R and PROMETEOII/2014/001.

ISSN 1135-5948

also semantic compatibility. In order to determine temporal compatibility, our approach uses temporal relationships between events extracted from a Temporal Information Processing system. Regarding semantic compatibility, in this paper we analyze two different distributional semantic models: LDA Topic Modeling (Blei, Ng, and Jordan, 2003) and Word2Vec Word Embeddings (Mikolov et al., 2013). We use this Semeval Timeline task as an evaluation and discussion framework.

This paper is organized as follows: next section (Section 2) presents the State of the Art in cross-document event ordering. Then, in Section 3, we explain our approach: how we formalize temporal and semantic compatibility, and how we apply distributional semantics to find coreferential events. Section 4 is devoted to the analysis, evaluation and discussion of the different approaches that have been implemented, and finally the main conclusions are presented in Section 5.

2 State of the Art

The most recent conferences about temporal information processing and temporal relation extraction were part of the SemEval challenges: *TempEval-1*, Temporal Relation Identification (Verhagen et al., 2007); *TempEval-2*, Evaluating Events, Time Expressions And Temporal Relations (Verhagen et al., 2010); and *TempEval-3*, Temporal Annotation (UzZaman et al., 2013). Challenges like the 6th i2b2 NLP Challenge (Sun, Rumshisky, and Uzuner, 2013) also focused on temporal relations but within a clinical context. All of theses challenges mainly focused on temporal relations of events, in order to: a) discover which of them occur before, simultaneously or after the others, and b) annotate all this temporal information (events, timex and relations) using the TimeML annotation scheme.

Regarding cross-document event coreference Bagga and Baldwin (1999) proposed one of the first approaches in this area. Ji et al., (2009) worked on a timeline task using the ACE 2005 training corpora. Bejan and Harabagiu (2014) performed cross- and within-document approaches using a rich set of linguistic features to model the event structure: lexical features such as head words and lemmas, class features such as PoS or event class, semantic features such as WordNet sense or semantic-roles frames, etc. Their proposal follows an unsupervised approach based on a non-parametrical Bayesian model. In the work presented by Li et al., (2011) the goal was to provide an event-fusion approach to obtain the most complete event possible by combining a set of coreference event mentions from different documents which were crawled from Websites. Another cross-document approach is proposed by Lee et al., (2012) introducing a novel coreference resolution system that models entities and events jointly. Cybulska and Vossen (2013) apply an event model based on four components: location, time, participant and action. They avoid the use of machine-learning methods in order to analyze how event components influence event coreference. Goyal et al., (2013) use a syntax-based distributional semantic approach on event coreference resolution. Lu and Ng (2016) present an event coreferent resolution system based on several sieves as similar lemmas, similar modifiers, hypernyms, etc. Finally, Yang et al., (2015) present a hierarchical distance-dependent Bayesian model for within- and cross-document event coreference resolution, concluding that it is a powerful approach to resolve the task in comparison to other state-of-the-art approaches.

Most recently, SemEval-2015 included the task that tried to combine temporal processing and event coreference in order to obtain a timeline of events related to a specific given entity from a set of documents (Minard et al., 2015). They proposed two different tracks on the basis of the data used as input. Track A, for which they provided only raw text sources, and Track B, for which they also made gold event mentions available. Track A had two participants: WHUNLP team and SPINOZAVU team. WHUNLP team processed the texts with Stanford CoreNLP[1] (Manning et al., 2014) and applied a rule-based approach to extract target entities and their predicates. They also performed temporal reasoning. The SPINOZAVU system (Caselli et al., 2015) is based on a pipeline developed in the NewsReader project. It addressed entity resolution, event detection, event-participant linking, coreference resolution, factuality profiling and temporal relation processing, first at document level, and then at cross-document level, in order to obtain timelines. Track B participants were the Heideltoul team and the GPLSIUA team. The Heideltoul approach (Moulahi et al., 2015) uses the HeidelTime tool for temporal information processing and the Standford CoreNLP for event coreference resolution. The GPLSIUA approach (Navarro-Colorado and Saquete, 2015) uses the OPENER language anal-

[1] http://stanfordnlp.github.io/CoreNLP/

ysis toolchain for entity detection, the TIPSem tool (Llorens, Saquete, and Navarro-Colorado, 2012) for temporal processing and two different approaches to detect event coreference: lexical semantics and a Topic Modeling algorithm over WikiNews corpus. Later works such as Laparra et al., (2015) showed that explicit temporal relations are not enough to obtain a full time-anchor annotation of events and evidenced the need for a temporal analysis at document level.

