STAR: A System of Argumentation for Story Comprehension and Beyond

Irene-Anna Diakidoy
Dept. of Psychology
University of Cyprus
eddiak@ucy.ac.cy

Antonis Kakas
Dept. of Computer Science
University of Cyprus
antonis@cs.ucy.ac.cy

Loizos Michael
Sch. of Pure & Applied Sciences
Open University of Cyprus
loizos@ouc.ac.cy

Rob Miller
Dept of Information Studies
University College London
rsm@ucl.ac.uk

Abstract

This paper presents the *STAR* system, a system for automated narrative comprehension, developed on top of an argumentation-theoretic formulation of defeasible reasoning, and strongly following guidelines from the psychology of comprehension. We discuss the system's use in psychological experiments on story comprehension, and our plans for its broader use in empirical studies concerning wider issues of commonsense reasoning.

Introduction

In this paper we present a new system, called *STAR*: STory comprehension through ARgumentation, for automated narrative comprehension. Its theoretical basis rests on the now established argumentation theory in Artificial Intelligence (Bench-Capon and Dunne 2007; Baroni, Caminada, and Giacomin 2011), uniformly applied to reason about actions and change in the presence of default background knowledge. Its practical development follows strongly guidelines from the psychology of comprehension both for its representation language and for its computational mechanisms for building and revising a comprehension model as the story unfolds. *STAR* has been used as part of psychological experiments to ascertain the background world knowledge that humans use in story comprehension and to examine the suitability of such knowledge for the automated comprehension of stories.

Although the original design of *STAR* was geared to story comprehension, the system provides the opportunity to carry out empirical studies to examine important questions on more general issues, such as: "What is an appropriate form of representation of commonsense knowledge used by humans in everyday tasks so that these tasks can be effectively automated?" and "How can such knowledge be acquired automatically from the Web?". Having a system that can evaluate empirically the computational properties together with the psychological relevance of possible answers to such questions can help significantly in developing automated systems with common sense. The system together with a growing corpus of benchmark stories is publicly available at: `http://cognition.ouc.ac.cy/narrative/`.

Theoretical Basis of *STAR*

The *STAR* system implements an argumentation-based semantics, appropriately adapted to account for the temporal aspects of reasoning necessary for story comprehension. We offer a high-level view of the semantics, and direct the reader to earlier work for more details (Diakidoy et al. 2014).

The semantics operates on a ***narrative***, a set of ***association rules***, and a ***priority relation*** among the latter. The first of these components corresponds to the story being comprehended, while the other two are typically story-independent (although story-specific additions can be dealt with as well). Implicit in the set of association rules is always a persistence rule for each literal and each time-point, stating that a literal that is true will remain so at the subsequent time-point.

Ignoring, for now, the existence of a priority relation, the semantics operates as one would typically expect. The narrative is interpreted as a set of time-stamped facts, and the association rules as a set of classical implications (over time-stamped literals), which can be applied to draw inferences. Both *modus ponens* and *modus tolens* can be used, supporting both forward and backward (i.e., contrapositive) reasoning through the association rules. Note that *modus tolens* on an association rule is used in a restricted manner: if the rule's head is known to be false, and every literal except one in the rule's body is known to be true, then the remaining literal in the rule's body is inferred to be false. Thus, both *modus ponens* and *modus tolens* are unit-propagating inference rules, adding single-literal inferences to what is already inferred.

A set comprising a subset of the narrative facts and a subset of the association rules is thought of as an ***argument***, supporting the inferences that are classically entailed by the set. Thus, following psychological guidelines, arguments are grounded on the explicit information in the story narrative.

Association rules are by their very nature defeasible, and subject to qualification by other association rules. The priority relation makes explicit their relative strength. Lifting this priority relation from individual rules to collections of rules gives rise to the notion of attacks between arguments that one needs to define as part of an argumentation framework.

The lifting from rule priorities to argument attacks is a key part of the semantics, as it has to carefully deal with the subtleties of using *modus tolens* on defeasible rules. In general, an argument attacks another if the two include association rules (or narrative facts, the mentioning of which is suppressed below) that are conflicting. Two association rules are ***conflicting*** when the inferences drawn through these rules are contradictory; such a conflict is called ***direct***. However, due to backward reasoning one needs to account for conflicts between rules that have contradictory heads, even if there is no direct conflict; such conflicts are called ***indirect***.

