A Consolidated Open Knowledge Representation for Multiple Texts

Rachel Wities* [1], Vered Shwartz[1], Gabriel Stanovsky[1], Meni Adler[1], Ori Shapira[1], Shyam Upadhyay[2], Dan Roth[2], Eugenio Martinez Camara[3], Iryna Gurevych[3,4] and Ido Dagan[1]

[1]Bar-Ilan University, Ramat-Gan, Israel
[2]University of Illinois at Urbana-Champaign, IL, USA
[3]Ubiquitous Knowledge Processing Lab (UKP), Technische Universitat Darmstadt
[4]Ubiquitous Knowledge Processing Lab (UKP-DIPF), German Institute for Educational Research

{rachelvov,vered1986,gabriel.satanovsky,meni.adler,obspp18}@gmail.com
{upadhya3,danr}@illinois.edu, {camara,gurevych}@ukp.informatik.tu-darmstadt.de, dagan@cs.biu.ac.il

Abstract

We propose to move from Open Information Extraction (OIE) ahead to Open Knowledge Representation (OKR), aiming to represent information conveyed jointly in a set of texts in an open text-based manner. We do so by consolidating OIE extractions using entity and predicate coreference, while modeling information containment between coreferring elements via lexical entailment. We suggest that generating OKR structures can be a useful step in the NLP pipeline, to give semantic applications an easy handle on consolidated information across multiple texts.

1 Introduction

Natural language understanding involves identifying, classifying, and integrating information about events and other propositions mentioned in text. While much effort has been invested in generic methods for analyzing single sentences and detecting the propositions they contain, little thought and effort has been put into the integration step: how to systematically consolidate and represent information contributed by propositions originating from multiple texts. Consolidating such information, which is typically both complementary and partly overlapping, is needed to construct multi-document summaries, to combine evidence when answering questions that cannot be answered based on a single sentence, and to populate a knowledge base while relying on multiple pieces of evidence (see Figure 1 for a motivating example). Yet, the burden of integrating information across multiple texts is currently delegated to downstream applications, leading to various partial solutions in different application domains.

This paper suggests that a common consolidation step and a corresponding knowledge representation should be part of the "standard" semantic processing pipeline, to be shared by downstream applications. Specifically, we pursue an Open Knowledge Representation (OKR) framework that captures the information expressed jointly in multiple texts while relying solely on the terminology appearing in those texts, without requiring pre-defined external knowledge resources or schemata.

As we focus in this work on investigating an *open* representation paradigm, our proposal follows and extends the Open Information Extraction (OIE) approach. We do that by first extracting textual predicate-argument tuples, each corresponding to an individual proposition mention. We then merge these mentions by accounting for proposition coreference, an extended notion of event coreference. This process yields consolidated propositions, each corresponding to a single fact, or assertion, in the described scenario. Similarly, entity coreference links are used to establish reference to real-world entities. Taken together, our proposed representation encodes information about events and entities in the real world, similarly to what is expected from structured knowledge representations. Yet, being an open text-based representation, we record the various lexical terms used to describe the scenario. Further, we model information redundancy and containment among these terms through lexical entailment.

In this paper we specify our proposed representation, while specifying the involved annotation sub-tasks from which our structures are composed. We then describe our annotated dataset, of news headline tweets about 27 news stories, which is

*Corresponding author

1. Turkey **forces down** Syrian plane.	4. Turkish PM says plane was carrying ammunition for Syria government.
2. Damascus sends note to Ankara over Syrian plane.	5. Last night Turkish F16s **grounded** a Syrian passenger jet.
3. Turkey Escalates Confrontation with Syria.	6. Russia angry at Turkey about Russian passengers.

Figure 1: A sample of news headlines, illustrating the need for information consolidation. Two mentions of the same proposition, for which event coreference holds, are highlighted, with the predicate in bold and the arguments underlined. Some information is redundant, but may be described at different granularity levels; for example, different mentions describe the interception target as a *plane* and as a *jet*, where *jet* entails *plane* and is accordingly more informative.

the first to be jointly annotated for all our required sub-tasks. We also provide initial predicted baseline results for each of the sub-tasks, pointing at the limitations of current state of the art.[1]

Overall, our main contribution is in proposing to create a consolidated representation for the information contained in multiple texts, and in specifying how such representation can be created based on entity and event coreference and lexical entailment. An accompanying contribution is our annotated dataset, which can be used to analyze the involved phenomena and their interactions, and as a development and test set for automated generation of OKR structures. We further note that while this paper focuses on creating an *open* representation, by consolidating Open IE propositions, future work may investigate the consolidation of other semantic sentence representations, for example AMR (Abstract Meaning Representation) (Banarescu et al., 2013), while exploiting similar principles to those proposed here.

2 Background: Relevant Component Tasks

In this section we describe the prior annotation tasks on which we rely in our representation, as described later in Section 3.

2.1 Open Information Extraction

Open IE (Open Information Extraction) (Etzioni et al., 2008) is the task of extracting coherent propositions from a sentence, each comprising a relation phrase and two or more argument phrases. For example, *(plane,* ***landed in****, Ankara).*

Open IE has gained substantial and consistent attention, and many automatic extractors were created (e.g., Fader et al. (2011); Del Corro and Gemulla (2013)). Open IE's extractions were also shown to be effective as intermediate sentence-level representation in various downstream applications (Stanovsky et al., 2015; Angeli et al., 2015). Analogously, we conjecture a similar utility of our OKR structures at the multi-text level.

Open IE does not assign roles to the arguments associated with each predicate, as in other single-sentence representations like SRL (Semantic Role Labeling) (Carreras and Màrquez, 2005; Palmer et al., 2010). While the former is not consistent in assigning argument slots to the same arguments across different propositions, the latter requires predefined thematic role ontologies. A middle ground was introduced by QA-SRL (He et al., 2015), where predicate-argument structures are represented using question-answer pairs, e.g. (what landed somewhere?, *plane*), (where did something land?, *Ankara*).

2.2 Coreference Resolution Tasks

In our representation, we use coreference resolution to consolidate mentions of the same entity or the same event across multiple texts.

