Narrative Text Comprehension: From Psychology to AI

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

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

Loizos Michael
Open University of Cyprus
loizos@ouc.ac.cy

Rob Miller
University College London
rsm@ucl.ac.uk

Abstract

This paper reports on how work in psychology can inform research into automated narrative text comprehension. Specifically, we describe the psychology's view of how human readers combine information from text with their own commonsense world knowledge to build a single coherent model of a narrative. We present a framework that formalizes, using a suitable form of preference-based argumentation, this notion of comprehension and outline how this might form the basis for partially mechanizing the process of comprehension.

1 Introduction and Motivation

Text comprehension has long been identified as a key test for the success of Artificial Intelligence (AI). Aside from its central position in many forms of the Turing Test, it is clear that human computer interaction could benefit enormously from this and other forms of natural language processing. The rise of computing over the internet, where so much data is in the form of textual information in web pages, has given even greater importance to this topic.

This paper reports on a research programme aiming to study how we can automate *narrative text comprehension* (NTC). A key component of this program is to learn from the (extensive) study of text comprehension in Psychology in order to draw guidelines for developing frameworks for automating NTC, and in particular *story comprehension* (SC).

In most theoretical models of SC in Psychology, comprehension involves the construction of a mental representation to capture the meaning of the text. This representation is a temporary mental model or situation model formed in working memory that accounts for the information contained in the text (Graesser, Millis, and Zwaan 1997; Kintsch 1988; Zwaan and Radvansky 1998). Although the situation model is to be distinguished from the more permanent representational structures of world knowledge in long-term memory, its formation depends critically on the activation and use of these existing knowledge structures. Propositions derived from the story and prior world knowledge are integrated in a way that contributes to the situation model's coherence and elaboration (McNamara and Magliano 2009). The outcome of successful comprehension is a situation model that represents key story elements along with implications and relations that were left implicit in the story.

Our current focus at this initial stage of our research is on this particular task of synthesis of the explicit information that the story text provides with the background common sense or world knowledge that the reader has. We will call this task the *integration problem* of NTC. For the moment we will assume solved the orthogonal issue of correctly parsing the natural language of the text into some information-equivalent structured (e.g., logical) form. This is not to say that this issue is not an important element of NTC. Indeed, it may need to be tackled in conjunction with the integration problem on which we are focusing (since, for example, the problem of de-referencing pronoun and article anaphora could depend on background world knowledge and hence possibly on the higher-level whole comprehension of the text (Levesque, Davis, and Morgenstern 2011)).

The Psychology of Story Comprehension

We illustrate the main challenges that the problem of integration of textual and background world knowledge poses by considering the following simple story.

Story: It was the night of Christmas Eve. After feeding the animals and cleaning the barn, Papa Joe took his shotgun from above the fireplace and sat out on the porch cleaning it. He had had this shotgun since he was young, and it had never failed him, always making a loud noise when it fired.

Papa Joe woke up early at dawn, picked up his shotgun and went off to forest. He walked for hours, until the sight of two turkeys in the distance made him stop suddenly. A bird on a tree nearby was cheerfully chirping away, building its nest. He aimed at the first turkey, and pulled the trigger.

After a moment's thought, he opened his shotgun and saw there were no bullets in the shotgun's chamber. He loaded his shotgun, aimed at the turkey and pulled the trigger again. Undisturbed, the bird nearby continued to chirp and build its nest. Papa Joe was very confused. Would this be the first time that his shotgun had let him down?