Based on the notion of conflict, we then say that an association rule ***strongly*** (resp., ***weakly***) ***qualifies*** another if the former is stronger (resp., not weaker) than the latter in terms of priority, and either they are in direct conflict, or they are in indirect conflict and the former is used in the forward direction while the latter in the backward direction.

Without going into details, we illustrate the definition via an example: If one observes a penguin and infers that it cannot fly, then one is not allowed to reason backward through the rule that birds can fly, to infer that it is not a bird. Indeed, the rule that penguins cannot fly is stronger and is in indirect conflict with the rule that birds can fly. Therefore, the latter is qualified and cannot be part of an acceptable argument.

The notion of qualification provides the necessary tool to define attacks between arguments. An argument ***attacks*** another if a rule in the former argument strongly qualifies a rule in the latter one; or if there is no strong qualification between rules of the two arguments, but there is weak qualification.

In accordance with another psychological guideline, that only skeptical inferences (but not necessarily all of them) are drawn when humans read stories, we define a ***comprehension model*** to be any subset of the unique grounded extension of the argumentation framework defined above.

The *STAR* System through an Example Story

The *STAR* system operates on a domain file, whose syntax and operational semantics we describe through an example "doorbell story", given in natural language text below. The full example and its tutorial presentation can be found at: `http://cognition.ouc.ac.cy/narrative/`.

> *Ann rang the doorbell. Mary, who was in the flat watching TV, got up from her chair and walked to the door. She was afraid. Mary looked through the keyhole. She saw Ann, her flatmate, at the door. Mary opened the door and asked Ann why she did not use her keys to come in the flat. Ann replied that she was upset that Mary did not agree to come with her at the shops. She wanted to get her up from her chair in front of the TV.*

To simplify the representation we are assuming in this story the existence of only one flat, door, doorbell, etc., and certain known entities, although as we shall discuss in the next section this is a computationally significant simplification that new versions of the system should be able to address.

A domain file for the *STAR* system is a file following and extending the typical Prolog syntax, and comprising:

- A series of sessions specified by the user, each representing the story narrative up to a certain story scene, and a set of questions to be answered at that point. This part facilitates the empirical investigation of reading incrementally the story and answering questions along the way.
- The world knowledge (WK) used as background knowledge for the story comprehension. This part amounts to a static representation of relevant commonsense knowledge about the physical and mental world of the story, which resides in the system's memory and is used in all sessions.

PART 1: Sessions and Narrative

The first part of the *STAR* domain file comprises a series of sessions. A ***session*** is a statement of the form:

```
session(s(#N),#Questions,#Visible).
```

where `#N` is a non-negative integer, and

- `#Questions` is a list of question names `q(#N)`, for any `#N`, to be answered by the system during this session;
- `#Visible` is a list of domain concepts that we wish to be shown on the screen as the system constructs its comprehension model. `#Visible` can alternatively take the value `all`, requesting that all domain concepts in the constructed comprehension model are shown on the screen.

A concept `#Concept` is a predicate name along with associated variables or constants for the predicate's arguments. A literal `#Literal` is either a concept `#Concept` or its negation `-#Concept` (i.e., the symbol for negation is "`-`").

A session statement as above gives the operational definition of the session. The ***narrative content*** of a session is given by a set of ***observation*** statements of the form:

```
s(#N) :: #GroundLiteral at #TimePoint.
```

where `#GroundLiteral` is a literal whose arguments are all constants, and `#TimePoint` is a positive integer.

Typically, we start with an initial / pre-story session statement `session(s(0),[],all).`, where background typing information (that remains unchanged across time) is given for the objects in the story. In the "doorbell story" example, such an initial session includes the following, stating that "ann" and "mary" are instances of / have type "person":

```
s(0) :: person(ann) at always.
s(0) :: person(mary) at always.
```

These observations are stated to hold at every time-point, and are available for reasoning everywhere in the story.