Entity Coreference Entity coreference resolution identifies mentions in a text that refer to the same real-world entity (Soon et al., 2001; Ng and Cardie, 2002; Bengtson and Roth, 2008; Clark and Manning, 2015; Peng et al., 2015). In the cross-document variant, Cross Document Coreference Resolution (CDCR), mentions of the same entity can also appear in multiple documents in a corpus (Singh et al., 2011).

Event Coreference Event coreference determines whether two event descriptions (mentions) refer to the same event (Humphreys et al., 1997). Cross document event coreference (CDEC) is a variant of the task in which mentions may occur in different documents (Bagga and Baldwin, 1999).

Compared to within document event coreference (Chen et al., 2009; Araki et al., 2014; Liu et al., 2014; Peng et al., 2016), the problem of cross document event coreference has been relatively under-explored (Bagga and Baldwin, 1999; Bejan and Harabagiu, 2014). Standard benchmarks for

[1]Our dataset, detailed annotation guidelines, the annotation tool and the baseline implementations are available at `https://github.com/vered1986/OKR`.

this task are the Event Coreference Bank (ECB) (Bejan and Harabagiu, 2008) and its extensions, that also annotate entity coreference: EECB (Lee et al., 2012) and ECB+ (Cybulska and Vossen, 2014). See (Upadhyay et al., 2016) for more details on cross document event coreference.

Differently from our dataset described in Section 4, ECB and its extensions do not establish predicate-argument annotations. A secondary line of work deals with aligning predicates across document pairs, as done in Roth and Frank (2012). PARMA (Wolfe et al., 2013) treated the task as a token-alignment problem, aligning also arguments, while Wolfe et al. (2015) added joint constraints to align predicates and their arguments.

Using Coreference for Consolidation Recognizing that two elements are corefering can help in consolidating textual information. In discourse representation theory (DRT), a proposition applies to all co-referring entities (Kamp et al., 2011). In recognizing textual entailment (Dagan et al., 2013), lexical substitution of co-referring elements is useful (Stern and Dagan, 2012). For example, in Figure 1, sentence (1) together with the coreference relation between *plane* and *jet* entail that "Turkey forces down Syrian jet."

2.3 Lexical Inference

Recognizing lexical inferences is an important component in semantic tasks, in order to bridge lexical variability in texts. For instance, in text summarization, lexical inference can help identifying redundancy, when two candidate sentences for the summary differ only in terms that hold a lexical inference relation (e.g. "the plane landed in *Ankara*" and "the plane landed in *Turkey*"). Recognizing the inference direction, e.g. that *Ankara* is more specific than *Turkey*, can help in selecting the desired granularity level of the description.

There has been consistent attention to recognizing lexical inference between terms. Some methods aim to recognize a general lexical inference relation (e.g. (Kotlerman et al., 2010; Turney and Mohammad, 2015)), others focus on a specific semantic relation, mostly hypernymy (Hearst, 1992; Snow et al., 2005; Santus et al., 2014; Shwartz et al., 2016), while recent methods classify a pair of terms to a specific semantic relation out of several (Baroni et al., 2012; Weeds et al., 2014; Pavlick et al., 2015; Shwartz and Dagan, 2016). It is worth noting that most existing methods are indifferent to the context in which the target terms occur, with the exception of few works, which were mostly focused on a narrow aspect of lexical inference, e.g. lexical substitution (Melamud et al., 2015).

Determining entailment between predicates is a different sub-task, which has also been broadly explored (Lin and Pantel, 2001; Duclaye et al., 2002; Szpektor et al., 2004; Schoenmackers et al., 2010; Roth and Frank, 2012). Berant et al. (2010) achieved state-of-the-art results on the task by constructing a predicate entailment graph optimizing a global objective function. However, performance should be further improved in order to be used accurately within semantic applications.

3 Proposed Representation

Our Open Knowledge Representation (OKR) aims to capture the consolidated information expressed jointly in a set of texts. In some analogy to structured knowledge bases, we would like the elements of our representation to correspond to entities in the described scenario and to statements (propositions) that relate them. Still, in the spirit of Open IE, we would like the representation to be open, while relying only on the natural language terminology in the given texts without referring to predefined external knowledge.

This section specifies our proposed structure, with a running example in Figure 2. The specification involves two aspects: the first is defining the component annotation sub-tasks involved in creating our representation, following those reviewed in Section 2; the second is specifying how we derive from these component annotations a consolidated representation. These two aspects are interleaved along the presentation, where for each step we first describe the relevant annotations and then how we use them to create the corresponding component of the representation.

3.1 Entities

To represent entities, we first annotate the text by entity mentions and coreference. Following the typical notion for these tasks, an *entity mention* corresponds to a word or multi-word expression that refers to an object or concept in the described scenario (in the broader sense of "entity"). Accordingly, we represent an *entity* in the described scenario by the coreference cluster of all its mentions. We represent the coreferring cluster of mentions by the multiset of its terms, keeping pointers

Original texts:

(1) Turkey forces down Syrian plane.

(2) Syrian jet grounded by Turkey carried munitions from Moscow.

(3) Intercepted Syrian plane carried ammunition from Russia.

Entities:
E_1 = {*Turkey*}, E_2 = {*Syrian*}, E_3 = {*plane*, *jet*}
E_4 = {*munitions*, *ammunition*}, E_5 = {*Moscow*}, E_6 = {*Russia*}

Proposition Mentions (+ coreference + argument alignment):
P_1: [a_1] **forces down** [a_2] (*Turkey* [E_1], *plane* [E_3])
P_2: [a_1] [a_2] (*Syrian* [E_2], *plane* [E_3]) (implicit)
P_2: [a_1] [a_2] (*Syrian* [E_2], *jet* [E_3]) (implicit)
P_1: [a_2] **grounded by** [a_1] (*jet* [E_3], *Turkey* [E_1])
P_3: [a_1] **carried** [a_2] **from** [a_3]
(*jet* [E_3], *munitions* [E_4], *Moscow* [E5])
P_1: **intercepted** [a_2] (*jet* [E_3])
P_2: [a_1] [a_2] (*Syrian* [E_2], *plane* [E_3]) (implicit)
P_3: [a_1] **carried** [a_2] **from** [a_3]
(*plane* [E_3], *ammunition* [E_6], *Russia* [E_6])

Consolidated Propositions:

