Composing Process Knowledge using Semantic Roles

Niranjan Balasubramanian

September 29, 2015

Contents

1 Introduction

AI systems have achieved impressive success with answering factual questions (e.g., "When was Bill Clinton born?"), using information from large-scale text or database resources [2, 10, 5, 28, 17]. To scale more challenging problems that go beyond factual answer retrieval [29, 7, 3], AI systems will need the ability to reason about the world. In particular they need knowledge of generalities and how it applies to the specific situations. Despite significant advances in AI, acquiring general knowledge and using it to reason effectively remains a huge challenge.

We propose to investigate this challenge in the context of learning and reasoning about *processes*. Knowledge about processes is a fundamental to part of our understanding of the world and it is essential to make predictions and answer questions about specific scenarios involving them. Our goal is to develop solutions to automatically construct a large repository of simple process knowledge automatically, and demonstrate its use to answer process questions that go beyond fact lookup.

1.1 Motivation

We motivate the need for process knowledge using an example from an actual 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 the ability to recognize a specific scenario involving a heat transfer process. 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 interprets 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 knowledge about conduction and heating in a suitable computational representation. TODO: [Reads very rough.]

At a high level the main bits of required knowledge are: what is undergoing the process, what is the result, what is the main action etc. These bits of information are naturally encoded as semantic roles. PropBank [19] and FrameNet [1] 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 [15, 33, 11, 30]. While these resources provide exhaustive coverage for modeling general open-domain actions, they only provide partial coverage on processes in the science domain. FrameNet, for instance, does 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.

In response we propose to investigate methods to automatically construct a large repository of simple knowledge about processes. Specifically, we will make three main contributions:

- A method for **automatic extraction of process knowledge** that combines extraction and joint inference.

- A framework for **iterative knowledge expansion**, which allows the system to discover new roles involved in a process and expand the process representation to accommodate them.
- The first comprehensive, large-scale **knowledge base of processes**, describing the roles and changes involved in that process.

To provide a concrete test-bed for development and evaluation purposes, we will initially develop this in the domain of elementary science, and evaluate it using unedited science exam questions about processes. We expect the resource to be useful to other researchers working in the areas of natural language processing, text processing, and question answering.

2 Proposal Synopsis

Our primary motivation is to design a representation that effectively supports reasoning while also being amenable for automatic extraction. We target a simple[1] form of knowledge that captures information about the entities involved and their semantic roles within the process.

We turn to the inferential needs of the target application to guide our choices. Our preliminary work in this domain suggests a mix of **pre-specified semantic roles** that apply to many processes, and a set of **automatically induced roles** that are process specific [24]. We also find that we need both **definitional** – describing the key classes of entities and types of actions involved – and **instance level** knowledge about specific scenarios in which the processes occur – providing details on the specific entities and actions.

Role	Types	Instances
Undergoer	Physical Object,	solid, pot, …
Enabler/Enabling Event	Physical Object, Heat Source	stove, flame, …
Theme	Energy, Heat	heat, radiation, …
Output	-	-
Purpose/Consequence	-	maintain temperature, …
Benefactive	-	-
Source	Physical Object	solid, pot, …
Target	Physical Object	handle, vessel, …
Medium	solid, contact	solid, contact, …

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

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. In addition to the *instance* role fillers, it also includes *definitional* type information which encodes class level information where possible.

Research Questions

There are many research challenges in automatically extracting this knowledge from text.

[1]Process knowledge is typically complex with sub-events, and temporal dependencies. They are beyond the scope of this investigation.

- SRL systems typically rely on large amounts of training data in order to generalize over sparse features. How can we leverage the abundance of information on the web to reduce the need for large training data?
- A priori it is not clear how to identify which set of sentences are likely to contain relevant information. How to gather sentences that convey the necessary knowledge?
- How do we account for process specific roles?

Our investigation aims to answer these questions.

