Composing Process Knowledge using Semantic Roles

Niranjan Balasubramanian

September 28, 2015

Contents

1 Introduction

Knowledge extraction and reasoning are central for many Artificial Intelligence systems. Recently introduced tasks such as *MCTest* reading comprehension challenge [24], grade-level science exams [6], and process comprehension tasks [2] serve as excellent benchmarks for developing reasoning-based question answering systems. These tasks test the ability of the systems to interpret and reason about scenarios and situations. Knowledge requirements analysis for grade-level science exams shows the need for deep inference supporting knowledge [8, 7]. Also, even with advanced state-of-the-art reasoning techniques, shallow text-derived knowledge is ineffective due to lexical and syntactic variations in language [14]. Scalable acquisition of deeper semantic knowledge is essential for effective reasoning in these tasks.

Much progress has been made on question answering involving simple facts [1, 9, 4, 23]. This is in large part due to the availability of large scale curated relational knowledge bases such as Freebase, coupled with significant advances in automatic relation extraction [20, 5, 26]. Similar advances in large scale inference-supporting knowledge is vital for making significant advances in reasoning-based QA tasks.

In this work, we focus on knowledge acquisition methods with grade science exams as an end application. In particular we focus on a subset of knowledge about physical, chemical, and other natural phenomena. The goal is to derive knowledge that allows effective reasoning about scenarios involving these phenomena. We refer to this as process knowledge. [1]

1.1 Motivation

We illustrate the need for inference supporting knowledge using the following example from a 4^{th} grade science exam.

A pot is heated on a stove. What causes the metal handle of the pot to become hot?
(A) conduction (B) convection (C) radiation (D) combustion.

The question tests reasoning about heat transfer mechanisms. To establish that conduction is the cause, a reasoning system must establish that there is some heat transfer happening through *direct contact*. We envision a system that first interpret the scenario and apply its knowledge about the heating process and the conduction process. In this case, the system needs to first identify *what is being heated* (the pot), and *what is the purported result* of the heating (the pot handle becoming hot). Then, knowledge about *heating* should allow it to conclude that the thing being heated (the pot) will become hot. Knowledge about *conduction* should suggest that anything *in direct contact* with a hot object (the pot handle) should also become hot. Combining these bits the system can verify that the described scenario indeed matches the conduction process. Replicating this type of reasoning requires the knowledge in a suitable computational representation.

The knowledge required – what is undergoing the process, what the result is and so on – are naturally encoded as semantic roles. PropBank and FrameNet are two of the highly successful efforts which provide this kind of semantic role based knowledge. They have spurred tremendous advances in automatic semantic role labeling and their applications []. While these resources provide exhaustive coverage for modeling general open-domain actions, they do not cover knowledge about scientific processes. FrameNet, for instance, does

[1] Our representation and acquisition methods will be targeted towards the grade science benchmarks but the methodology is general and can be applied to other tasks and domains.

not have entries for nearly half of the processes described in 4th grade science exams. The coverage is likely worse for higher grade levels with deeper knowledge domains.

1.2 Proposal

In response, we propose to build a knowledge base about physical, chemical, and biological processes from their textual descriptions.

Representation

Our primary motivation is to design a representation that facilitates reasoning while balancing extractability. We propose to compose knowledge about each process using a set of pre-determined vocabulary of semantic roles and derive additional new roles automatically as needed. While many competing role sets exist, we propose a role vocabulary that covers most inferential needs in the target process recognition dates [19]. By leveraging a common vocabulary we expect to gain from shared learning for the same roles across different processes.

Similar to FrameNet we propose to determine roles with respect to the semantics of the process rather than with respect to the specific verb (or predicate) that is used to describe the process. This canonicalization is desirable as it reduces reasoning time burden by eliminating need for steps that only establish equivalence of information realized via different predicates.