P_1 : { [a_1] **forces down** [a_2], [a_2] **grounded by** [a_1], **intercepted** [a_2] } [a_1: {E_1}, a_2: {E_3}]

P_2: { [a_1] [a_2] (implicit) } [a_1: {E_2}, a_2: {E_3}]

P_3: { [a_1] **carried** [a_2] **from** [a_3] } [a_1: {E_3}, a_2: {E_4}, a_3: {E_5, E_6}]

Entailment Graphs:

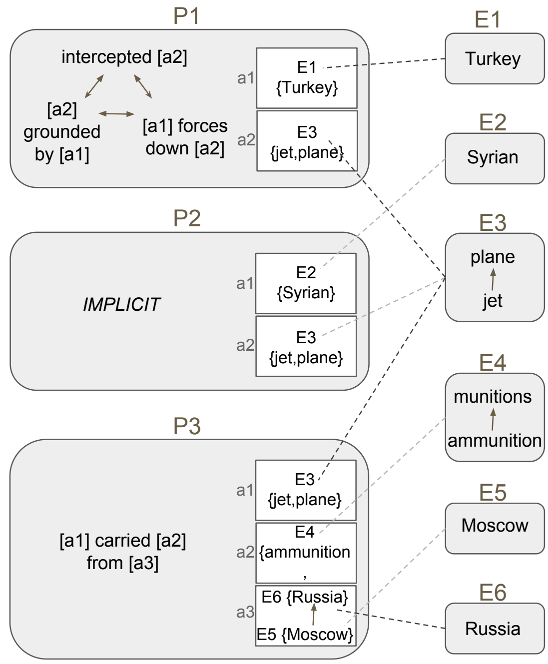
P_1 :

Propositions: [a_1] **forces down** [a_2] ⟷ [a_2] **grounded by** [a_1]; [a_1] **forces down** [a_2] → **intercepted** [a_2]; [a_2] **grounded by** [a_1] → **intercepted** [a_2]

Entities: E_3: **jet** ⟶ **plane**; E_4: **ammunition** ⟶ **munitions**

Arguments: $P_3.a_3$: **Moscow** ⟶ **Russia**

(a)



(b)

Figure 2: An illustration of our OKR formalism **(a)**, with a corresponding graphical view of the consolidated structure **(b)**. In (b), dashed lines connect entities to their instantiation within arguments, while allowing graph-traversal inferences such as: what is the relation between *Turkey* and *Russia*? *Turkey* intercepted a plane that carried ammunition from *Russia* (the path from E_1 to E_6 via the darker dashed lines).

to each term's mentions (see Entities in Figure 2; to avoid clutter, pointers are not presented in the figure). We note that we take an inclusive view which regards concepts as entities, for example the adjective *Syrian* is considered an entity mention that may corefer with *Syria*.

3.2 Proposition Mentions and Consolidated Propositions

To represent propositions, we first annotate Open IE style extractions, which we term *proposition mentions*. Each mention consists of a predicate expression, e.g. around verbs or nominalizations, and a set of arguments (see Proposition Mentions in Figure 2). We deviate slightly from standard Open IE formats by representing the predicate expression as a template, with place holders for the arguments (marked with brackets in the figure). This follows the common representation of predicates within predicate inference rules, as in DIRT (Lin and Pantel, 2001), and allows the span of entity arguments to correspond exactly to the entity term. Further, as typical in Open IE, modalities and negations become part of the lexical elements of the predicate. Notice that at this stage an argument mention is already associated with its corresponding entity. Further, we annotate implicit predicates when a predication between two entities is implied, without an explicit predicate expression, as common for noun modifications (P_2 in the figure). Nested propositions are represented by having one proposition mention as an argument of the other (e.g. "the [plane] **was forced to** [land in Ankara]").

To link different mentions of the same real world fact, we annotate *proposition coreference*, which generalizes the notion of event coreference to cover all types of predications (e.g., *John is Mary's brother* would co-refer with *Mary is John's sister*). This annotation specifies the coreference relation for a cluster of proposition mentions (denoted by the same proposition index P_i in Figure 2), as well as an alignment of their arguments, (denoted by matching argument indexes within the same proposition cluster). We then consider a *proposition* to correspond to a coreference cluster of proposition mentions, which jointly describe the referred real-world fact.

Yet, a cluster of co-referring proposition mentions does not provide a succinct representation for the aggregated textual description of a proposition. To that end, we aggregate the information in the cluster into a *Consolidated Proposition*, composed of a *consolidated predicate* and *consolidated arguments*. Similar to entity representation, a consolidated predicate is represented by the set of all predicate expressions appearing in the cluster. A consolidated argument is specified by the set of all entities (or propositions, in case of having one proposition being an argument of another one) that occupy this argument's slot in the different mentions. As with entities, each element in this representation is accompanied by a set of pointers to all its original mentions (omitted from the figure). A graphical illustration of this structure is given in Figure 2(b) (for now, ignore the arrows within some of the nodes).

A consolidated proposition encodes compactly all possible textual descriptions for the referred proposition, which can be generated from its mentions taken jointly. Each description can be generated by picking one possible predicate expression and then picking one possible lexical choice for each argument. For example, P_1 may be described as *Turkey intercepted a plane*, *Turkey forces down a jet* etc. Some of these descriptions correspond to original mentions in the text, while others can be induced through coreference (as reviewed at the end of Section 2.2). The representation of a consolidated proposition thus does not depend on the particular way in which lexical choices were split across the different proposition mentions.

3.3 Lexical Entailment Graphs

The set of descriptions encoded in a consolidated proposition is highly redundant. To make it more useful, we would like to model the information overlap between different lexical choices. For example, we want to know that *Turkey intercepted a plane* is more general than, or equivalently, is entailed by, *Turkey intercepted a jet*. To that end, we annotate the lexical entailment relations between the elements in each component of our representation, that is, within each consolidated predicate, consolidated argument and entity. This yields a *lexical entailment graph* within each component (see figure 2), which models the information containment relationships between different descriptions.