In this paper we reanalyze the impact of combining temporal inference with two of the (current) most important distributional models: LDA Topic Modeling (Blei, Ng, and Jordan, 2003) and Word2Vec Word Embeddings (Mikolov et al., 2013) in the cross-document event ordering task. In next section, we explain how both distributional semantic models have been adapted to event coreference resolution.

3 Our approach

Our approach to Cross-Document Event Ordering is based on the idea that two events *e1* and *e2* are coreferent if they have not only temporal compatibility but also semantic compatibility (they refer, in some way, to the same facts) (Navarro-Colorado and Saquete, 2016). Formally:

$$coref(e1, e2) \rightarrow (e1_t = e2_t) \wedge (e1_s \simeq e2_s)$$

From this idea, our system is structured in four steps:

1. Event and temporal relation extraction using Temporal Information Processing;
2. Target Entity Filtering in order to select the events related to the target entity;
3. Temporal clustering through temporal compatibility inference;
4. Semantic clustering using two different distributional semantic models: LDA Topic Modeling and Word2Vec Word Embeddings.

Each step is explained in depth in the next subsections.

3.1 Temporal Information Processing

The first module of the proposed architecture performs Temporal Information Processing using TIPSem system. It automatically annotates all the temporal information according to TimeML standard annotation scheme (Saurí et al., 2006), which means annotating all the temporal expressions (TIMEX3), events (EVENT) and links between them.

3.2 Target Entity Filtering

Considering that not all the events annotated by the previous module are necessary to build the timeline, but only the ones related to a target entity, a Target Entity Filtering needs to be performed in order to avoid those events that are annotated but not related to the given entity.

The Target Entity Filtering requires resolving name entity recognition and entity coreference resolution. Since this is not the main challenge of our research, this task is performed using an external tool. That is why the OPENER[2] web services were integrated in our proposal. More specifically, the NER and the coreference resolution component.

To determine if an event must be part of the timeline or not, this module selects the events in which a target entity (or a target entity coreference) explicitly participates in a *has_participant* relation with the semantic role A0 (i.e. agent) or A1 (i.e. patient), as defined in the Propbank Project (Palmer, Gildea, and Kingsbury, 2005).

3.2.1 Temporal Compatibility Clustering

As we have explained before, a Temporal Information Processing system such as TIPSem works at document level and is able to extract from each document all the explicit temporal information as well as establish temporal relations between times and events or between events. However, in order to establish a cross-document timeline of events, this is not enough. It is necessary to know explicitly the time at which each event occurs, and to perform cross-document event clustering.

One must infer the time-anchoring of all the selected events from the temporal information extracted by TIPSem in each document (within-document temporal inference). Through this inference, the temporal clustering of all the events of the different documents is performed (cross-document temporal inference). As previously stated, we consider two events to be clustered when they are temporally compatible, that is, if they happen at the same time.

Our model infers temporal compatibility in two steps:

[2] http://www.opener-project.eu/webservices

- **Within-document temporal clustering**: For each document, the temporal information of each event is going to be extracted. Each event is anchored to a time anchor[3] when a SIMULTANEOUS temporal link exists between this event and a temporal expression. After this, two events are considered part of the same cluster if they are temporally compatible, meaning that: a)they are anchored to the same time anchor, or b) they have a SIMULTANEOUS temporal link between them.
- **Cross-document temporal clustering**: From a set of documents (related by the same topic), and considering that in the previous step all the events were assigned a time anchor, all the events in the different documents that are temporally compatible, that is, are anchored to the same time anchor, are clustered together.

3.2.2 Semantic Compatibility Clustering: distributional semantics and event coreference

All those events occurring at the same time and being semantically compatible must be part of the same cluster in the timeline of a specific entity. The problem is that it is not exactly clear which components of the event structure are determinant in event coreferent resolution (Cybulska and Vossen, 2013).

Rather than creating a complex feature matrix to represent the semantics of the argument as Bejan and Harabagiu (2014) does, we propose a compact, use-based distributional representation of the semantics of the arguments. Moreover, contrary to Goyal et al., (2013), who use a syntax-based distributional representation, we use the argument structure of the event.