Role	Types	Instances	Patterns
Undergoer	Physical Object	solid, pot, …	when <x> is heated, <x> conducts heat, …
Enabler/Enabling Event	Physical Object	stove, flame, …	heated on the <x>, heated by <x>, …
Theme	Energy	heat, radiation, …	conducts <x>, transfers <x>, …
Output	-	-	-
Purpose/Consequence	-	maintain temperature, …	helps to <x>, …
Benefactive	-	-	-
Source	Physical Object	solid, pot, …	when <x> is heated, <x> conducts heat, …
Target	Physical Object	handle, vessel, …	becomes <x> hot, …
Medium	solid, contact	solid, contact, …	via <x>, in touch with <x>, …

Figure 1: An example entry for the process *thermal conduction*.

With each role we include role fillers which are the frequent entries for that role, types which encodes class level information about the fillers, and syntactic argument realization patterns for these fillers. We also include the source sentences from which the knowledge was derived. Figure 1 shows an example for the envisioned knowledge. The table includes a set of pre-specified general purpose roles such as *Undergoer*, *Enabler* and also a process specific role *Medium* that is automatically derived from inspecting sentences that describe the process.

Iterative Role Extraction and Joint Inference

We devise an iterative procedure that exploits the abundance and variety of information on the web to acquire the necessary knowledge.

General semantic role labeling task is challenging because of the lexical and syntactic variations in role realizations. Handling the variations requires learning from large amounts of training data, which is laborious and requires expert labor. Also as discussed earlier, existing semantic resources such as FrameNet or PropBank cannot be directly used for training as they do not cover the target concepts. Rather than build a role labeler that can work on any sentence, we propose a simpler approach that works well on a smaller subset of sentences.

We tackle the role acquisition problem by explicitly searching for sentences that convey role information using specific patterns. Leveraging the vastness of the web greatly reduces the need to accurately identify roles in difficult to process sentences. Using a set of manually constructed query patterns we find sentences which convey the roles using expected lexico-syntactic constructs. For instance "<process name> causes <x>" is a simple pattern that can be used to find the *result* role for a process. This strong expectation for how arguments are realized allows us to turn SRL into a simpler task of assessing the confidence on the extraction using features that generalize better across different processes. This allows us to substantially reduce the burden of annotating sentences for each process. Even though PropBank and FrameNet data are not directly usable, the predicate-based roles and the frame structures can provide evidence for composing our target semantic roles. We propose to exploit these features to build a local sentence-level extractor.

Despite the targeted acquisition, local sentence-level extraction alone is not adequate, because not all lexico-syntactic patterns are unambiguous. For example, "evaporates into <x>" extracts steam, a *result*, as well as atmosphere, a *location*.

A second key idea is to perform joint inference over multiple sentences to avoid errors in the local extraction. Because our goal is to acquire knowledge about a process, we turn the variety of expression on the web to our advantage. We build on prior work in Integer Linear Program (ILP) based semi-supervised and unsupervised induction approaches. The ILP attempts to find an assignment of roles that maximizes the sum of extraction confidences, while also minimizing disagreements in labels for similar text spans in similar sentences.

Assessment and Discovery

While we start with a set of general purpose roles, *a priori* we don't assume which roles are applicable to each process as this manual assignment is not desirable from a scalability standpoint. Also not all critical information may be covered by the set of roles that hand.

A third contribution of this proposal includes methods that identify the subset of roles that apply to a process and discover new roles not part of the general vocabulary. As part of the targeted acquisition, we propose an iterative approach with methods to i) assess the coverage of the roles, ii) induce new roles by identifying consistently repeated arguments that don't fit existing roles, and iii) relevance feedback based query pattern generation to expand the knowledge.

Evaluation

The generated knowledge base will be evaluated for intrinsic quality and external utility. For internal evaluations we will perform manual evaluation of the resulting KB. As part of the evaluation, **we will also create a large scale curated KB, which relies on correcting the outputs from the system, rather than careful annotation of each sentence**. We will also generate a gold standard of the desired roles and role fillers for evaluating coverage. As an external evaluation we will use the knowledge base for answering process recognition questions. All the resources and evaluation test beds will be shared with the research community for further research.