One can separate a given story in any number of sessions one wishes. One could have a single session (in addition to, or including, the initial session), in which case one gives the whole story narrative to the system for it to answer questions after "reading and comprehending the whole story". Alternatively, one could have a session after each sentence of the story where the system reads the story sentence by sentence and answers questions after "comprehending the story up to the end of the current sentence read in the last session".

For example, we may want to set the first scene, and hence the first session, of the "doorbell story" to include the narrative that comes from just the first sentence of the story. The narrative content of the first scene / session could then be:

```
s(1) :: rang(ann,doorbell) at 1.
```

This is humanly extracted from the story text "*Ann rang the doorbell.*" using the concepts and objects explicitly used in the text. Nevertheless, this process of extracting narrative facts from the text already contains a part of the comprehension process, since the natural language processing tasks of pronoun and article de-referencing, and, importantly, of setting the relative timing of these observations, is assumed to be carried out in extracting the narrative from the story.

The operative part of this first session is represented by the following statement / instruction to the *STAR* system to read up to scene `s(1)` and answer questions `q(1)` and `q(2)`:

```
session(s(1),[q(1),q(2)],all).
```

Questions are defined using the following syntax as in the example questions of the "doorbell story" below:

```
q(1) ?? has(ann,doorkeys) at 1.
q(2) ?? is_a(ann,visitor) at 1;
        is_a(ann,resident) at 1.
```

Question `q(1)` is a true / false question: "Does Ann have the door keys?". Question `q(2)` is a multiple-choice question: "Is Ann a visitor or a resident?". The aim of this session, hence, is to test the elaboration inferences that the system performs "after reading the first sentence of the story".

The same question can be readily reused and asked as part of several distinct sessions. This is typically done to check if there has been a revision in the comprehension model of the system as the story unfolds. For each answer choice to a question, the system returns one of the following: ***accepted***, meaning that this choice holds in the comprehension model, ***rejected***, meaning that its negation holds in the comprehension model, and ***possible***, when neither of the above holds.

A second session and scene could be represented as follows, again through a humanly extracted narrative, based on the text of the second and third sentences of the story:

```
session(s(2),[q(3)],all).

s(2) :: in_flat(mary) at 3.
s(2) :: watch(mary,tv) at 3.
s(2) :: getup(mary,chair) at 3.
s(2) :: walk_to(mary,door) at 4.
s(2) :: afraid(mary) at 4.

q(3) ?? wants(mary,see_who_at(door)) at 4;
        wants(mary,open(door)) at 4.
```

Question `q(3)` corresponds to "Why did Mary walk to the door?", probing to see what explanation, if any, has the system formed in its comprehension model (this far) for Mary walking to the door. This explanatory question is not open-ended, but asks to select from two possible explanations, by checking if any of the two presented choices are already part of the comprehension model of the system. Thus, the system does not behave like a human reader where it might be the case that the question itself, rather than the story text, causes the mind to form an explanation in its comprehension model, when none might have been present before the question was posed. Therefore, the user needs to anticipate the answering of the particular questions by providing related background knowledge in the second part of the story representation.

Along the same lines, we can complete the sessions and narrative of the "doorbell story" with new scenes as follows:

```
s(3) :: look_through(mary,keyhole) at 5.
s(3) :: see_at(mary,ann,door) at 6.
s(3) :: flatmate(mary,ann) at 6.
s(4) :: open(mary,door) at 8.
s(4) :: ask(mary,ann,use(doorkeys)) at 8.
s(5) :: upset_with(ann,mary) at 9.
s(5) :: -agree(mary,request_from(ann,
                    go_to(shops))) at 1.
s(6) :: wants(ann,getup(mary,chair)) at 1.
```

The exact time-points in the narrative are largely inconsequential, other than indicating the ordering of observations. However, they should be spaced sufficiently apart to allow causal laws to bring about their indirect / knock-on effects.

PART 2: Background World Knowledge

The second part of the *STAR* domain file includes commonsense world knowledge (WK) that we judge is appropriate for a particular story, or more generally for a theme in which many stories can be written; e.g., the general theme of "life at home" of which the "doorbell story" is an example.