In this regard, when we apply distributional semantic models we are considering the context of the events as the main component that contributes to establish the semantic compatibility and, therefore, the event coreference. Current computational models of distributional semantics are based on the word/document model of Information Retrieval. In order to increase the semantic representativity of the vector space and to resolve data sparseness problems, different models have been proposed such as, among others, LSA Latent Semantic Analysis (Landauer and Dumais, 1997), LDA Latent Dirichlet Allocation (Topic Modeling) (Blei, Ng, and Jordan, 2003) or, recently, Word2Vec Word Embeddings (Mikolov et al., 2013). In this paper we apply Topic Modeling and Word Embeddings to the event coreference resolution task.

LDA Topic Modeling extracts a predefined set of topics from large corpora. Each topic is a distribution over a fixed vocabulary. In order to assign words to topics, LDA uses two values: the topic assigned to a word in other texts and the most frequent topic in the text where the word appears. Through several iterations, in the end the corpus is represented as a word-topic matrix, in which each topic is composed of the weight of each word in it. Through LDA Topic Modeling, a high dimensional vector space is reduced to a k topics vector space (Blei, Ng, and Jordan, 2003).

In contrast, Word Embeddings Word2Vec is a predictive model that works better with very high vector spaces. It learns about distributed representation of words based on neural networks. From the point of view of semantic representation, instead of trying to reduce the dimensionality of the vector space as LSA and LDA do, Word2Vec tries to optimize the representation of the context where a word appears: on one hand, through the continuos skip-gram model, Word2Vec maximizes relevant contexts; on the other hand, through negative sampling it assigns high probability to relevant words and low probability to noise words (Mikolov et al., 2013). Word2Vec does not use a linguistic-motivated context size: it applies a window of size k. LDA Topic Modeling, in contrast, tends to the establish the text as context, considering relevant the most frequent topic of each text to specify the topic of each word (Blei, Ng, and Jordan, 2003).[4]

We use distributional semantic models for event coreference resolution as follows. Firstly, we apply both LDA Topic Modeling and Word2Vec to English Wikipedia. That way, we obtain two distributional knowledge-bases in which each word is represented as a contextual vector in a high dimensional space.

Through LDA Topic Modeling, the distributional knowledge-base is a vector space made up of 500 topics. Each word is, then, represented as a 500-dimension vector in which each value is the weight of the word in each topic. Through Word2Vec the distributional knowledge-base is the embeddings of each word in a space of 1000

[3] A time anchor is always a DATE (as defined in TimeML) and its format follows the ISO-8601 standard: YYYY-MM-DD, being the maximum granularity admitted in the task DAY.

[4] For a more systematic comparison of distributional models, see (Baroni, Dinu, and Kruszewski, 2014).

dimensions.[5]

Secondly, each event structure is represented as a compositional vector. Applying a Part of Speech tagger and a Semantic Role labeling, the event structure is extracted.[6] It is made up of the nouns, verbs and adjectives of the event head and the main arguments[7]. Then the contextual vector of each word is extracted from the distributional knowledge-base (Topic Modeling Knowledge-base on one hand, Word2Vec knowledge-base on the other). Finally, following the additive model (Mitchell and Lapata, 2010), all word vectors are added up into a single compositional vector that represents the distributional meaning of the whole event structure.

Formally, the event structure is represented as a tuple of three elements: two arguments ($A0$ and $A1$) and one event head (H):

$$ES = < A0, A1, H >$$

Each argument is a compositional vector $\overrightarrow{V}(A)$ formed by the sum of the contextual vector $\overrightarrow{V}(w_n)$ of each word in the argument (w_n):

$$\overrightarrow{V}(A) = \sum^{n} \overrightarrow{V}(w_n)$$

The event head H is the contextual vector of a single word. Finally, the compositional vector of the whole event structure $\overrightarrow{V}(ES)$ is:

$$\overrightarrow{V}(ES) = \overrightarrow{V}(A0) + \overrightarrow{V}(A1) + \overrightarrow{V}(H)$$

Finally, in order to detect if two events are coreferential, the system calculates the cosine similarity between both event vectors. If the cosine similarity between two event vectors is higher than 0.9[8], the system concludes that there is a coreference between them and hence they are grouped together in the same cluster. Formally:

$$coref(\overrightarrow{V}(ES_1), \overrightarrow{V}(ES_2)) \Longrightarrow sim(\overrightarrow{V}(ES_1), \overrightarrow{V}(ES_2)) \geq 0.9$$