A method used by psychologists to assess the level of comprehension is to pose inferential questions (at various stages in the exposition of the story) to examine the extent to which the reader is able to distinguish between answers that can plausibly be implied by the story and those that cannot.[1]

---

[1] See www.ucl.ac.uk/infostudies/rob-miller/papa-joe/ for a full questionnaire used with a set of human readers in the context of a

• After reading the first paragraph we can ask "Where does Papa Joe live?". The reader should understand that Papa Joe lives in the countryside. Successful comprehension requires the recognition of such an implied "truth" about the physical setting of the story (van den Broek 1994; Zwaan, Langston, and Graesser 1995), and its distinction from non-implied truths (e.g., Papa Joe lives in a city or a boat). Hence the reader should be able to *elaborate* further from what is given explicitly in the story by drawing on and connecting to relevant background world knowledge, e.g., barns with animals are normally found in the countryside, fireplaces and porches are normally in people's homes, etc. In psychological terminology, that Papa Joe lives in the countryside is a *bridging inference* that integrates story elements with each other and background knowledge (Gerrig and O'Brien 2005; Graesser, Millis, and Zwaan 1997).

• At the end of the second paragraph we can ask "Why was Papa Joe in the forest?", expecting the reader to be able to distinguish between the intention to hunt and other intentions such as that of bird watching or practicing shooting. This shows two important aspects of story comprehension:

1. Elaboration in comprehension is expected to link with *world knowledge about mental states* that the reader has about peoples' typical motivations, desires, intentions, feelings and thoughts, so that the reader can reason in these terms about the actors in the story. Such mental world knowledge is as important in comprehension as physical world knowledge. Actions and events in a story are structured on the basis of spatiotemporal and causal relations (Mandler and Johnson 1977; Rumelhart 1975). Inferences about motives and intentions supply important causal information that may be left implicit but which is critical in situating and explaining the actions described (Zwaan, Langston, and Graesser 1995).

2. Comprehension needs to establish *explanatory coherence* between the various pieces of information in the story. The first sentence of the second paragraph creates an expectancy of intentionality, but the reader is not expected to make explanatory guesses at this stage. Only after reading the entire second paragraph will the typical reader weed out less likely intentions and arrive at the explanatory inference that Papa Joe intends to hunt in the forest (rather than e.g., practice shooting). The need to establish explanatory coherence leads to inferences that can link a number of story actions and situations in a *coherent whole* that accounts for all of them (Albrecht and O'Brien 1993; Graesser, Millis, and Zwaan 1997). In contrast, an explanatory inference that the reason for being in the forest is to hear birds sing would indicate that the reader had linked only a few pieces of information, and had neglected to account for actions such as carrying the shotgun, stopping at the sight of turkeys, aiming and pulling the trigger.

• *Integration processes* are needed to consolidate the network of text-based and knowledge-based (inferred) propositions in a compact way so that only those that are in

"think-aloud" experiment to collect a sample of world knowledge and associated inferences typically used to comprehend this story.

some way strongly inter-related with each other remain in the model (Kintsch 1988; 2005; McNamara and Magliano 2009). After reading the second paragraph, a coherent model will no longer contain information about where the shotgun is kept (above the fireplace) or where it was cleaned (out on the porch) as no links are indicated between these facts and subsequent information. Also, a coherent model will not contain new pieces of information that can be drawn through a single isolated connection with some concept in the background world knowledge — e.g., in our story we would not draw the extra conclusion that the weather was mild out on the porch that night (an isolated association at best only peripheral to the story).

• Comprehension requires *cognitive economy*, because the limited cognitive resource of readers necessitates the activation of only a small restricted subset of the available conceptual world knowledge (Kintsch 1988; Gerrig and O'Brien 2005). So, after reading the first paragraph, readers are expected to activate knowledge they have about shotgun firing, firing sounds, and tasks related with shotgun maintenance (e.g., cleaning), but not knowledge about kinds or parts of a shotgun or the purpose of using a shotgun since this is not supported by explicit textual information. Cognitive economy is achieved through *selection or deletion operators* applied during integration. *Generalization* operators further reduce the network through knowledge-based inferences that summarize parts of it. Key elements, identified by repeated reference, and key relations, as indicated by readers' world and *story knowledge* (e.g., that stories describe events, events are spatially and temporally related, actions are causally related both to each other and to the intentions, desires and goals of the actors), are likely to remain in the model and, when appropriate, to be replaced by superordinate inferences that generalize across some of them. The intentionality inference of "hunting" represents such a generalization that subsumes most of the actions and states described in the second paragraph, as well as the spatiotemporal and causal relations between them.