This world knowledge comprises simple rules, which the system then interprets as premises on which to build the arguments that support its comprehension model. An important feature of these rules is that they express typical associations between concepts. Hence, they are not strict in nature, but rather they are inherently default rules. Rules could be thought of as properties that normally hold under certain conditions, e.g., "if a person rings the doorbell then normally that person does not have the keys". This frees the user from the heavy representational burden of detailed rules that take into account all circumstances of their possible qualification.

The first step in representing WK is to declare those concepts that are fluent, i.e., persist over time and change across time. All other concepts are taken to hold instantaneously at the point of their observation or inference. The declaration of fluents is done using the statement `fluents(#Fluents)`, where `#Fluents` is a list of fluents. This declaration for the "doorbell story" example is as follows, where `_` is (as in Prolog) a free unspecified argument of the concept:

```
fluents([ in_flat(_), watch(_,_), afraid(_),
    flatmate(_,_), upset_with(_,_), has(_,_),
    is_a(_,_), expect(_,_), wants(_,_),
    knows(_,_), agree(_,_), refused(_,_) ]).
```

WK is represented in terms of associations between concepts. There are three types of such association rules, ***property, causal, preclusion***, of the following form respectively:

```
p(#N) :: #Body implies #Literal.
c(#N) :: #Body causes #Literal.
r(#N) :: #Body precludes #Literal.
```

where `p(#N)`, `c(#N)`, `r(#N)` are unique identifiers of their respective association rules, and rule bodies are of the form

```
#Body = true | #Literal | #Literal,#Body
```

Concepts in rules need to be range-restricted, and this is ensured by insisting that every variable appearing in a rule appears in at least two concepts. Since the application of rules can be traced back to the grounded concepts in the narrative, and since each rule is unit-propagating, this range-restriction ensures that all drawn inferences are also grounded.

For property rules the association between their body and their head is at the same time point, while for causal rules this is between time points t and $t+$, where $t+$ is the time-point immediately following t (i.e., $t+1$ for discrete time). A preclusion rule represents an association between its body holding at time t and its head *not* holding at time $t+$. These preclude the causation of their head by some other causal rule, as preclusion rules are stronger than causal rules.

The background knowledge can also contain priority (i.e., relative strength) statements between association rules. Domain-specific priorities are entered by the user in the domain in the form `#A1 >> #A2.`, where `#A1` and `#A2` are the names of two association rules. Note also that in general the system treats causal (labeled `c(#N)`) associations rules stronger than (competing) inertia rules, inertia rules stronger than (competing) property (labeled `p(#N)`) association rules, and preclusion (labeled `r(#N)`) association rules stronger than (competing) causal association rules, but also weaker than any other (competing) association rule.

We illustrate this type of knowledge representation with some examples of commonsense knowledge pertaining to the particular "doorbell story", and questions asked in this.

"Those with doorkeys do not normally ring the doorbell."
```
p(11) :: has(Person,doorkeys) implies
         -ring(Person,doorbell).
```

"Visitors do not normally have doorkeys, but residents do."
```
p(12) :: is_a(Person,visitor) implies
         -has(Person,doorkeys).
p(13) :: is_a(Person,resident) implies
         has(Person,doorkeys).
```

These are properties that hold in typical situations in the world of "doorbell ringing". They are the type of knowledge that humans turn to or activate when primed by the first sentence of the story and / or when probed by questions such as `q(1)` and `q(2)`. We present empirical results on psychological aspects of story comprehension in the next section.

Other relevant knowledge is given by the following rules:

"Walking to the door normally gets one close to the door."
```
c(21) :: walk_to(Person,door) causes
         close_to(Person,door).
```

"Walking to the door normally stops one from watching TV."
```
c(22) :: watch(Person,tv),
         walk_to(Person,door) causes
         -watch(Person,tv).
```