Event coreference is considered a transitive relation:

$$coref(\overrightarrow{V}(ES_a), \overrightarrow{V}(ES_b)) \wedge coref(\overrightarrow{V}(ES_b), \overrightarrow{V}(ES_c)) \Longrightarrow$$
$$coref(\overrightarrow{V}(ES_a), \overrightarrow{V}(ES_c))$$

4 Experiments and evaluation

In order to evaluate our approach, we have used the dataset provided for Task4 at SemEval 2015.[9] This dataset is composed of articles from Wikinews about three topics: 1) Airbus and Boeing; 2) General Motors, Chrysler and Ford; and 3) Stock Market. All the experiments shown in this section were performed using Track B input[10].

As a baseline, we have implemented a lexical (non-distributional) WordNet-based approach. With this baseline we assume that two events are coreferent if their event heads express the same concept. Therefore, two events are clustered together as coreferential if both event heads are the same word (that is, they have the same lemma), or both event heads are synonyms (that is, they share the same synset in WordNet).

Regarding the distributional semantic models applied to cross-document event ordering, the system has been run with six different configuration:

1. TC+TM0505: Temporal clustering + LDA Topic Modeling Semantic clustering considering the event head and the arguments in the same proportion.
2. TC+TM1000: Temporal clustering + LDA Topic Modeling Semantic clustering considering only distributional similarity between heads.
3. TC+TM0010: Temporal clustering + LDA Topic Modeling Semantic clustering considering only distributional similarity between arguments.
4. TC+W2V0505: Temporal clustering + Word2Vec Words Embedding Semantic clustering considering the event head and the arguments in the same proportion.
5. TC+W2V1000: Temporal clustering + Word2Vec Words Embedding Semantic clustering considering only distributional similarity between heads.
6. TC+W2V0010: Temporal clustering + Word2Vec Words Embedding Semantic clustering considering only distributional similarity between arguments.

The results are shown in Table 1. These data show the performance of the system according to the evaluation measures of Semeval2015-Task 4 Track B. Furthermore, the comparison between our results and the results obtained by

[5] In order to create these distributional knowledge bases, we have used Gensim (`https://radimrehurek.com/gensim/`) both to compute Topic Modeling and Word2Vec, and Wiki2Vec (`https://github.com/idio/wiki2vec`).

[6] We use a Python implementation of Collobert's SENNA (Collobert et al., 2011) (`https://pypi.python.org/pypi/practnlptools/1.0`).

[7] A0 and A1 following PropBank tagset (`https://verbs.colorado.edu/~mpalmer/projects/ace.html`).

[8] After some tests, we have settled a threshold of 0.9 over 1.

[9] http://alt.qcri.org/semeval2015/task4/index.php?id=data

[10] SemEval 2015 Task 4, two different tracks were proposed on the basis of the data used as input: Track A for which they provided only raw text sources, and Track B, for which they also provided available gold events mentions.

other systems can only be done against the GPLSIUA and HEIDELTOUL teams' outcome, since those were the teams that participated in the same Track at SemEval 2015-Task 4. Given that this is a very novel task, no other competition has been held.

Approach	Micro-F1	Micro-P	Micro-R
TC+LCV2 (baseline)	29.95	25.17	36.98
TC+TM0505	30.23	25.56	36.98
TC+TM1000	30.23	25.56	36.98
TC+TM0010	**30.28**	25.64	36.98
TC+W2V0505	**30.41**	25.82	36.98
TC+W2V1000	30.35	25.74	36.98
TC+W2V0010	**30.43**	25.85	36.98
GPLSI_Run1	25.36	21.73	30.46
HEIDELTOUL_Run2	18.34	13.58	28.23

Table 1: Cross-document Event Ordering Results

The best results were achieved by the Word2Vec Word Embeddings model, with an F1-measure higher than 30.40% both in the experiment considering the event head and the arguments in the same proportion, and in the experiment considering only distributional similarity of the event heads. Regarding Topic Modeling approaches, the best results for event ordering were achieved by the TM0010 experiment, in which only the similarity between arguments was taken into account. In any case, F1-score results are quite similar throughout the different distributional semantics models, with the same recall and very slight improvements in precision. The same recall is obtained because all the experiments share the same selection of events and temporal inference. These data show that, in this evaluation framework, there are no great differences between lexical-based and distributional-based models on one hand (only 0.5 points of improvement), and Topic Modeling and Word Embeddings on the other hand. This happens because the task is more focused on event ordering and the event clusters are quite small in general, therefore, the impact of the event coreference resolution modules is not outstanding.