• Integration can be seen as an iterative *general updating mechanism* of the comprehension model (Kintsch 1988). As each new piece of text information is encountered, its integration has the potential to revise the model, particularly when a "discontinuity" is encountered in the story — some unaccounted change involving the physical and temporal setting, the causal effects of actions, the story characters, or their intentions (Zwaan, Langston, and Graesser 1995). Readers typically respond to discontinuities by lowering the importance of previous elements in the model that are affected by the change, and constructing new substructures to account for it (Gernsbacher 1990; Rapp and Taylor 2004).

In our story we have an example of such a "discontinuity" after the first sentence of the third paragraph. Until then, we expect the reader to understand that the first turkey is injured or dead while the second one is alive (drawing from world knowledge about the firing of guns), but the reader is now expected to *revise* this and accept that both turkeys are alive. (In AI terms a revision like this comes from activating an *endogenous qualification* for the rule that pulling the trigger causes the shotgun to fire, based on the stronger asso-

ciation that unloaded guns do not fire.) The implied "truth" that the gun fired and any ramifications of this, such that the first turkey is injured, all need to be retracted. In other cases the discontinuity is not (immediately) explained by the story. (In AI terms this typically indicates an *exogenous qualification*.) Here, a minimal revision in the comprehension model is performed until story provides further explanation. For example, after the second sentence of the third paragraph in our story, the reader would again comprehend that the first turkey is shot. But after reading the third sentence this will conflict with the information that the bird nearby has not been startled by the noise of firing. Hence the implied "truth" that the shotgun has fired needs to be retracted from the comprehension model together with the ramification that the first turkey is injured or dead. But the reader is not immediately required to guess why the gun did not fire.

• Different readers may have *different, equally legitimate comprehension models*. Model differences can arise from differences in conceptual knowledge and differences in mental world knowledge (e.g., variability regarding likelihoods of motive, especially if these are not clarified by the text). But these are primarily differences in elaboration, not coherence. A reader who is a hunter herself might include in her representation an elaboration regarding the shotgun that would be missing from a less knowledgeable reader's model. Different models can be equally successful in terms of their coherence despite differences in elaboration.

2 Knowledge Representation and Reasoning

Our aim is to be guided by and build on the psychological research referred to above in order to develop a Knowledge Representation and Reasoning (KRR) framework specifically for NTC and SC. This KRR framework will be built using the technical know-how developed in AI for Reasoning about Action and Change and Default Reasoning and the integration of these areas (see (van Harmelen, Lifschitz, and Porter 2008) for an overview), but adapted according to the different perspective given by the task of SC. In particular, in order to reflect the central psychological SC task of *integration*, all types of reasoning (e.g., causal, persistence, predictive, property elucidation) in our system will be defeasible. Our initial conjecture has been that some of the other psychological features of SC, such as *coherency* and *cognitive economy*, can be modeled as heuristics, applied as refinements to an underlying process of default reasoning.

Clearly, a fundamental requirement of our framework is that the representation of world knowledge should facilitate efficient computation of comprehension models. A key observation from Psychology that can help us with this is that the background world knowledge used in SC is not in the form of an elaborate formal theory, but is better regarded as a collection of (relatively loose) semantic associations between concepts, reflecting typical rather than absolute facts and rules. Thus rule-based knowledge need not be fully qualified at the representation level, since it can be qualified via the reasoning process by the relative strength of other (conflicting) associations/rules in the knowledge.

As regards reasoning, again the psychological perspective gives us reason to depart from a standard view of drawing conclusions based on truth in all (preferred) models consistent with the story and world knowledge. Instead the emphasis can be on building one model from a collection of safe or *sceptical* properties that follow from the text as unqualified conclusions. Moreover, this single model need not be complete. It is a subset of conclusions grounded on the text of the story and based on a subset, which the reader can choose, of the available world knowledge. As the story progresses the development of the model should allow for its minimal revision in the face of new opposing information from the text.