This knowledge reflects associations between actions and some of their effects. They are causal rules, as those found in frameworks of Reasoning about Actions and Change: the original Situation Calculus (McCarthy and Hayes 1969) and Event Calculus (Kowalski and Sergot 1986; Miller and Shanahan 2002), or subsequent action languages and calculi (Gelfond and Lifschitz 1993; McCain and Turner 1997; Kakas and Miller 1997; Thielscher 1999; Kakas, Michael, and Miller 2011). Unlike in most of these frameworks, these rules are inherently default in nature, and do not need to contain explicit qualification conditions, such as the condition `-carry(Person,tv_on)` in rule `c(22)` that the person is not carrying the TV switched on. Should this precondition be needed in the context of some story, then some other commonsense knowledge, such as "One carrying a TV switched on is normally watching the TV." will be activated to provide the qualification. This could be done by either of these rules:

```
r(23) :: carry(Person,tv_on) precludes
         -watch(Person,tv).
p(24) :: carry(Person,tv_on) implies
         watch(Person,tv).
```

The first rule will qualify the termination of watching TV by any (opposing) causal rule. With the second rule the user will need to additionally make its strength explicit by including in the domain the statement `p(24) >> c(22).`.

Note that we have the commonsense knowledge that "One does not normally carry a TV switched on.". Hence, adding the condition `-carry(Person,tv_on)` in the causal rule `c(22)` would be unnecessary. In other words, should we decide to add this precondition explicitly in `c(22)` it will be trivially satisfied by the commonsense knowledge

```
p(25) :: person(Person) implies
         -carry(Person,tv_on).
```

Note that not even the precondition `watch(Person,tv)` is needed in `c(22)`. Removing this would appear to have the effect that whenever a person walks to the door, we would derive that that person is not watching TV. Although this is typically true and the system could include such conclusions in its comprehension model, one would expect that such an inference would not be drawn (or would be quickly dropped) by humans due to lack of *coherence* with the rest of the story, unless of course the story is about watching TV.

Reasoning About the Mental World

Additional parts of commonsense world knowledge used for comprehending the "doorbell story" are the following:

"If the doorbell rings and a person is not expecting visitors this normally makes the person afraid."
```
c(31) :: rings(doorbell),
         -expect(Person,visitors) causes
         afraid(Person).
```

"If the doorbell rings and a person is afraid then normally the person is not expecting visitors."
```
p(32) :: rings(doorbell), afraid(Person)
         implies -expect(Person,visitors).
```

"One who is afraid normally does not want to open the door."
```
p(33) :: afraid(Person) implies
         -wants(Person,open(door)).
```

"Walking to the door normally implies wanting to see who is there."
```
p(34) :: rings(doorbell),
         walk_to(Person,door) implies
         wants(Person,see_who_at(door)).
```

"Walking to the door normally implies wanting to open it."

```
p(35) :: rings(doorbell),
         walk_to(Person,door) implies
         wants(Person,open(door)).
p(33) >> p(35).
```

"If it is a flatmate at the door then normally this stops fear."

```
c(36) :: see_at(Person,Other,door),
         flatmate(Person,Other) causes
         -afraid(Person).
c(36) >> c(31).
```

Knowledge above concerns the "mental aspects of opening the front door", written down in an informal commonsense way, in terms of behavior rules of how desires, expectations and emotions are generated in people from stimuli in their environment. There are two important comments to make here that relate to the nature of the representation.

The first comment relates to that — analogously to the case of knowledge about the physical world of the story, so in the case of the mental world of the story — the knowledge is specialized in the context of "doorbell ringing". As we shall discuss in the next section, such specialized world knowledge is typical of what human readers would volunteer, when asked to verbalize the knowledge they are using to comprehend a story. Indeed, the knowledge volunteered by humans could be specialized even more, by dropping the `rings(doorbell)` and `flatmate(Person,Other)` conditions, and the given knowledge would still suffice to properly make sense of a story in the "doorbell ringing" context.

However, knowledge in such a specialized form would not be useful in different contexts under which expectations and fear of visitors would be relevant. This raises the important question of whether there is a more general form of the knowledge, from which one could generate in some automated way this specialized contextual knowledge. On the psychological front, one could ask whether there is evidence that such a process of specialization operates during story comprehension, or whether knowledge is "stored" in the human mind directly in its specialized contextual form.

The second, related, comment concerns the degree of formality of that knowledge; i.e., whether it formally relates to a grand theory of desires and emotions (Murphy and Medin 1985). Perhaps we would not expect commonsense reasoning to rely (at least consciously) on a detailed scientific-like theory. Instead, we would expect humans in their typical everyday reasoning to use behavior rules, like the ones above, to capture the specific aspects of the mental world phenomena that are relevant at the time, relying on a phenomenological understanding and description of our mental world.