In comparison with other approaches at Semeval2015-Task 4 competition, all our experiments outperform the state-of-the-art systems in all their runs and all metrics (see Table 1), with a remarkable difference in F1-score of 12.09 points with HEIDELTOUL and 5.07 points with GPLSI when we compare the best solutions given by both systems. These data show clearly that the fully distributional models are suitable for the event ordering task.

To conclude, after a qualitative analysis of the results obtained by our approach, with the SEMEVAL 2015 Task 4 corpus and its evaluation framework, we can say that the performance of the distributional (contextual) representation of each word in event coreference is quite similar to lexical representations. However, regarding event ordering, if we compare the results of our distributional approach with the systems presented at SemEval 2015 task 4, the improvement of distributional models is significant. Besides, in this evaluation framework, there are not any significant differences between the performance of LDA Topic Models and Word2Vec Word Embeddings distributional models. Therefore, it is not possible to conclude that for event coreference resolution an optimized local context is better than an (abstract) textual topic model or the other way round. Both contextual representations are similar regarding event coreference resolution in this framework, and further experiments with specific event coreference corpora like ECB and ECB+ are required.

5 *Conclusions*

The main aim of this paper is to determine what is the contribution of temporal inference and distributional semantic models to the cross-document event ordering task.

Regarding temporal clustering, in order to determine the time-anchoring of the events and to cluster together those which happen at the same time, our approach uses the temporal relationships between events obtained by a Temporal Information Processing system called TIPSem.

In order to analyze the impact of applying distributional semantic approaches to the task, two different methods are analyzed: LDA Topic Modeling and Word2Vec Word Embeddings. Each distributional model has been run with three different configurations: 1. considering distributional similarity between event heads and between arguments in the same proportion, 2. considering only distributional similarity between event heads, and 3. considering only distributional similarity between arguments. All of them include the temporal clustering since it is impossible for two events to be coreferent if they occur at different times.

Regarding cross-document event ordering, the different experiments have been evaluated under the framework proposed at SemEval-2015 Task 4 Track B. Results show that timeline creation is a very challenging task (Best F1-Score of 30.43%) but with our approach we are outperforming the results of the state-of-the-art systems (+12.09 points than HEIDELTOUL and +5.09 points than GPLSI in F1-score) and we consider that combining temporal inference with distributional semantic methods is a feasible approach to tackle the event ordering task.

Therefore, with this corpus and these data we can conclude first that fully distributional models are suitable for the event ordering task; second that merely the distributional vector of the event head is enough to represent the distributional meaning of the event structure and, finally, that for this specific evaluation framework, there are not significant differences between Topic Modeling and Word2Vec Word Embeddings in these tasks, as the way they are currently developed.

As Future Work, we plan to compare our system with the state-of-the-art event coreference systems using ECB or ECB+ corpora. Furthermore, we want to analyze if there is some relation between the kind of event and its event structure, and a further study of another distributional models.

References

Bagga, A. and B. Baldwin. 1999. Cross-document event coreference: Annotations, experiments, and observations. In *In Proc. ACL-99 Workshop on Coreference and Its Applications*, pages 1–8.

Baroni, M., G. Dinu, and G. Kruszewski. 2014. Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 238–247, Baltimore, Maryland, June. Association for Computational Linguistics.

Bejan, C. A. and S. Harabagiu. 2014. Unsupervised Event Coreference Resolution. *Computational Linguistics*, 40(2):311–347.

Blei, D. M., A. Y. Ng, and M. I. Jordan. 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3:993–1022.

Caselli, T., A. Fokkens, R. Morante, and P. Vossen. 2015. SPINOZA_VU: An NLP Pipeline for Cross Document TimeLines. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 787–791, Denver, Colorado, June. Association for Computational Linguistics.

Collobert, R., J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa. 2011. Natural Language Processing (Almost) from Scratch. *Journal of Machine Learning Research*, pages 41–71.

Cybulska, A. and P. Vossen. 2013. Semantic relations between events and their time, locations and participants for event coreference resolution. In *RANLP*, pages 156–163. RANLP 2011 Organising Committee / ACL.