To capture these features of representation and reasoning we will use preference-based argumentation (e.g., (Modgil and Prakken 2012)) to give us a unified approach for causal as well as default reasoning. The reasoning to construct a comprehension model will be based on building arguments and counter arguments, and qualification at all levels will be captured uniformly through an acceptability requirement on those arguments that support the conclusions in the model.

The Argumentation-Based Framework

We use a typical RAC language of Fluents, Actions, Times, with an extra sort of Actors. An actor-action pair is an ***event***, and a fluent/event or its negation is a ***literal***. For this work it suffices to represent times as natural numbers and to assume that time-points are sufficiently dense between story elements to allow for the realization of indirect effects.

For any fluent literal L, any fluent/event literal X, and any set S of fluent/event literals, we consider ***argument types*** as follows: ***property arguments*** $\mathrm{pro}(X, S)$; ***causal arguments*** $\mathrm{cau}(X, S)$; ***persistence arguments*** $\mathrm{per}(L, \{L\})$ (which we sometimes write as $\mathrm{per}(L, \cdot)$). A general argument of any type is denoted by $\mathrm{arg}_i(H_i, B_i)$.

A ***world knowledge theory*** $\mathcal{W}$ is a set of property and causal arguments together with a (partial) irreflexive priority relation on them. A ***narrative*** $\mathcal{N}$ is: a set of observations $\mathrm{OBS}(X, T)$ for a fluent/event literal X, and a time-point T; and a set of (story specific) property or causal arguments.

Definition 1. A ***story representation*** $\mathcal{SR} = \langle \mathcal{W}, \mathcal{N}, \succ \rangle$ comprises a world knowledge theory $\mathcal{W}$, a narrative $\mathcal{N}$, and a (partial) irreflexive priority relation $\succ$ extending that in $\mathcal{W}$ so: *(i)* $\mathrm{cau}(H, B_1) \succ \mathrm{per}(\neg H, B_2)$; *(ii)* $\mathrm{per}(H, B_1) \succ \mathrm{pro}(\neg H, B_2)$; *(iii)* $\succ$ may prioritize between arguments in $\mathcal{N}$ and those in $\mathcal{W}$ (typically the former over the latter).

A representation $\mathcal{SR}$ of our example story (focusing on its ending) may include the following arguments in $\mathcal{W}$ and $\mathcal{N}$ (where pj is short for "Papa Joe", our main story character):

$c1 : \mathrm{cau}(fired\_at(pj, X), \{aim(pj, X), pull\_trigger(pj)\})$
$c2 : \mathrm{cau}(\neg alive(X), \{fired\_at(pj, X), alive(X)\})$
$c3 : \mathrm{cau}(noise, \{fired\_at(pj, X)\})$
$c4 : \mathrm{cau}(\neg chirp(bird), \{noise, nearby(bird)\})$
$c5 : \mathrm{cau}(gun\_loaded, \{load\_gun\})$
$p1 : \mathrm{pro}(\neg fired\_at(pj, X), \{\neg gun\_loaded\})$
$p2 : \mathrm{pro}(\neg fired\_at(pj, X), \{\neg noise\})$ (story specific)
$p3 : \mathrm{pro}(\neg noise, \{chirp(bird)\})$

with $p1 \succ c1$, $p2 \succ c1$, $p3 \succ c3$; and the following in $\mathcal{N}$:

$\mathrm{OBS}(alive(turkey), 1),$ $\mathrm{OBS}(aim(pj, turkey), 1),$
$\mathrm{OBS}(pull\_trigger(pj), 1),$ $\mathrm{OBS}(\neg gun\_loaded, 4),$

OBS($load\_gun$, 5),   OBS($pull\_trigger(pj)$, 6),
OBS($chirp(bird)$, 10),   OBS($nearby(bird)$, 10),

with the exact time-point choices being inconsequential.