Notice that in our setting the lexical entailment relation is considered within the given context (see Section 2.3). For example, *grounded* and *forced down* may not be generically synonymous, but they do convey equivalent information in a given context of forcing a flying plane to land. Contradictions are modeled to a limited extent, by annotating contradiction relations (in context) between elements of our entailment graphs, for example when different figures are reported for the number of casualties in a disaster. This is a natural representation, since contradiction is often modeled within a three-way entailment classification task. Modeling of broader cases of contradiction is left for future work.

The entailment graphs yield better modeling of the supporting text mentions (and their total count) for each possible description. For example, knowing that *Moscow* entails *Russia*, we can assume in P_3 two supporting mentions for knowing that the ammunition was carried from Russia, while having only one supporting mention for the more detailed information regarding Moscow being the origin. Such frequency support often correlates with confidence and prominence of information, which, together with generality modeling, may be very useful in applications such as multi-document summarization or question answering. Finally, the graphical view of our representation lends itself to graph-based inferences, such as looking for all connections between two entities, similar to aggregated inferences over structured knowledge graphs (see example in Figure 2(b)).

In summary, our open knowledge representation consists of the following: *entities*, generated by detecting entity mentions and coreference; *consolidated propositions*, composed of consolidated predicates and arguments, which are generated by detecting proposition mentions and coreference relations between them; *lexical entailment graphs* for entities, consolidated predicates and consolidated arguments, which specify the inference relations between the elements within each of these components. This yields a compact representation of all possible descriptions of the statements jointly asserted by the set of texts, as induced via coreference-based inference, while tracking information containment between different descriptions as well as tracking their (induced) supporting mentions.

4 News-Related Tweets Dataset

Following the formal definition of our OKR structures, we compiled a corpus with gold annotations of our 5 subtasks (listed in Table 1). As outlined in the previous section, our structures follow deterministically from these annotations. Specifically, we make use of the news-related tweets collected in the Twitter Event Detection Dataset (McMinn et al., 2013), which clusters tweets from major news networks and other sources discussing the same event (for example, the grounding of a Syrian plane by the Turkish government). We chose to annotate news related tweets in this first dataset for several reasons: (1) they represent self contained assertions, (2) they tend to be relatively factual and succinct, and (3) by looking at several news sources we can obtain a corpus with high redundancy, which our representation aims to address. We note that while this dataset exhibits a limited amount of linguistic complexity, making it suitable for a first investigation, it still represents a very practical use case of consolidating information in a large stream of tweets about a news story.

This annotation serves two main purposes. First, it validates the feasibility of our annotation scheme in terms of annotator requirements, training and agreement. Second, to the best of our knowledge, this is the first time these core NLP annotations are annotated *in parallel* over the same texts. Following, this annotation has the potential of becoming a useful resource spurring future research into *joint prediction* of these annotations. For instance, predicate argument structures may benefit from co-reference signals, and entity extraction systems may exploit signals from lexical entailment.

Overall, we annotated 1257 tweets from 27 clusters. We release the dataset both in full OKR format, as well as ECB-like "light" format, containing only the annotated co-reference chains. Overall corpus statistics are depicted in Table 2.

4.1 Dataset Characteristics

An analysis of the annotations reveals interesting and unique characteristics of our annotated corpus.

First, the part of speech distribution of entities and predicates (Table 3) shows that our corpus captures information beyond the current focus on verb-centric applications and corpora in NLP. Namely, our corpus contains a vast number of non-verbal predications (67%), and a relatively large number of adjectival entities, owing to the fact that our structure annotates concepts such as "northern" or "Syrian" as entities in an implicit relation.

Second, the average number of unique lemmas per entity and proposition chains (2.00 and 2.24, respectively) shows that our corpus exhibits a fair amount of non-trivial lexical variability.

Finally, roughly 96% of our entailment graphs (entity and proposition) form a connected component. This data provides an interesting potential for investigating and modeling lexical inference relations within coreference chains.

4.2 Annotation Procedure and Agreement

The annotation was performed by two native English speakers with linguistic academic background, which had 10 hours of in house training. The entire annotation process took 200 person-hours using a graphical tool purposely-designed to facilitate the incremental annotation for all subtasks . We employ the QA-SRL annotation methodology to help determining Open IE predicate and argument spans in the gold standard, for its intuitiveness for non-expert annotators (He et al., 2015). Five clusters were annotated independently by both annotators and were used to measure their agreement on the task. The other clusters were annotated by one annotator and reviewed by an expert.

We measure agreement separately on each annotation subtask. After each task in our pipeline we keep only the consensual annotations. For example, we measure entity coreference agreement only for entity mentions that were annotated by both annotators. For entity, predicate and argument mention agreement, we average the accuracy

Task	Entity Ment.	Entity Co-reference				Prop. Mentions		Proposition Co-Reference								Entailment	
						Pred.	Arg.	Predicate				Argument				Entity	Prop
	avg. acc	MUC	B^3	CEAF	CoNLL $\mathbf{F_1}$	avg. acc	avg. acc	MUC	B^3	CEAF	CoNLL $\mathbf{F_1}$	MUC	B^3	CEAF	CoNLL $\mathbf{F_1}$	F_1	F_1
IAA	.85	.87	.92	.92	.90	.74 (.93, .72)†	.85	.86	.88	.76	.83	.99	.99	.98	.99	.70	.82
Pred	.58	.84	.89	.81	.85	.41 (.73, .25)†	.37	.47	.67	.56	.56	.93	.97	.94	.95	.44	.56

Table 1: Inter-Annotator Agreement (top) and off-the-shelf state-of-the-art predicted performance (bottom, see Section 5) for the OKR subtasks: (1) Entity mention extraction (for prediction we use F1 score) (2) Entity co-reference (3) Proposition Extraction (predicate identification and argument detection) (4) Proposition Co-reference (predicate coreference and argument alignment), and (5) Entailment graphs (entity and proposition entailment; argument entailment figures are not presented due to very low statistics). † Numbers in parenthesis denote verbal vs. non-verbal predicates, respectively.

# Entities	1262
# Entity mentions	5074
# Entity singeltons	777
# Propositions	1406
# Proposition mentions	4311
# Proposition Singletons	949
Avg. mentions per entity chain	8.86
Avg. distinct lemmas per entity chain	2.00
Avg. mentions per proposition chain	7.35
Avg. distinct lemmas per prop. chain	2.24
Avg. number of elements per arg. chain	1.08

Table 2: Twitter dataset statistics. Distinct lemma terms per proposition chain were calculated only on explicit propositions. Average number of elements per argument chain measures how many distinct entities or propositions were part of the same argument.