Such knowledge in the form of behavior rules could even be a-causal in nature. One could posit that through experience humans turn typical causal explanations into rules of inference, effectively compiling and compacting typically activated parts of their knowledge into a more efficient form. Such an example is rule `p(32)` where the typical explanation "of not expecting visitors" for the observation that a person is afraid when they hear the doorbell (see rule `c(31)`) is turned into a rule for "not expecting visitors". Similarly, `p(34)` and `p(35)` are behavior rules capturing typical explanations of why one walks to the door when the doorbell rings; these would otherwise be derived via abductive reasoning from the causal rules where desires generate actions.

In fact, a new form of abductive reasoning is relevant to comprehension when seeking to explain a given information in the story, e.g., when asked an (open) explanatory question such as "Why did Ann ring the doorbell?" in our example story. This new form of abduction is best described by *inference to the most coherent explanation*, namely inference to a hypothesis that together with the background world knowledge can logically derive the observation, and that is "closely dependent" on the rest of the inferences drawn in the comprehension model. In cognitive psychology this property of the comprehension model is called *explanatory coherence* (Graesser, Millis, and Zwaan 1997; Thagard 1989; Hobbs 1979) and is important for maintaining a level of *cognitive economy* in human comprehension.

Empirical Studies with *STAR*

As evidenced by our discussion so far, especially in the last part of the preceding section, the opportunity to interact with a developed system forces one to start thinking more concretely about some deep questions regarding our particular approach to story comprehension, but also more generally, regarding the logical nature of commonsense human reasoning. Beyond triggering one's thinking about these questions, the *STAR* system can also be used as an experimental tool to study those aspects of the questions that are ultimately empirical. In turn, gathered empirical evidence will help in the iterative improvement of our understanding and the further development and extension of the *STAR* system itself.

Our high-level empirical direction is to present the system with increasingly demanding stories. A corpus of stories can be either prepared by us (a task that we have, in fact, been engaged in), or obtained from existing repositories used for developmental tests in reading comprehension. At this stage the main aims of our experiments are the following:

Nature of representation: What are "natural representations" of world knowledge as used by humans for comprehension? What is the logical / argumentative nature of human reasoning in the practice of commonsense reasoning? What is an appropriate granularity of commonsense knowledge representation? Can natural representations as used by humans in comprehension be extracted or projected from more general and abstract representations?

Effective use of knowledge: How is world knowledge effectively retrieved and / or specialized to the context at hand? How can this process be automated so that the same WK can be used across different stories and contexts?

Coherence in comprehension: How can the notion of coherence be exploited to develop computationally more effective comprehension in automated systems? How is coherence related to the above issues: Does it affect the nature of representation and the effectiveness of WK use?

To make progress towards these aims, we are examining two classes of experiments, which we discuss next.

Psychological Experiments

In the first class of experiments emphasis lies on the human aspects of comprehension: How do humans represent and activate knowledge and which knowledge do they appeal to when comprehending stories? To what extent is this knowledge logical / argumentative in nature? How deep or shallow is the chaining of rules / arguments during comprehension? How can answers to these questions be gathered through established psychological experimental methodologies?

In an initial experiment, we have carried out a psychological study to ascertain the world knowledge that is activated to successfully comprehend example stories. We developed a set of inferential questions to follow the reading of the story in pre-specified sessions. These questions assessed the extent to which readers connected, explained, and elaborated key story elements. Readers were instructed to answer each question and to justify their answers using a "think-aloud" method of answering questions while reading, in order to reveal the world knowledge that they had used.

The qualitative data from the readers was pooled together and analysed as to the frequencies of the types of responses, in conjunction with the information given in justifications and think-aloud protocols. The analysis of these combined question and think-aloud data revealed: both the concepts and story elements that were the focus of encoding efforts and knowledge activation, and the activated knowledge that provided the basis for the inferences (generalizations, connections, associations, explanations, and predictions).