To make sense of stories, we define an **interpretation** Δ of $\mathcal{SR}$ to be a sequence of tuples $\langle \text{arg}(H, B), T^h, d, X, T \rangle$, for an argument $\text{arg}(H, B)$ in $\mathcal{SR}$, a fluent/event literal X, a direction $d \in \{\text{F}, \text{B}\}$, time-points T^h, T. Δ **supports** a set of fluent/event literals M **at** T, if for every $X \in M$, either $\langle \text{arg}(H, B), T^h, d, X, T \rangle \in \Delta$ or OBS(X, T) $\in \mathcal{N}$. Δ captures the inferences that one may draw (from the narrative or through arguments), which, in the spirit of cognitive economy, may be a subset of all possible ones. T^h captures the time-point at which the head of the argument applies, while X and T capture the inference drawn from that argument.

To make precise the inference process, say that argument $\text{arg}(H, B)$ **on** T^h: **forward activates** X **at** T^h **under** Δ if $X = H$ and Δ supports B at $T - \{\langle Y, T \rangle \mid Y \in B\}$ is the activation **condition**; and **backward activates** X **at** T **under** Δ if $\neg X \in B$ and Δ supports $\{\neg H\}$ at T^h and $B \setminus \{\neg X\}$ at $T - \{\langle \neg H, T^h \rangle\} \cup \{\langle Y, T \rangle \mid Y \in B \setminus \{\neg X\}\}$ is the activation **condition**; in either case, $T = T^h$ if $\text{arg}(H, B)$ is a property argument, and $T = T^h - 1$ for the other arguments.

The use of arguments over only neighboring time-points aids in our later development of a computational model in terms of a stepwise transition between time-points.

Definition 2. An interpretation Δ is **grounded** by $\mathcal{SR}$ if it can be obtained by starting from an empty Δ and repeating the following: Choose any $\text{arg}(H, B)$ on T^h that forward / backward activates X at T under Δ, and append $\langle \text{arg}(H, B), T^h, d, X, T \rangle$ to Δ, with $d = \text{F}/\text{B}$, respectively. A tuple's **condition** is the associated activation condition. Δ is **safe** if it includes no $\langle \text{arg}_1(H_1, B_1), T_1^h, d_1, X_1, T_1 \rangle$ and $\langle \text{arg}_2(H_2, B_2), T_2^h, d_2, X_2, T_2 \rangle$ with $H_1 = \neg H_2, T_1^h = T_2^h$.

Much like the activation network suggested by psychologists (cf. Section 1), the iterative construction of Δ effectively builds a graph of inferences, grounded in the story narrative, and guided by the available world knowledge. Due to the default nature of arguments, certain ways of chaining them are "unsafe"; we shall illustrate this notion later on.

Consider the end of the second paragraph of our example story, corresponding to time-points $1 - 3$ in our example narrative. Note that the empty Δ supports $aim(pj, turkey)$ and $pull\_trigger(pj)$ at 1. Hence, $c1$ on 2 forward activates $fired\_at(pj, turkey)$ at 2 under the empty Δ. We can thus populate Δ with $\langle c1, 2, \text{F}, fired\_at(pj, turkey), 2 \rangle$. Similarly, we can include $\langle \text{per}(alive(turkey), \cdot), 2, \text{F}, alive(turkey), 2 \rangle$ in the new Δ. Under this latter Δ, $c2$ on 3 forward activates $\neg alive(turkey)$ at 3, allowing us to further extend Δ with $\langle c2, 3, \text{F}, \neg alive(turkey), 3 \rangle$. The resulting Δ is a grounded interpretation that supports $\neg alive(turkey)$ at 3. It is based on this inference that we expect readers to respond that the first turkey is dead, when asked about its status at this point.

Reading the first sentence of the third paragraph, we learn that OBS($\neg gun\_loaded$, 4). We expect that this new piece of evidence will lead readers to revise their inferences so far. The following definitions aid in formalizing the situation.