POS	Nouns	Verbs	Adj's	Impl.	Others
Ent. Dist.	.85	.01	.09	–	.05
Pred. Dist.	.40	.33	.04	.18	.04

Table 3: Entity and Predicate distribution across part of speech tags: nouns, verbs, adjectives, non-lexicalized (implicit) and all others.

Disagreement Type	**Examples**
Phrasal verbs	[*placed to leave*]$_{pred.}$ **vs.** [*placed to*]$_{pred.}$[*leave*]$_{pred.}$ [*faces charges*]$_{pred.}$ **vs.** [*faces*]$_{pred.}$ [*charges*]$_{arg.}$
Nominalizations	[*suspect*]$_{ent.}$ *plane* **vs.** [*suspect*]$_{pred.}$ *plane* [*terror*]$_{ent.}$ *attack* **vs.** [*terror*]$_{pred.}$ *attack* *U.S.* [*elections*]$_{ent.}$ **vs.** *U.S.* [*elections*]$_{pred.}$
Entailment	*fuel*$\rightarrow$*gas* **vs.** *gas*$\rightarrow$*fuel* *scandal*$\rightarrow$*case* **vs.** *case*$\rightarrow$*scandal*

Table 4: Typical cases of annotator disagreements. Annotated spans are denoted by square brackets, subscript denotes label for the mention (predicate, argument or entity).

of the two annotators, each computed while taking the other as a gold reference.

For entity, predicate, and argument co-reference we calculated coreference resolution metrics: the link-based MUC (Vilain et al., 1995), the mention-based B^3 (Bagga and Baldwin, 1998), the entity-based CEAF, and the widely adopted CoNLL F_1 measure which is an average of the three. For entity and proposition entailment we compute the F_1 score over the annotated directed edges in each entailment graph, as is common for entailment agreement metrics (Berant et al., 2010).

We macro-averaged these scores to obtain an overall agreement on the 5 events annotated by both annotators. The agreement scores for the two annotators are shown in Table 1, and overall show high levels of agreement. A qualitative analysis of the more common disagreements between annotators is shown in Table 4.

Overall, this shows that our parallel annotation is indeed feasible; agreement on each of the subtasks is relatively high and on par with reported inter-annotator agreement on similar tasks.

5 Baselines

As we have shown in previous sections, our structure is derived from known "core" NLP tasks, extended where needed to fit our consolidated representation. Subsequently, a readily available means of automatically recovering OKR is through a pipeline which uses off-the-shelf models for each of the subtasks.

To that end, we employ publicly available tools and simple baselines which approximate the current state-of-the-art in each of these subtasks. For brevity sake, in the rest of the section we briefly describe each of these baselines. For a more detailed technical description see the OKR repository (`https://github.com/vered1986/OKR`).

For *Entity Mention* extraction we use the spaCy NER model[2] in addition to annotating all of the nouns and adjectives as entities. For *Proposition Mention* detection we use Open IE propositions extracted from PropS (Stanovsky et al., 2016), where non-restrictive arguments were reduced following Stanovsky and Dagan (2016). For *Proposi-*

[2] https://spacy.io/

tion and *Entity coreference*, we clustered the entity mentions based on simple lexical similarity metrics (e.g., lemma matching and Levenshtein distance), shown to be effective on our news tweets. [3]

For *Argument Mention* detection we attach the components (entities and propositions) as arguments of predicates when the components are syntactically dependent on them. *Argument Coreference* is simply predicted by marking coreference if and only if the arguments are both mentions of the same entity co-reference chain. For *Entity Entailment* purposes we used knowledge resources (Shwartz et al., 2015) and a pretrained model for HypeNET (Shwartz et al., 2016) to obtain a score for all pairs of Wikipedia common words (unigrams, bigrams, and trigrams). A threshold for the binary entailment decision was then calibrated on a held out development set. Finally, for *Predicate Entailment* we used the entailment rules extracted by Berant et al. (2012).

5.1 Results and Error Analysis

Using the same metrics used for measuring inter-annotator agreement, we evaluated how well the presented models were able to recover the different facets of the OKR gold annotations. The performance on the different subtasks is presented in Table 1 (bottom).

We measure the performance of each component separately, while taking the annotations for all previous steps from the gold human annotations. This allows us to examine the performance of the current component, alleviating any incurred errors from previous steps. Thus, we can identify technological "bottle-necks" – the steps which most significantly lower predicted OKR accuracy using current off-the-shelf tools.

First, we noticed that non-verbal predicates pose a challenge for current verb-centric systems. This primarily manifests in low scores for identifying entities, predicates and arguments. Many entity mention errors are due to nominalizations mistakenly annotated as entities. When excluding gold nominalizations, the entity mention baseline F1 score rises from 0.58 to 0.63. As mentioned earlier (Section 4.2) nominalizations were also one of the main challenges for the annotators. Furthermore, recognizing nominalizations and other non-verbal predicates, which are very common in our dataset (see Table 3), proves to be a difficult task. Indeed, we see a significant improvement in performance when comparing verbal predicate mention performance to non-verbal performance (accuracy of 0.73 vs. 0.25). Finally, argument identification was hard mainly because of inconsistencies in verbal versus nominal predicate-argument structure in dependency trees.[4]

The low performance in predicate coreference compared to entity coreference can be explained by the higher variability of predicate terms. The argument co-reference task becomes easy given gold predicate-argument structures, as most arguments are singletons (i.e. composed of a single element).

Finally, while the performance of the predicate entailment component reflects the current state-of-the-art (Berant et al., 2012; Han and Sun, 2016), the performance on entity entailment is much worse than the current state-of-the-art in this task as measured on common lexical inference test sets. We conjecture that this stems from the nature of the entities in our dataset, consisting of both named entities and common nouns, many of which are multi-word expressions, whereas most work in entity entailment is focused on single word common nouns. Furthermore, it is worth noting that our annotations are of naturally occurring texts, and represent lexical entailment in real world coreference chains, as opposed to synthetically compiled test sets which are often used for this task.