1.3 Prior Work and Contributions

The motivation and direction for this proposal stems directly from our prior work on grade science exams. Our earlier work studied knowledge requirements [8], developed inference-supporting rule knowledge bases [7], and investigated sophisticated state-of-the-art probabilistic reasoning methods for QA [14]. Our ongoing preliminary work investigated the value of a handful of semantic roles in answering process recognition questions and identified the knowledge, representation, and reasoning gaps [19]. This proposal aims to address the central challenges that we've identified through these prior works.

Upon successful completion this project would have made the following contributions:

- Release of a knowledge base covering semantic representations of processes for grade level science.
- Methods for gathering high quality sentences that cover a target set of roles.
- A joint inference method that reconciles roles across different sentences.
- Methods for assessing applicability, coverage and discovery of new roles.

2 Background and Preliminary Work

TODO: Needs a rewrite. Don't read yet. Much of SRL work has focused on interpreting sentences i.e., identifying roles expressed in a sentence. Our goal is different. We seek to acquire knowledge about processes and represent them in terms of semantic roles. SRL is a challenging structure prediction task which often requires

Prior work either resorted to supervised learning techniques or to semi-supervised approaches However, obtaining training data is often difficult and laborious, especially for complex tasks such as SRL. Several prior approaches have used semi-supervised learning (SSL) approaches to address this issue.

We propose a different approach, where we explicitly find sentences that express the roles of interest and do a joint inference of role labels across all sentences. This allows us greater flexibility in choosing which sentences to use (we can maximize role coverage, diversity of instances etc.). Aligning the various spans within these sentences allows us to do collective labeling (similar in spirit to transduction learning).

Our central premise is that the entities involved in a process and the roles they play provide a powerful representation for reasoning and QA. Similar representations have been shown to be useful for Open-domain factoid question answering [25, 21], and reading comprehension tasks on process descriptions [2].

2.1 Preliminary Work

As preliminary work, we first analyzed the knowledge requirements for a set of questions targeting around 150 processes. While a small collection of general purpose roles (e.g., Undergoer, Result, Enabler, Trigger) capture the key semantic elements for a majority of the processes [19], we also find that a set of domain specific roles (e.g., Direction, Medium, Physical_Property, Chemical_Property) are also critical. We propose to compose a representation that combines both general and domain specific roles.

Note that these roles are specified based on a canonical view of the process. One of the key challenges in question answering is in overcoming the variability of expression. Labeling with respect to a canonicalized view of the process is a deliberate attempt at addressing the variability problem.

Manually specifying roles for each process is not desirable from a scalability standpoint. Lexical and syntactic cues are indicative of certain roles in sentences. For example a *[process] happens by [event]* pattern suggests the presence of a manner role, and the presence of locative prepositions can indicate location or orientation roles.

3 ProcIteRoles Architecture

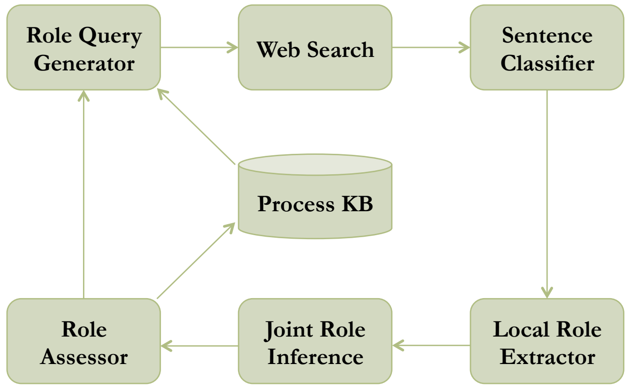

Figure 2: Process KB Acquisition: Proposed Architecture

ProcIterRoles allows for targeted iterative acquisition and refinement of process roles. The main idea is to collect high quality sentences and roles and iteratively expand the acquisition to include additional sentences and other roles.