Fix $\alpha_1 = \langle \text{arg}_1(H_1, B_1), T_1^h, d_1, X_1, T_1 \rangle$ and $\alpha_2 = \langle \text{arg}_2(H_2, B_2), T_2^h, d_2, X_2, T_2 \rangle$. Tuple α_1 **undercuts** tuple α_2 if $d_1 = \text{F}$, $\text{arg}_2(H_2, B_2) \not\succ \text{arg}_1(H_1, B_1)$, $H_1 = \neg H_2$, $T_1^h = T_2^h$. An undercut exists when an argument used in the forward direction takes priority over another argument, thus undermining the *process* of using the latter argument to draw an inference, regardless of the inference itself. Tuple α_2 is **disputed** (**on** X_2 **at** T_2) **by** $\mathcal{SR}$ if OBS($\neg X_2, T_2$) $\in \mathcal{N}$. Tuple α_2 is **disputed** (**on** X_2 **at** T_2) **by** tuple α_1 if $X_1 = \neg X_2$, $T_1 = T_2$. A dispute exists when the *inference* activated by an argument is in conflict with either the story or the inference of another argument, regardless of the arguments used.

A grounded interpretation A_1 of $\mathcal{SR}$ **disputes** (on X at T) a second one A_2, if a tuple in A_2 is disputed (on X at T) either by the story or a tuple in A_1. A_1 **attacks** A_2 if: either some tuple in A_1 undercuts some tuple in A_2; or some tuple in A_2 is disputed by the story; or some tuple $\alpha_2 \in A_2$ is disputed by a tuple $\alpha_1 \in A_1$, but α_2 does not undercut α_1.

Definition 3 (Admissible Interpretation). Let $\mathcal{SR}$ be a story representation. A safe grounded interpretation Δ of $\mathcal{SR}$ is **admissible** if Δ does not attack itself, and it attacks any other safe grounded interpretation of $\mathcal{SR}$ that attacks Δ.

Continuing with our illustration, we wish to show that Δ from earlier is not admissible. Consider the grounded interpretation A that includes the following in the given order:
$\langle \text{per}(gun\_loaded, \cdot), 4, \text{B}, \neg gun\_loaded, 3 \rangle$
$\langle \text{per}(gun\_loaded, \cdot), 3, \text{B}, \neg gun\_loaded, 2 \rangle$
$\langle p1, 2, \text{F}, \neg fired\_at(pj, turkey), 2 \rangle$.
To reason that the gun was unloaded even before it was observed to be so, we appeal to the *backward persistence* of $\neg gun\_loaded$, by using $\text{per}(gun\_loaded, \cdot)$ contrapositively to backward activate an inference, capturing a form of proof by contradiction: had the gun been loaded at 3, it would have been so at 4 which would contradict the story.[2]

Since $c1 \not\succ p1$, the last tuple in A undercuts the first one in Δ; so A attacks Δ. Since Δ does not attack A, Δ is not admissible. One could extend Δ to counterattack A. Since $p1$ on 2 backward activates gun_loaded at 2 under Δ, we can include $\langle p1, 2, \text{B}, gun\_loaded, 2 \rangle$ in Δ. This new Δ attacks A, since $\langle \text{per}(gun\_loaded, \cdot), 3, \text{B}, \neg gun\_loaded, 2 \rangle \in A$ is disputed by $\langle p1, 2, \text{B}, gun\_loaded, 2 \rangle$ without the former undercutting the latter. Thus, Δ counterattacks through the same argument that A used to attack! This would trivialize the reasoning process, had it been not for an important point: this Δ is no longer safe, and hence it is not admissible.