While several tasks achieve reasonable performance on our datasets, most tasks leave room for improvement. These bottle-necks are bound to hinder the performance of a pipeline end-to-end system. Future research into OKR should first target these areas; either as a pipeline or in a joint learning framework.

6 Applications and Related Work

The need to consolidate information originating from multiple texts is common in applications that summarize multiple text into some structure, such as multi-document summarization and knowledge-base population. Currently, there is no

[3] We chose simple metrics over complex state-of-the-art entity coreference models since they target different scenarios from ours: first, they focus on named entities, and are likely to overlook common nouns like *plane* and *jet*. Second, since we work in the context of the same news story, it is reasonable to assume that, for example, two mentions of a person with the same last name belong to the same entity.

[4] E.g., "Facebook's acquisition of Instagram" is represented differently than "Facebook acquired Instagram".

systematic solution, and the burden of integrating information across multiple texts is delegated to downstream applications, leading to partial solutions which are geared to specific applications.

Multi-Document Summarization (MDS) (Barzilay et al., 1999) is a task whose goal is to produce a concise summary from a set of related text documents, such that it includes the most important information in a non-redundant manner. While extractive summarization selects salient sentences from the document collection, abstractive summarization generates new sentences, and is considered a more promising yet more difficult task.

A recent approach for abstractive summarization generates a graphical representation of the input documents by: (1) parsing each sentence/document into a meaning representation structure; and (2) merging the structures into a single structure that represents the entire summary, e.g. by identifying coreferring items.

In that sense, this approach is similar to OKR. However, current methods applying this approach are still limited. Gerani et al. (2014) parse each document to discourse tree representation (Joty et al., 2013), aggregating them based on entity coreference. Yet, they work with a limited set of (discourse) relations, and rely on coreference only between entities, which was detected manually.

Similarly, Liu et al. (2015) parse each input sentence into an individual AMR graph (Banarescu et al., 2013), and merge those into a single graph through identical concepts. This work extends the AMR formalism of canonicalized representation per entity or event to multiple sentences. However, they only focus on certain types of named entities, and collapse two entities based on their names rather than on coreference.

Event-Centric Knowledge Graphs (ECKG) (Vossen et al., 2016; Rospocher et al., 2016) is another related work which represent news articles as graphs. Event nodes are linked to DBPedia (Auer et al., 2007), with the goal of enriching entities and events with dynamic knowledge. For example, an event describing the interception of the Syrian plane by Turkey will be linked in DBPedia to *Syria* and *Turkey*.

We propose that OKR can help the described applications by providing a general underlying representation for multiple texts, obviating the need to develop specialized consolidation methods for each application. We can expect the use of OKR structures in MDS to shift the research efforts in this task to other components, e.g. generation, and eventually contribute to improving state of the art on this task. Similarly, an algorithm creating the ECKG structure can benefit from building upon a consolidated structure such as OKR, rather than working directly on free text.

7 Conclusions

In this paper we advocate the development of representation frameworks for the consolidated information expressed in a set of texts. The key ingredients of our approach are the extraction of proposition structures which capture individual statements and their merging based on entity and event coreference. Coreference clusters are proposed as a handle on real world entities and facts, while still being self-contained within the textual realm. Lexical entailment is proposed to model information containment between different textual descriptions of the same real world components.

While we developed an "open" KR framework, future work may investigate the creation of similar models based on structures that do refer to external resources (such as PropBank, as in Abstract Meaning Representation – AMR). Gradually, fine grained semantic phenomena may be addressed, such as factuality, attribution and modeling subevents and cross-event relationships. Finally, we plan to investigate performing the core annotation sub-tasks via crowdsourcing, for scalability.

Acknowledgments

This work was supported in part by grants from the MAGNET program of the Israeli Office of the Chief Scientist (OCS) and the German Research Foundation through the German-Israeli Project Cooperation (DIP, grant DA 1600/1-1), and by Contract HR0011-15-2-0025 with the US Defense Advanced Research Projects Agency (DARPA).

References

Gabor Angeli, Melvin Jose Johnson Premkumar, and Christopher D. Manning. 2015. Leveraging linguistic structure for open domain information extraction. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages

344–354, Beijing, China, July. Association for Computational Linguistics.

Jun Araki, Zhengzhong Liu, Eduard Hovy, and Teruko Mitamura. 2014. Detecting subevent structure for event coreference resolution. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014)*, pages 4553–4558, Reykjavik, Iceland. European Language Resources Association (ELRA).

Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. 2007. The semantic web. In *Lecture Notes in Computer Science*, volume 4825, chapter Dbpedia: A nucleus for a web of open data, pages 722–735. Springer Berlin Heidelberg.

Amit Bagga and Breck Baldwin. 1998. Algorithms for scoring coreference chains. In *Proceedings of the First International Conference on Language Resources and Evaluation (LREC'98)*, volume 1, pages 563–566, Granada, Spain. European Language Resources Association (ELRA).

Amit Bagga and Breck Baldwin. 1999. Cross-document event coreference: Annotations, experiments, and observations. In *Proceedings of the Workshop on Coreference and its Applications*, pages 1–8, College Park, Maryland,US. Association for Computational Linguistics.

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract meaning representation for sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186, Sofia, Bulgaria, August. Association for Computational Linguistics.

Marco Baroni, Raffaella Bernardi, Ngoc-Quynh Do, and Chung-chieh Shan. 2012. Entailment above the word level in distributional semantics. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 23–32, Avignon, France, April. Association for Computational Linguistics.

Regina Barzilay, Kathleen R. McKeown, and Michael Elhadad. 1999. Information fusion in the context of multi-document summarization. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, pages 550–557, College Park, Maryland, USA, June. Association for Computational Linguistics.

Cosmin Bejan and Sanda Harabagiu. 2008. A linguistic resource for discovering event structures and resolving event coreference. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, pages 2881–2887, Marrakech, Morocco, May. European Language Resources Association (ELRA).

Cosmin A. Bejan and Sanda Harabagiu. 2014. Unsupervised event coreference resolution. *Computational Linguistics*, 40(2):311–347.

Eric Bengtson and Dan Roth. 2008. Understanding the value of features for coreference resolution. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 294–303, Honolulu, Hawaii. Association for Computational Linguistics.