Figure 2 shows the architecture of the acquisition process. Using simple query patterns we search for sentences that express roles with highly regular lexical cues. A sentence filter addresses sense issues and to remove malformed sentences. The filtered sentences are then processed through a pattern-based extractor that identifies and scores the candidate roles. In parallel, the sentences are also aligned to identify lexical units that should play similar semantic roles in the sentences. A joint inference module then provides a collective label assignment for all the sentences. The extracted roles are then assessed in conjunction with the existing roles. The assessor determines which roles should be added to the KB and also determines which roles need further additions. New query patterns are created for the roles that need addition and whole procedure is repeated again.

In the subsequent sections, we describe each component in greater detail.

4 Sentence Gathering

ProcIterRoles will leverage the vastness of the web to build a targeted collection of sentences that expresses roles of interest. There are two key challenges to address in gathering relevant sentences:

- **Relevance** – We want to find sentences that describe the target process of interest. The vastness of the web also means that there is information on nearly any possible interpretation of the words used to describe the process. For instance if we are interested in the process *crop fertilization*, we might also find information on *fertilization* in the reproductive sense, or in other metaphorical uses such as *cross fertilization of ideas*. Also, the target sense may not be a dominant sense on the web. Constructing effective keyword queries is therefore critical for finding relevant information.
- **Role Coverage** – We want to find sentences that cover all applicable roles that convey the desired information via simple expected constructions. Some roles are often expressed via highly regular lexico-syntactic constructions. On the other hand, roles such as x can be expressed in many different ways. Again with the vastness of the web, constructing effective role patterns is critical for finding useful sentences.

To address these challenges we propose two strategies. First, we target definitional sentences which convey the most salient information. Second, we espouse a feedback strategy that leverages high quality sentences gathered in the earlier phases to guide search in the later stages.

4.1 Querying

We propose to investigate simple but effective query pattern formulation methods. On a related task, our prior work had explored techniques that can find sentences that convey information about different aspects with respect to a topic (e.g., biographical aspects of a person). Our preliminary work shows that simple lexical templates e.g., "<process name> is the process by which" can yield high quality *definitional* sentences about processes. We use additional lexical templates for roles with regular lexicon-syntactic constructions e.g., "<process name> causes" is an effective pattern to find sentences that express the *result* roles of processes. The key challenge is to figure out querying patterns for roles with diverse lexical realizations. We propose adopting the standard bootstrapping approach used in relation extraction techniques to borrow functional patterns that introduce roles in other processes. Bootstrapping is known to introduce noise and topic drift issues. However, our approach doesn't entirely rely on the patterns alone. Rather we propose strong scoring and filtering mechanisms that can remove noise introduced via bootstrapping.

4.2 Scoring and Filtering

To account for the challenges in relevance, we seek to build a distributional context model that is seeded with some domain corpus. This model is then refined iteratively to allow for role coverage. Sentences from the web have high variance in quality and relevance.

5 Role Extraction

Many different approaches have been investigated for role labeling. The learning formulations studied range from pipelined classification approaches [13, 3], efficient structured and joint inference [16, 27], to end-to-end deep learning architectures [10]. Many different lexico-syntactic features, such as dependency paths and n-gram contexts, provide weak evidences for determining semantic roles [13]. Because these path-based and n-gram features are sparse, these supervised techniques require large amounts of training data. Semi-supervised and unsupervised approaches have been proposed as a means to address the training data problem [11, 15].