Using this information as the background world knowledge in the domain files for the *STAR* system, we confirmed that the system was able to produce a comprehension model that answered questions according to the majority of human answers, but also identified questions where humans had exhibited significant variability in their provided answers. The data from this experiment and the corresponding *STAR* domain files can all be found online on the system's website at: `http://cognition.ouc.ac.cy/narrative/`.

Another psychological experiment that we have recently completed aims to use a more systematic process to collect world knowledge from 60 participants across four different stories, all on the same theme of "life at home". The stories were presented sentence by sentence, but in different ways to each participant, e.g., different order, with or without a title, etc., and participants were asked to report in an informal structured natural language version of conditional rules the world knowledge they activated as part of their comprehension of each sentence. We then monitored the process of translating the world knowledge as produced by the readers into a machine readable format for the *STAR* system. We are currently examining to what extent this knowledge is appropriate and sufficient for the system to comprehend successfully the four stories, and / or seeking to identify what additional knowledge and input might be required for this.

We will also measure the correlation between the participants' level of natural language skills (ascertained through a pre-test experiment) and the quality of the knowledge they volunteered for the *STAR* system. Will the *STAR* system perform better in comprehending stories with the world knowledge from participants with higher levels of natural language skills? Such a positive correlation would be an encouraging indication of the ability of the *STAR* system to work with natural representations of commonsense knowledge.

Automated WK Acquisition

In the second class of experiments emphasis lies on the automated acquisition of WK: What representation facilitates best the acquisition process? What are the pros and cons of learning directly in the language of the *STAR* system, as opposed to learning in the language used by human participants in the psychological experiments and then translating that into the language of the *STAR* system? Is the gathered knowledge general / specific enough to be used for comprehending actual stories? What types of knowledge cannot be extracted automatically due to lack of (reliable) training data? Is the *distributional hypothesis*, that "a word is characterized by the company it keeps" (Firth 1957), sufficient to justify that the learned knowledge is commonsensical?

The importance of these experiments rests not only in offering an approach for collecting and representing commonsense knowledge in a machine readable form that could be more viable than crowdsourcing approaches (Rodosthenous and Michael 2014), but also in that they can provide knowledge in a general form, independent of particular stories.

As a first step, we will gather a small corpus of short stories on the same general theme, and crawl the Web for web-pages containing words found in the stories. Using the text in those web-pages, we will train machine learning algorithms to learn rules appropriate for drawing inferences (Michael and Valiant 2008; Michael 2009; 2013a), including rules that are causal in nature (Michael 2011), along with the relative priorities of the learned rules (Dimopoulos and Kakas 1995).

We will then experiment with different methods of reducing — at runtime, as stories are read by the *STAR* system — the learned story-independent world knowledge to a specific story context, and study the system's degree of comprehension, and its ability to solve other related tasks (Michael 2013b). In line with formal results suggesting that learning and reasoning cannot proceed independently (Michael 2014), a two-way feedback of the performance of the system to the knowledge acquisition methods will be used to iteratively improve the form of the learned knowledge and the comprehension performance of the *STAR* system.

Future Work

Working with the *STAR* system to date has admittedly raised more questions than what it has answered. However, it has also given us an empirical tool to investigate and start making progress towards answering these questions, following closely the scientific methodology of observing, hypothesizing, and evaluating. How should the acquisition of general knowledge and its contextual specialization be integrated into the system? Does the representation of the mental world through behavior rules, rather than a formal theory, suffice to replicate the way humans comprehend stories according to psychological experiments? We plan to push forward with empirical research, continuing to be guided by psychology, and attempting to understand how logic can serve to describe and computationally capture actual human reasoning.

References

Baroni, P.; Caminada, M.; and Giacomin, M. 2011. An Introduction to Argumentation Semantics. *Knowledge Engineering Review* 26(4):365–410.

Bench-Capon, T. J. M., and Dunne, P. E. 2007. Argumentation in Artificial Intelligence. *Artificial Intelligence* 171(10–15):619–641.

Diakidoy, I.-A.; Kakas, A. C.; Michael, L.; and Miller, R. 2014. Story Comprehension through Argumentation. In *Proceedings of the 5th International Conference on Computational Models of Argument (COMMA 2014)*, volume 266 of *Frontiers in Artificial Intelligence and Applications*, 31–42. Scottish Highlands, U.K.: IOS Press.