Goyal, K., S. K. Jauhar, H. Li, M. Sachan, S. Srivastava, and E. H. Hovy. 2013. A structured distributional semantic model for event co-reference. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, ACL 2013, 4-9 August 2013, Sofia, Bulgaria, Volume 2: Short Papers*, pages 467–473. The Association for Computer Linguistics.

Ji, H., R. Grishman, Z. Chen, and P. Gupta. 2009. Cross-document event extraction and tracking: Task, evaluation, techniques and challenges. In *RANLP*, pages 166–172. RANLP 2009 / ACL.

Landauer, T. K. and S. T. Dumais. 1997. A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, 104(2):211–240.

Laparra, E., I. Aldabe, and G. Rigau. 2015. Document level time-anchoring for timeline extraction. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2015)*, Beijing, China.

Lee, H., M. Recasens, A. Chang, M. Surdeanu, and D. 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*, EMNLP-CoNLL '12, pages 489–500, Stroudsburg, PA, USA. Association for Computational Linguistics.

Li, P., Q. Zhu, and X. Zhu. 2011. A clustering and ranking based approach for multi-document event fusion. In *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 2011 12th ACIS International Conference on*, pages 159–165, July.

Llorens, H., E. Saquete, and B. Navarro-Colorado. 2012. Automatic System for Identifying and Categorizing Temporal Relations in Natural Language. *International Journal of Intelligent Systems*, 27(7):680–703.

Lu, J. and V. Ng. 2016. Event coreference resolution with multi-pass sieves. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*.

Manning, C. D., M. Surdeanu, J. Bauer, J. Finkel, S. J. Bethard, and D. McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60.

Mikolov, T., I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. 2013. Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, , and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems (NIPS 2013)*, volume 26. pages 3111–3119.

Minard, A.-L., M. Speranza, E. Agirre, I. A. adn Marieke van Erp, B. Magnini, G. Rigau, and R. Urizar. 2015. Semeval-2015 task 4: Timeline: Cross-document event ordering. In *Proceedings of the 9th International Workshop on Semantic Evaluation*, SemEval '15, pages 778–786. Association for Computational Linguistics.

Mitchell, J. and M. Lapata. 2010. Composition in Distributional Models of Semantics. *Cognitive Science*, 34:1388–1429.

Moulahi, B., J. Strötgen, M. Gertz, and L. Tamine. 2015. Heideltoul: A baseline approach for cross-document event ordering. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 825–829, Denver, Colorado, June. Association for Computational Linguistics.

Navarro-Colorado, B. and E. Saquete. 2015. GPLSIUA: Combining Temporal Information and Topic Modeling for Cross-Document Event Ordering. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 820–824, Denver, Colorado, June. Association for Computational Linguistics.

Navarro-Colorado, B. and E. Saquete. 2016. Cross-document event ordering through temporal, lexical and distributional knowledge. *Knowledge-Based Systems*, 110:244–254, October.

Palmer, M., D. Gildea, and P. Kingsbury. 2005. The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, 31.

Saurí, R., J. Littman, R. Knippen, R. Gaizauskas, A. Setzer, and J. Pustejovsky, 2006. *TimeML Annotation Guidelines 1.2.1 (http://www.timeml.org/)*.

Sun, W., A. Rumshisky, and O. Uzuner. 2013. Evaluating temporal relations in clinical text: 2012 i2b2 challenge. In *J Am Med Inform Assoc.*, pages 806–13, September-October.

UzZaman, N., H. Llorens, L. Derczynski, J. Allen, M. Verhagen, and J. Pustejovsky. 2013. Semeval-2013 task 1: Tempeval-3: Evaluating time expressions, events, and temporal relations. In *Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 1–9, Atlanta, Georgia, USA. ACL. ISBN: 978-1-937284-49-7.

Verhagen, M., R. Gaizauskas, M. Hepple, F. Schilder, G. Katz, and J. Pustejovsky. 2007. Semeval-2007 task 15: Tempeval temporal relation identification. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 75–80, Prague. ACL.

Verhagen, M., R. Saurí, T. Caselli, and J. Pustejovsky. 2010. Semeval-2010 task 13: Tempeval-2. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 57–62, Uppsala, Sweden. ACL.

Yang, B., C. Cardie, and P. I. Frazier. 2015. A hierarchical distance-dependent bayesian model for event coreference resolution. *TACL*, 3:517–528.