The focus of these approaches have been to build a SRL system that can identify the roles mentioned in a sentence. Our requirement is subtly different. We need to build a mechanism for acquiring the typical role fillers for a given process. First, we formulate a simpler local classification task that avoids the need for learning over role and predicate specific patterns. Starting with sentences that are likely to contain a specific role and a candidate text span from the sentence, we pose a classification task to determine if the candidate is indeed expressing role of interest. Then, we pose a joint inference task over multiple sentences, which allows us to use role decisions on similar text spans to influence each other.

5.1 Local Role Extraction

We set up a local (within sentence per role) extraction task. Relying on the patterns alone is problematic. Hand authored patterns, especially specifying the expected syntactic structure of the argument is quite limiting. Instead we generate many possible arguments that match a range of weakly indicative argument patterns and train a classifier.

The inputs are a sentence S, the role R for which it was retrieved, and the role pattern X_R that it matches, and a set of candidate spans C. The task is to predict if for each span if it is expressing the role of interest.

We adopt the standard SRL features such as clause, dependency path features, and n-gram context features [13, 16]. We explore two types of extensions that are specific to our setting:

- Different from a standard SRL setting, we seek identification of roles with respect to a canonical realization of the process. One can view this task as finding a mapping from predicate-specific semantic role to the process-specific role. To this end, we use an SRL system trained on PropBank data to identify predicate level semantic roles and use those as features to derive this mapping. Similarly, the frames that are evoked by the predicates in the sentence also provide important signals. For example, knowing that there is a conversion frame in a sentence increases the possibility of finding a result of a change of state of process like evaporation.

- Also we have strong expectations on *how* the argument is realized because the sentence is retrieved via a specific query pattern. This allows us to encode features that test if the expectations are met. [**Explain w/ an example**]

5.2 Joint Role Inference

Within sentence joint role inference has been shown to help SRL [22, 16]. Since our objective is to extract knowledge from multiple sentences, we propose to also exploit joint inference of roles across sentences. Local role extraction allows to reliably identify whether the specified role is expressed by the candidate text spans. However, this local classification is often inadequate because some cue patterns are ambiguous. For example, "evaporates into <x>" can match "steam" which is a *result* or "atmosphere" which is a *location*. Relying on the extraction pattern alone is problematic. Therefore, we propose to leverage role predictions on other similar text spans to improve inference.

Formalism

The key premise is that aligned text spans in similar sentences should be assigned same roles. This idea had been successfully used in semi-supervised and unsupervised settings to increase training data for SRL [11, 12, 18]. We adopt it for joint inference over test sentences, similar to the transductive learning approaches [**?**].

We extend an ILP-based formalism which has been shown to successfully model within sentence joint inference for SRL [22]. We add 1) a penalty term to the maximization objective that penalizes assignments that violate smoothness of labeling, 2) constraints that effectively fix labels from the local extractor for certain roles.

As noted earlier, the local extractor scores each text span $t_{i,k}$ from sentence S_k on how likely it is to belong to the role r_j. We use indicator variables to denote role assignment. z_{ijk} represents if the text span t_{ik} in sentence S_k is assigned the role r_j. Formally, the inference aims to find the best joint assignment to set of indicator variables F that maximizes the following objective function:

$$\arg\max_{\mathbf{z}} \sum_{i,j,k} z_{ijk} \cdot \rho(t_{ik}, r_j) \cdot \lambda_j - \beta \left\{ \sum_{i,k,l,m} \sigma(t_{ik}, t_{lm}) \left(\sum_{c \in |R|} |z_{ick} - z_{lcm}| \right) \right\}$$

subject to

$$\forall z_{ick} \in \mathbf{z}, \sum_{c=1}^{|R|} z_{ick} \leq 1 \text{[A span gets only one role.]}$$

$$\forall S_k \forall c \in |R|, \sum_{t_{ik} \in S_k} z_{ick} \leq 1 \text{[Roles are not repeated.]}$$

$$\cdots \text{[Other within sentence constraints.]}$$

Alignment

The effectiveness of the joint inference relies on the ability to identify text spans in different sentences that should get the same role. Prior work explored a dependency graph-based approach to align predicate-argument structures in sentences. A linear combination of the overlap in lexical and syntactic structures of the candidate text spans is used to evaluate whether they should get the same roles [11, 12, 17]. This scoring function is used to transfer roles from a labeled sentence to an unlabeled sentence (semi-supervised setting) or to induce roles as clusters of arguments (unsupervised setting).