Dimopoulos, Y., and Kakas, A. 1995. Learning Non-Monotonic Logic Programs: Learning Exceptions. In *Proceedings of the 8th European Conference on Machine Learning (ECML 1995)*, volume 912 of *LNAI*, 122–137. Berlin: Springer.

Firth, J. R. 1957. A Synopsis of Linguistic Theory, 1930–1955. *Studies in Linguistic Analysis* 1–32. Reprinted in F.R. Palmer, ed. (1968). Selected Papers of J.R. Firth 1952-1959. London: Longman.

Gelfond, M., and Lifschitz, V. 1993. Representing Action and Change by Logic Programs. *Journal of Logic Programming* 17(2/3–4):301–321.

Graesser, A. C.; Millis, K. K.; and Zwaan, R. A. 1997. Discourse Comprehension. *Annual Review of Psychology* 48:163–189.

Hobbs, J. R. 1979. Coherence and Coreference. *Cognitive Science* 3(1):67–90.

Kakas, A. C., and Miller, R. 1997. A Simple Declarative Language for Describing Narratives With Actions. *Journal of Logic Programming* 31(1–3):157–200.

Kakas, A. C.; Michael, L.; and Miller, R. 2011. Modular-$\mathcal{E}$ and the Role of Elaboration Tolerance in Solving the Qualification Problem. *Artificial Intelligence* 175(1):49–78.

Kowalski, R., and Sergot, M. 1986. A Logic Based Calculus of Events. *New Generation Computing* 4(1):67–95.

McCain, N., and Turner, H. 1997. Causal Theories of Action and Change. In *Proceedings of the 14th AAAI Conference on Artificial Intelligence (AAAI 1997)*, 460–465. Menlo Park: AAAI Press.

McCarthy, J., and Hayes, P. J. 1969. Some Philosophical Problems from the Standpoint of Artificial Intelligence. *Machine Intelligence* 4:463–502.

Michael, L., and Valiant, L. G. 2008. A First Experimental Demonstration of Massive Knowledge Infusion. In *Proceedings of the 11th International Conference on Principles of Knowledge Representation and Reasoning (KR 2008)*, 378–389. Sydney, Australia: AAAI Press.

Michael, L. 2009. Reading Between the Lines. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI 2009)*, 1525–1530.

Michael, L. 2011. Causal Learnability. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI 2011)*, 1014–1020. Barcelona, Catalonia, Spain: IJCAI/AAAI.

Michael, L. 2013a. Machines with WebSense. In *Working notes of the 11th International Symposium on Logical Formalizations of Commonsense Reasoning (Commonsense 2013)*.

Michael, L. 2013b. Story Understanding... Calculemus! In *Working notes of the 11th International Symposium on Logical Formalizations of Commonsense Reasoning (Commonsense 2013)*.

Michael, L. 2014. Simultaneous Learning and Prediction. In *Proceedings of the 14th International Conference on Principles of Knowledge Representation and Reasoning (KR 2014)*. Vienna, Austria: AAAI Press.

Miller, R., and Shanahan, M. 2002. Some Alternative Formulations of the Event Calculus. *Lecture Notes in Artificial Intelligence* 2408:452–490.

Murphy, G. L., and Medin, D. L. 1985. The Role of Theories in Conceptual Coherence. *Psychological Review* 92(3):289–316.

Rodosthenous, C. T., and Michael, L. 2014. Gathering Background Knowledge for Story Understanding through Crowdsourcing. In Finlayson, M. A.; Meister, J. C.; and Bruneau, E. G., eds., *Proceedings of the 5th Workshop on Computational Models of Narrative (CMN 2014)*, volume 41 of *OASIcs - OpenAccess Series in Informatics*, 154–163. Quebec City, Canada: Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik.

Thagard, P. 1989. Explanatory Coherence. *Behavioral and Brain Science* 12(3):435–467.

Thielscher, M. 1999. From Situation Calculus to Fluent Calculus: State Update Axioms as a Solution to the Inferential Frame Problem. *Artificial Intelligence* 111:277–299.