Observe that Δ uses $\langle c1, 2, \text{F}, fired\_at(pj, turkey), 2 \rangle$ and $\langle p1, 2, \text{B}, gun\_loaded, 2 \rangle$ to infer gun_loaded at 2, as follows: assume $\neg gun\_loaded$ at 2; by $p1$ on 2 it follows that $\neg fired\_at(pj, turkey)$ at 2; by $c1$ on 2 it follows that $fired\_at(pj, turkey)$ at 2; a contradiction, therefore, gun_loaded at 2. Although using $c1$ in the forward direction, and then chaining $p1$ in the backward direction is valid when reasoning with classical rules, this is not the case when reasoning with arguments. Indeed, $c1$ on 2 cannot be used to infer $fired\_at(pj, turkey)$ at 2, because under the assump-

---

[2] As one would expect, this forward persistence of gun_loaded could be undercut by a causal argument for $\neg gun\_loaded$ on 4.

tion $\neg gun\_loaded$ at 2, $p1$ undermines the use of $c1$. This is precisely captured by saying that Δ is not safe.

The process of understanding our story may then proceed by building a Δ that includes the tuples in A, along with $\langle per(alive(turkey), \cdot), T, \textsc{f}, alive(turkey), T \rangle$ for $T = 2, 3, 4$, resulting in an admissible interpretation that supports $alive(turkey)$ at 4. It is based on this inference that we expect readers to respond that the first turkey is alive.

Making sense of the remaining story proceeds analogously. After Papa Joe loads the gun and fires again, one draws the inferences that the first turkey is dead, that noise was caused, and that the bird stopped chirping as a result, through the arguments $c1, c2, c3, c4$. The observation that the bird is still chirping offers an attack on all these inferences through their backward use. Attempting to remedy the situation leads one to build a Δ that includes persistence and the arguments $p2, p3$ to undercut the firing of the gun.

One of the aims of our framework is to distinguish between the inferences that are accepted, rejected, or allowed given a story. The distinction of $\{accepted, allowed\}$ from $\{rejected\}$ is accommodated by the notion of admissibility. But given psychological evidence suggesting that human readers draw only inferences that are necessary (or sceptical), we wish to single-out the case of accepted inferences.

Definition 4 (Comprehension Model)**.** Let $\mathcal{SR}$ be a story representation. An admissible interpretation Δ of $\mathcal{SR}$ is a ***comprehension model*** of $\mathcal{SR}$ if there exists no admissible interpretation of $\mathcal{SR}$ that disputes Δ (on X at T).

A comprehension model, thus, accommodates the distinction of $\{accepted\}$ from $\{allowed, rejected\}$. One may wish to partly re-introduce the distinction between $allowed$ and $rejected$, as this would be useful in the empirical evaluation, where the multiple-choices of a question might not include any accepted answers, and the reader might be expected to discriminate between allowed and rejected answers. We can do so by adapting our definition of a comprehension model, so that instead of disallowing the existence of other admissible interpretations that dispute Δ, we simply ask that these disputes are only on one of the choices C_i offered as possible answers to a question, and that Δ disputes back on some other choice from within C_i; this gives rise to a non-deterministic split in a controlled manner.

Computing Comprehension Models

Following psychological evidence, we wish to compute a comprehension model by considering the narrative incrementally, and updating Δ as each new part of the story is made available. We do so by considering graph structure that succinctly encodes all interpretations that are relevant.

A grounded interpretation of $\mathcal{SR}$ is effectively a directed acyclic graph $\mathbb{G}$ with tuples $\langle \texttt{arg}(H, B), T^h, d, X, T \rangle$ being its vertices, and such that for each $\langle X, T \rangle$ in the condition of any given tuple in $\mathbb{G}$, either there exists an edge to that tuple from a tuple $\langle \texttt{arg}(H, B), T^h, d, X, T \rangle$ in $\mathbb{G}$, or $\textsc{obs}(X, T) \in \mathcal{N}$. We shall call such a graph ***grounded***.

Consider the operation that removes from $\mathbb{G}$ the edge between any pair of tuples that violate the safeness condition, and subsequently iteratively removes from $\mathbb{G}$ every vertex (and its outgoing edges) whose presence in $\mathbb{G}$ would make it non-grounded. We shall call the resulting graph ***safe***, as it effectively corresponds to a safe grounded interpretation.