A key difference in our setting is that there are different sub-groups of sentences with different alignment characteristics. *Definition* sentences describe the process in terms of classes of entities and *instance* sentences which involve specific entities. Aligning definitional sentences is quite different from aligning definitions with instances or instances with other instances. Instances involve completely different entities which may not align via direct entailment but may align as substitutable siblings. Also, instances often tend to involve other non-essential information with respect to the process, whereas definitions are compact and tend to contain the most salient bits of information. To account for these differences we consider two extensions. Use different sets of weights and different similarity functions to combine the scores based on the types of sentences being aligned.

$$\sigma(t_{ik}, t_{lm}, u, v) = \alpha_{uv} \cdot lexsim_{uv}(t_{ik}, t_{lm}) + (1 - \alpha_{uv}) \cdot synsim_{uv}(t_{ik}, t_{lm})$$

[**Consider changing this to a trained classifier. Use sentence patterns rather than definition or instance sentence distinction.**]

6 Role Assessment and Discovery

Inference yields a set of role fillers that can be reliably identified from the input set of sentences. The yield is limited by the set of roles and the query patterns used to retrieve sentences for these roles.

Assessment

To expand the knowledge further, we propose an iterative procedure that learns from the inferred roles to find new query patterns. Traditional bootstrapping methods coupled with query expansion techniques from information retrieval can be exploited to instantiate query templates with new patterns derived from the retrieved sentences. The new query patterns are then used to repeat the entire procedure to derive new fillers.

The iterative procedure however can yield role fillers that are inconsistent with roles obtained from previous iterations. While it is possible to infer new roles jointly with the current KB at each iteration, it can introduce too many variables in inference and render it inefficient. Therefore, we propose a simple aggregation procedure that consolidates the roles by effectively resolving any inconsistencies between the different iterations.

Role Discovery

At each iteration we will also inspect if there are any new roles that need to be added to the role set. For many processes there are specific important roles that do not fit any of the general roles. Some examples:

- *Phototropism* is the mechanism by which plants grow towards a light source. The notion of a target *light source* and the *direction* or *orientation* of the growth are critical for distinguishing positive and negative phototropism. However, neither notion fits with any of the existing roles.
- *Heat transfer* processes such as radiation have notion of a *medium* which is critical for distinguishing between instances such as convection and radiation Again medium doesn't fit with any existing roles.

We propose a graph-based clustering formulation that attempts to identify frequently repeated arguments and assess how well they fit with existing roles. If a frequently repeated argument is assigned no role or is not assigned to any particular role with high confidence then we propose to induce a new role for such arguments.

Formulation

The role fillers can be viewed as entities that share specific role-specific relations with each other. We propose to identify new roles by finding role fillers that share many connections with other extracted role fillers. To this end, we create a graph where the nodes are role fillers and edges are role labels or relations connecting the role fillers. Using the dependency graphs of the sentences, we locate other arguments (chunks) that are associated with the extracted role fillers and add them to the graph.

[To be filled out...]

7 Evaluation

Our primary motivation is to compose knowledge that is necessary for grade-level science questions. Therefore we propose both an intrinsic evaluation that measures the accuracy of the extracted roles, and an external evaluation where we assess the utility of the roles in question answering.