Jonathan Berant, Ido Dagan, and Jacob Goldberger. 2010. Global learning of focused entailment graphs. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1220–1229, Uppsala, Sweden, July. Association for Computational Linguistics.

Jonathan Berant, Ido Dagan, and Jacob Goldberger. 2012. Learning entailment relations by global graph structure optimization. *Computational Linguistics*, 38(1):73–111.

Xavier Carreras and Lluís Màrquez. 2005. Introduction to the CoNLL-2005 shared task: Semantic role labeling. In *Proceedings of the Ninth Conference on Computational Natural Language Learning (CoNLL-2005)*, pages 152–164, Ann Arbor, Michigan, June. Association for Computational Linguistics.

Zheng Chen, Heng Ji, and Robert Haralick. 2009. A pairwise event coreference model, feature impact and evaluation for event coreference resolution. In *Proceedings of the Workshop on Events in Emerging Text Types*, pages 17–22, Borovets, Bulgaria, September. Association for Computational Linguistics.

Kevin Clark and Christopher D. Manning. 2015. Entity-centric coreference resolution with model stacking. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1405–1415, Beijing, China, July. Association for Computational Linguistics.

Agata Cybulska and Piek Vossen. 2014. Using a sledgehammer to crack a nut? lexical diversity and event coreference resolution. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014)*, pages 4545–4552, Reykjavik, Iceland. European Language Resources Association (ELRA).

Ido Dagan, Dan Roth, and Mark Sammons. 2013. *Recognizing textual entailment*. Morgan & Claypool Publishers, San Rafael, CA.

Luciano Del Corro and Rainer Gemulla. 2013. Clausie: Clause-based open information extraction. In *Proceedings of the 22Nd International Conference on World Wide Web*, WWW '13, pages 355–366, Rio de Janeiro, Brazil. Association for Computing Machinery.

Florence Duclaye, François Yvon, and Olivier Collin. 2002. Using the web as a linguistic resource for learning reformulations automatically. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC'02)*, volume 2, pages 390–396, Las Palmas, Canary Islands - Spain. European Language Resources Association (ELRA).

Oren Etzioni, Michele Banko, Stephen Soderland, and Daniel S. Weld. 2008. Open information extraction from the web. *Communications of the ACM*, 51(12):68–74, December.

Anthony Fader, Stephen Soderland, and Oren Etzioni. 2011. Identifying relations for open information extraction. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 1535–1545, Edinburgh, Scotland, UK., July. Association for Computational Linguistics.

Shima Gerani, Yashar Mehdad, Giuseppe Carenini, Raymond T. Ng, and Bita Nejat. 2014. Abstractive summarization of product reviews using discourse structure. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1602–1613, Doha, Qatar, October. Association for Computational Linguistics.

Xianpei Han and Le Sun. 2016. Context-sensitive inference rule discovery: A graph-based method. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 2902–2911, Osaka, Japan, December. The COLING 2016 Organizing Committee.

Luheng He, Mike Lewis, and Luke Zettlemoyer. 2015. Question-answer driven semantic role labeling: Using natural language to annotate natural language. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 643–653, Lisbon, Portugal, September. Association for Computational Linguistics.

Marti A. Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. In *COLING 1992 Volume 2: Proceedings of the 15th International Conference on Computational Linguistics*, volume 2, pages 539–545. Association for Computational Linguistics.

Kevin Humphreys, Robert Gaizauskas, and Saliha Azzam. 1997. Event coreference for information extraction. In *Proceedings of a Workshop on Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts*, pages 75–81, Madrid, Spain, July. Association for Computational Linguistics.

Shafiq Joty, Giuseppe Carenini, Raymond Ng, and Yashar Mehdad. 2013. Combining intra- and multi-sentential rhetorical parsing for document-level discourse analysis. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 486–496, Sofia, Bulgaria, August. Association for Computational Linguistics.

Hans Kamp, Josef Van Genabith, and Uwe Reyle. 2011. Discourse representation theory. In *Handbook of philosophical logic*, volume 15, pages 125–394. Springer Netherlands.

Lili Kotlerman, Ido Dagan, Idan Szpektor, and Maayan Zhitomirsky-Geffet. 2010. Directional distributional similarity for lexical inference. *Natural Language Engineering*, 16(04):359–389.

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, Jeju Island, Korea, July. Association for Computational Linguistics.

Dekang Lin and Patrick Pantel. 2001. Dirt - discovery of inference rules from text. In *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '01, pages 323–328, San Francisco, California. Association for Computing Machinery.

Zhengzhong Liu, Jun Araki, Eduard Hovy, and Teruko Mitamura. 2014. Supervised within-document event coreference using information propagation. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014)*, pages 4539–4544, Reykjavik, Iceland. European Language Resources Association (ELRA).

Fei Liu, Jeffrey Flanigan, Sam Thomson, Norman Sadeh, and Noah A. Smith. 2015. Toward abstractive summarization using semantic representations. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1077–1086, Denver, Colorado, May–June. Association for Computational Linguistics.

Andrew J. McMinn, Yashar Moshfeghi, and Joemon M. Jose. 2013. Building a large-scale corpus for evaluating event detection on twitter. In *Proceedings of the 22Nd ACM International Conference on Information & Knowledge Management*, CIKM '13, pages 409–418, San Francisco, California, USA. Association for Computing Machinery.

Oren Melamud, Ido Dagan, and Jacob Goldberger. 2015. Modeling word meaning in context with substitute vectors. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 472–482, Denver, Colorado, May–June. Association for Computational Linguistics.

Vincent Ng and Claire Cardie. 2002. Improving machine learning approaches to coreference resolution.

In *Proceedings of 40th Annual Meeting of the Association for Computational Linguistics*, pages 104–111, Philadelphia, Pennsylvania, USA, July. Association for Computational Linguistics.

Martha Palmer, Daniel Gildea, and Nianwen Xue. 2010. Semantic role labeling. *Synthesis Lectures on Human Language Technologies*, 3(1):1–103.

Ellie Pavlick, Johan Bos, Malvina Nissim, Charley Beller, Benjamin Van Durme, and Chris Callison-Burch. 2015. Adding semantics to data-driven paraphrasing. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1512–1522, Beijing, China, July. Association for Computational Linguistics.