2.1 Automatic Extraction of Process 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.

Our key premise is that to gather role knowledge we don't need a semantic role labeler that works well on all sentences. We are interested in role acquisition and not role interpretation. We exploit the abundance and variety of information available on the web to target extraction from sentences that convey the same information in expected (easy to extract) constructs and use joint inference to further improve performance. Our approach includes:

- **Targeted Pattern-based Extraction** – We build a simple pattern-based local role extractor augmented with a classifier. Using a set of manually constructed query patterns we search the web to find sentences and extract role fillers. A simple classifier then assesses if the extraction is valid. Unlike traditional SRL tasks, here we have a strong expectation of the type of role and where the filler is likely to be located. This expectation allows us to design simple structural features that generalize better across different roles and processes, thereby reducing need for large amounts of training data.
- **Joint Inference Across Sentences** – We propose a joint inference model that operates over multiple sentences to avoid errors in the local extraction. Despite the targeted acquisition, local sentence-level extraction alone is not adequate, because not all patterns are unambiguous. For example, "evaporates into <x>" extracts steam, a *result*, as well as atmosphere, a *location*. Redundancy and variety of expression on the web can help us: If the *atmosphere* were in fact a *result* then we might expect to see other sentences where it is expressed as a result with unambiguous patterns. We propose a joint inference formulation that favors roles minimizing disagreements in labels for similar text spans in similar sentences.

2.2 Iterative Knowledge Expansion

Determining the best set of knowledge bearing sentences *a priori* is extremely difficult. The quality and scope of the extracted knowledge is effectively determined by the set of sentences retrieved by the query patterns. The query patterns may have limited coverage in some cases especially for roles that were not part of the pre-specified vocabulary. We propose an iterative expansion of the knowledge to improve coverage of existing roles, and to discover new roles. An inspection of the gathered knowledge can provide valuable guidance in expanding and refining the knowledge. We propose to investigate methods to 1) assess the coverage of the roles, 2) induce new roles by identifying consistently repeated arguments that don't fit existing roles, 3) bootstrap extractors and query patterns using relevance feedback techniques, 4) refine the knowledge by organizing the knowledge in clusters of scenarios/instances.

2.3 Knowledge Base of Processes

Our target domain is grade level science. Based on initial analysis we estimate to generate about entries for around 2000 processes encompassing simple physical, chemical, biological, and other natural processes in this domain. As part of the proposed work, we propose to curate a subset of this knowledge base. Recent work has shown an effective question-based mechanism for acquiring semantic role labels via crowd sourcing [16]. To foster further research, we will release the knowledge base, and open source the code, and host web services that will allow dynamic construction of knowledge for new unseen processes. We anticipate this knowledge base will also be useful for QA in the science domain, for communities interested in AI systems, and to the semantic role labeling community at large.

The proposed methods will be evaluated for intrinsic quality and external utility. For internal evaluations we will perform manual evaluation of the resulting KB. As an external evaluation we will use the knowledge base for answering process recognition questions.

The remainder of this proposal provides details on these three main contributions.

3 ProcIterRolesArchitecture

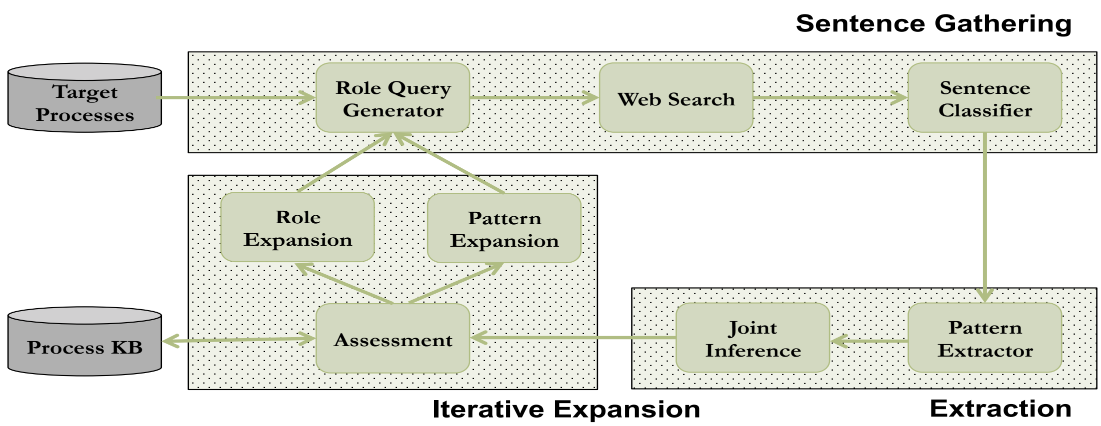

Figure 2: Process KB Acquisition: Proposed Architecture

ProcIterRoles allows for targeted iterative extraction and expansion 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 Role Query Generator