7.1 KB Evaluation

Our goal is to acquire knowledge about processes along relevant role dimensions. [**To be filled...]**

7.2 Question Answering using Semantic Roles

Testbed Following Clark et al., 2014, we propose to build a large collection of multiple-choice questions at different grade levels. In prior work we leveraged existing collections to select questions. To effectively test the utility of our approach, we propose to create two types of questions in increasing levels of difficultly:

1) Questions that test ability to recognize instances of processes. 2) Questions that require ability to reason about the instance. We will limit the reasoning questions to those that are effectively achieved by following simple axioms defined over the semantic roles. [Example???]

Creating these questions is a difficult task for non-experts. We propose to work with the graduate students in the Department of Education at Stony Brook to collect questions and question templates. Once we have generate question templates that specify what kind of understanding is to be tested, we can scale out the question acquisition process via crowd sourcing.

Method

We use a simple approach to evaluate the utility of semantic roles for question answering. For the recognition questions we will follow our prior work, where we build a supervised role labeler and use it to parse the question to extract a role based representation. Then for each answer option we match the extracted roles against the roles available in the database.

For reasoning questions, we will utilize the above approach to first recognize the process and identify the roles mentioned in the question. We will utilize hand-written axioms defined over the roles. [Example.?]

8 Development Plan and Timeline

The project will proceed in three phases. In the first phase, we will design the representation and methods for role discovery and acquisition. In the second phase, we will start curating the generated knowledge and build a semantic role labeler for questions. In the third phase we will refine and make necessary adjustments to the pipeline, finish the curation and release the resources. The research plan in calendar years is shown below:

Year 1:

1.
2.
3.

Year 2:

1.
2.
3.

9 Broader Impact

10 Curriculum Development Activities

I plan to teach a course centered around the core concepts of knowledge representation, and scalable extraction techniques for knowledge. This course is relevant for both Masters and PhD students. Most NLP-based technology companies and technology companies with a large web presence have a need for extracting and organizing knowledge from their user engagement data. This course will provide a basic overview of a distributed information extraction pipeline, persistence, and building applications that rely on the extracted data.

11 Prior NSF Support

Dr. Niranjan Balasubramanian has broad expertise in information extraction, esp. in large scale knowledge generation and its application to complex NLP tasks but has not received prior support from NSF.

References

[1] Jonathan Berant, Andrew Chou, Roy Frostig, and Percy Liang. Semantic parsing on freebase from question-answer pairs. In *EMNLP*, pages 1533–1544, 2013.

[2] Jonathan Berant, Vivek Srikumar, Pei-Chun Chen, Abby Vander Linden, Brittany Harding, Brad Huang, Peter Clark, and Christopher D. Manning. Modeling biological processes for reading comprehension. In *Proceedings of EMNLP*, 2014.

[3] Anders Björkelund, Love Hafdell, and Pierre Nugues. Multilingual semantic role labeling. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning: Shared Task*, pages 43–48. Association for Computational Linguistics, 2009.

[4] Antoine Bordes, Jason Weston, and Nicolas Usunier. Open question answering with weakly supervised embedding models. In *Machine Learning and Knowledge Discovery in Databases*, pages 165–180. Springer, 2014.

[5] Andrew Carlson, Justin Betteridge, Bryan Kisiel, Burr Settles, Estevam R Hruschka Jr, and Tom M Mitchell. Toward an architecture for never-ending language learning. In *AAAI*, volume 5, page 3, 2010.

[6] Peter Clark. Elementary school science and math tests as a driver for ai: Take the aristo challenge. *to appear*, 2015.

[7] Peter Clark, Niranjan Balasubramanian, Sumithra Bhakthavatsalam, Kevin Humphreys, Jesse Kinkead, Ashish Sabharwal, and Oyvind Tafjord. Automatic construction of inference-supporting knowledge bases. In *AKBC2014*, Montreal, Canada, December 2014.

[8] Peter Clark, Phil Harrison, and Niranjan Balasubramanian. A study of the AKBC/requirements for passing an elementary science test. In *Proc. of the AKBC-WEKEX workshop at CIKM*, 2013.