Haoruo Peng, Kai-Wei Chang, and Dan Roth. 2015. A joint framework for coreference resolution and mention head detection. In *Proceedings of the Nineteenth Conference on Computational Natural Language Learning*, pages 12–21, Beijing, China, July. Association for Computational Linguistics.

Haoruo Peng, Yangqiu Song, and Dan Roth. 2016. Event detection and co-reference with minimal supervision. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 392–402, Austin, Texas, November. Association for Computational Linguistics.

Marco Rospocher, Marieke van Erp, Piek Vossen, Antske Fokkens, Itziar Aldabe, German Rigau, Aitor Soroa, Thomas Ploeger, and Tessel Bogaard. 2016. Building event-centric knowledge graphs from news. *Web Semantics: Science, Services and Agents on the World Wide Web*, 37:132–151.

Michael Roth and Anette Frank. 2012. Aligning predicate argument structures in monolingual comparable texts: A new corpus for a new task. In **SEM 2012: The First Joint Conference on Lexical and Computational Semantics – Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*, pages 218–227, Montréal, Canada, 7-8 June. Association for Computational Linguistics.

Enrico Santus, Alessandro Lenci, Qin Lu, and Sabine Schulte im Walde. 2014. Chasing hypernyms in vector spaces with entropy. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, volume 2: Short Papers*, pages 38–42, Gothenburg, Sweden, April. Association for Computational Linguistics.

Stefan Schoenmackers, Jesse Davis, Oren Etzioni, and Daniel Weld. 2010. Learning first-order horn clauses from web text. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 1088–1098, Cambridge, MA, October. Association for Computational Linguistics.

Vered Shwartz and Ido Dagan. 2016. Path-based vs. distributional information in recognizing lexical semantic relations. In *Proceedings of the 5th Workshop on Cognitive Aspects of the Lexicon (CogALex - V)*, pages 24–29, Osaka, Japan, December. The COLING 2016 Organizing Committee.

Vered Shwartz, Omer Levy, Ido Dagan, and Jacob Goldberger. 2015. Learning to exploit structured resources for lexical inference. In *Proceedings of the Nineteenth Conference on Computational Natural Language Learning*, pages 175–184, Beijing, China, July. Association for Computational Linguistics.

Vered Shwartz, Yoav Goldberg, and Ido Dagan. 2016. Improving hypernymy detection with an integrated path-based and distributional method. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2389–2398, Berlin, Germany, August. Association for Computational Linguistics.

Sameer Singh, Amarnag Subramanya, Fernando Pereira, and Andrew McCallum. 2011. Large-scale cross-document coreference using distributed inference and hierarchical models. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 793–803, Portland, Oregon, USA, June. Association for Computational Linguistics.

Rion Snow, Daniel Jurafsky, and Andrew Y. Ng. 2005. Learning syntactic patterns for automatic hypernym discovery. In *Advances in Neural Information Processing Systems*, volume 17, pages 1297–1304. MIT Press.

Wee M. Soon, Hwee T. Ng, and Daniel C. Y. Lim. 2001. A machine learning approach to coreference resolution of noun phrases. *Computational linguistics*, 27(4):521–544.

Gabriel Stanovsky and Ido Dagan. 2016. Annotating and predicting non-restrictive noun phrase modifications. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1256–1265, Berlin, Germany, August. Association for Computational Linguistics.

Gabriel Stanovsky, Ido Dagan, and Mausam. 2015. Open IE as an intermediate structure for semantic tasks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 303–308, Beijing, China, July. Association for Computational Linguistics.

Gabriel Stanovsky, Jessica Ficler, Ido Dagan, and Yoav Goldberg. 2016. Getting more out of syntax with props. *arXiv preprint*.

Asher Stern and Ido Dagan. 2012. Biutee: A modular open-source system for recognizing textual entailment. In *Proceedings of the ACL 2012 System Demonstrations*, pages 73–78, Jeju Island, Korea, July. Association for Computational Linguistics.

Idan Szpektor, Hristo Tanev, Ido Dagan, and Bonaventura Coppola. 2004. Scaling web-based acquisition of entailment relations. In Dekang Lin and Dekai Wu, editors, *Proceedings of EMNLP 2004*, pages 41–48, Barcelona, Spain, July. Association for Computational Linguistics.

Peter D. Turney and Saif M. Mohammad. 2015. Experiments with three approaches to recognizing lexical entailment. *Natural Language Engineering*, 21(03):437–476.

Shyam Upadhyay, Nitish Gupta, Christos Christodoulopoulos, and Dan Roth. 2016. Revisiting the evaluation for cross document event coreference. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics*.

Marc Vilain, John Burger, John Aberdeen, Dennis Connolly, and Lynette Hirschman. 1995. A model-theoretic coreference scoring scheme. In *Proceedings of the 6th conference on Message understanding*, pages 45–52, Columbia, Maryland. Association for Computational Linguistics.

Piek Vossen, Rodrigo Agerri, Itziar Aldabe, Agata Cybulska, Marieke van Erp, Antske Fokkens, Egoitz Laparra, Anne-Lyse Minard, Alessio Palmero Aprosio, German Rigau, et al. 2016. Newsreader: Using knowledge resources in a cross-lingual reading machine to generate more knowledge from massive streams of news. *Knowledge-Based Systems*, 110:60–85.

Julie Weeds, Daoud Clarke, Jeremy Reffin, David Weir, and Bill Keller. 2014. Learning to distinguish hypernyms and co-hyponyms. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 2249–2259, Dublin, Ireland, August. Dublin City University and Association for Computational Linguistics.

Travis Wolfe, Benjamin Van Durme, Mark Dredze, Nicholas Andrews, Charley Beller, Chris Callison-Burch, Jay DeYoung, Justin Snyder, Jonathan Weese, Tan Xu, and Xuchen Yao. 2013. Parma: A predicate argument aligner. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 63–68, Sofia, Bulgaria, August. Association for Computational Linguistics.

Travis Wolfe, Mark Dredze, and Benjamin Van Durme. 2015. Predicate argument alignment using a global coherence model. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 11–20, Denver, Colorado, May–June. Association for Computational Linguistics.