We propose to investigate simple but effective query pattern formulation methods. 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 Sentence Classifier

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 Extraction

Many different approaches have been investigated for role labeling. The learning formulations studied range from pipelined classification approaches [15, 4], efficient structured and joint inference [21, 34], to end-to-end deep learning architectures [12]. Many different lexico-syntactic features, such as dependency paths and n-gram contexts, provide weak evidences for determining semantic roles [15]. 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 [13, 20].

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 Pattern-based 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 [15, 21]. 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 [27, 21]. 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 [13, 14, 23]. 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 [27]. 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:

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

$$\text{subject to}$$

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

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

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

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 [13, 14, 22]. 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 and Refinement

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.

Our objective is similar to the goals of prior work on unsupervised role induction. They use a deterministic procedure for candidate argument identification and clustered syntactic signatures of these arguments to induce roles. However, unlike the standard unsupervised setting, we have a set of roles that have been identified already. Further, we find that most of these process specific new roles tend to be realized via prepositional, noun-noun, or other noun modifier relations that attach to one of the existing roles. Information about how these candidates are related to currently identified roles is likely to help.

- We propose to investigate methods for breaking coarse-grained roles into multiple sub-roles. and preposition relation extraction tools

  We use two sources for identifying candidates. First, we obtain candidates from the local extraction pipeline that were assigned low scores by the inference and then we extend it with candidates from a separate PropBank style argument identification pipeline []. Second, we also inspect role fillers for existing roles that can be broken up into fine-grained roles. For example, "towards a light source" can be further split into two roles one relating to the "direction" of the growth and the other relating to the goal "light source".
- Having identified candidates we iterate through the pipeline to find additional sentences that contain these candidates and the core roles or predicates for the process. Following prior role induction work, we extract a context signature for each candidate and cluster the candidates that are realized with similar contexts. As mentioned earlier, we propose to also use the semantic role context of the candidates.

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] Collin F Baker, Charles J Fillmore, and John B Lowe. The berkeley framenet project. In *Proceedings of the 17th international conference on Computational linguistics-Volume 1*, pages 86–90. Association for Computational Linguistics, 1998.

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

[3] 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.

[4] 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.

[5] 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.

[6] 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.

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

[8] 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.

[9] 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.

[10] 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.

[11] Parvin Sadat Feizabadi and Sebastian Padó. Combining seemingly incompatible corpora for implicit semantic role labeling. *Proc. of* SEM*, pages 40–50, 2015.

[12] 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.

[13] 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.

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

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

[16] Luheng He, Mike Lewis, and Luke Zettlemoyer. Question-answer driven semantic role labeling: Using natural language to annotate natural language. 2015.

[17] Rob High. The era of cognitive systems: An inside look at ibm watson and how it works.

[18] 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.

[19] Paul Kingsbury and Martha Palmer. Propbank: the next level of treebank. In *Proceedings of Treebanks and lexical Theories*, volume 3. Citeseer, 2003.

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

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] Michael Roth and Mirella Lapata. Context-aware frame-semantic role labeling. *Transactions of the Association for Computational Linguistics*, 3:449–460, 2015.

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

[32] 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.

[33] Mihai Surdeanu, Richard Johansson, Adam Meyers, Lluís Màrquez, and Joakim Nivre. The conll-2008 shared task on joint parsing of syntactic and semantic dependencies. In *Proceedings of the Twelfth Conference on Computational Natural Language Learning*, pages 159–177. Association for Computational Linguistics, 2008.

[34] 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.