[9] Anthony Fader, Luke Zettlemoyer, and Oren Etzioni. Open question answering over curated and extracted knowledge bases. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1156–1165. ACM, 2014.

[10] William R Foland Jr and James H Martin. Dependency-based semantic role labeling using convolutional neural networks. In *Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics*, pages 279–288, 2015.

[11] Hagen Fürstenau and Mirella Lapata. Graph alignment for semi-supervised semantic role labeling. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 11–20, Singapore, 2009.

[12] Hagen Fürstenau and Mirella Lapata. Semi-supervised semantic role labeling via structural alignment. *Computational Linguistics*, 38(1):135–171, 2012.

[13] Daniel Gildea and Daniel Jurafsky. Automatic labeling of semantic roles. *Computational linguistics*, 28(3):245–288, 2002.

[14] Tushar Khot, Niranjan Balasubramanian, Eric Gribkoff, Ashish Sabharwal, Peter Clark, and Oren Etzioni. Exploring markov logic networks for question answering. In *Empirical Methods in Natural Language Processing*, 2015.

[15] Ivan Titov Alexandre Klementiev. Semi-supervised semantic role labeling: Approaching from an unsupervised perspective. In *Proceedings of the COLING Conference.*, 2012.

[16] Peter Koomen, Vasin Punyakanok, Dan Roth, and Wen-tau Yih. Generalized inference with multiple semantic role labeling systems. In *Proceedings of the Ninth Conference on Computational Natural Language Learning*, pages 181–184. Association for Computational Linguistics, 2005.

[17] Joel Lang and Mirella Lapata. Unsupervised induction of semantic roles. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 939–947, Los Angeles, California, June 2010. Association for Computational Linguistics.

[18] Joel Lang and Mirella Lapata. Unsupervised semantic role induction with graph partitioning. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1320–1331, Edinburgh, Scotland, UK., July 2011. Association for Computational Linguistics.

[19] Samuel Louvan, Chetan Naik, Veronica Lynn, Ankit Arun, Niranjan Balasubramanian, and Peter Clark. Semantic role labeling for process recognition questions. In *K-CAP Scientific Knowledge Workshop*, 2015.

[20] Mausam, Michael Schmitz, Robert Bart, Stephen Soderland, and Oren Etzioni. Open language learning for information extraction. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 523–534. Association for Computational Linguistics, 2012.

[21] Luiz Augusto Pizzato and Diego Mollá. Indexing on semantic roles for question answering. In *Coling 2008: Proceedings of the 2nd workshop on Information Retrieval for Question Answering*, pages 74–81. Association for Computational Linguistics, 2008.

[22] Vasin Punyakanok, Dan Roth, Wen-tau Yih, and Dav Zimak. Semantic role labeling via integer linear programming inference. In *Proceedings of the 20th International Conference on Computational Linguistics*, COLING '04, Stroudsburg, PA, USA, 2004. Association for Computational Linguistics.

[23] Siva Reddy, Mirella Lapata, and Mark Steedman. Large-scale semantic parsing without question-answer pairs. *Transactions of the Association for Computational Linguistics*, 2:377–392, 2014.

[24] Matthew Richardson, Christopher JC Burges, and Erin Renshaw. Mctest: A challenge dataset for the open-domain machine comprehension of text. In *EMNLP*, volume 1, page 2, 2013.

[25] Dan Shen and Mirella Lapata. Using semantic roles to improve question answering. In *EMNLP-CoNLL*, pages 12–21, 2007.

[26] Fabian M Suchanek, Gjergji Kasneci, and Gerhard Weikum. Yago: a core of semantic knowledge. In *Proceedings of the 16th international conference on World Wide Web*, pages 697–706. ACM, 2007.

[27] Oscar Täckström, Kuzman Ganchev, and Dipanjan Das. Efficient inference and structured learning for semantic role labeling. *Transactions of the Association for Computational Linguistics*, 3:29–41, 2015.