Let $\mathcal{SR}[T']$ be obtained from $\mathcal{SR}$ by restricting the narrative up to time-point T'. Let $\mathbb{G}[T']$ be the safe graph obtained as specified above from the maximal grounded graph when considering observations in $\mathcal{SR}[T']$ only, and such that $T^h \leq T'$ for every $\langle \texttt{arg}(H, B), T^h, d, X, T \rangle \in \mathbb{G}[T']$.

Given a comprehension model $\Delta[T - 1]$ of $\mathcal{SR}[T - 1]$, the algorithm sets $\Delta[T] = \Delta[T - 1]$, and then repeats the following until some externally provided condition is met:

- Set $\Pi[T] = \texttt{retract}(\Delta[T], \mathbb{G}[T])$
- Set $\Delta[T] = \texttt{expand}(\Pi[T], \mathbb{G}[T])$

Process $\texttt{retract}(I, G)$ is defined to return the grounded interpretation I' obtained from I after removing every element from I that is attacked by G, and then iteratively removing from I' every tuple $\langle \texttt{arg}(H, B), T^h, d, X, T \rangle$ whose presence in I' would make it non-grounded. Process $\texttt{expand}(\cdot, \cdot)$ can be chosen to accommodate the variability in comprehension that is observed even among humans with effectively the same world knowledge. In particular, $\texttt{expand}(I, G)^3$ can return *any* interpretation I' that extends I by choosing an argument $\texttt{arg}(H, B)$ on T^h that forward / backward (with $d = \textsc{f}/\textsc{b}$) activates X at T under I, such that I' is safe, not self-attacking, and $\langle \texttt{arg}(H, B), T^h, d, X, T \rangle$ is not attacked by any $\langle \texttt{arg}_2(H_2, B_2), T_2^h, d_2, X_2, T_2 \rangle \in G$.

Theorem 1. Each iteration returns a comprehension model $\Delta[T]$ of $\mathcal{SR}[T]$, in time polynomial in the size of $\mathcal{SR}[T]$.

3 Evaluation through Empirical Studies

A widely-used method of evaluating the level of comprehension is through multiple-choice questions after reading the story ("offline"). However, this provides no information as to how the unfolding of events in a story contributes to the actual model construction and the world knowledge that is used to support it. "Online" measures, such as thinking aloud after reading certain text segments are more likely to provide data about the actual processes and their content (Crain-Thoreson, Lippman, and McClendon-Magnuson 1997).

For our investigation, we developed a set of inferential multiple-choice questions for the Papa Joe story and carried out an initial "think-aloud experiment" with a small group of people using a mixed ("online" and "offline") methodology: inferential questions were inserted after certain text segments and readers were asked to justify their answers, and encouraged to reveal the world knowledge that they had used. Out of the 14 college students that participated in this experiment, 6 demonstrated gaps in their representation. The other 8 readers who successfully comprehended the story served as an example of its coherent comprehension. This low performance can be explained on the basis of motivational grounds: findings have highlighted the pervasive influence of readers' goals and motivations for reading in determining to a large extent the level of comprehension achieved

---

[3]This effectively corresponds to computing a partial subset of the standard grounded extension of the argumentation framework.

(van den Broek et al. 2001). The students in our sample were instructed to simply read and understand the story in the context of voluntary study participation. The story was neither part of their own reading choices nor part of any course reading requirements.

With this empirical data we can test our framework's ability to capture the majority answers, and to account for their variability. We illustrate this through Question 6: "What was Papa Joe doing in the forest?", asking about Papa Joe's motive, and offering "Practice Shooting", "Hunting" "Catch Turkeys" and "Bird Watching"as candidate answers.

Although a common sense world knowledge could indeed sanction all these motives by containing simple unconditional property arguments that support them, it was evident from the responses of the students that these motives can be "derived" from higher-level desires or goals of the actor. Following theories of agency, e.g., (Rao and Georgeff 1995; Kakas et al. 2008), we can link these motives and intentions to the actor's desires through the property arguments

$\text{pro}(intention(Person, hunt\_for(Object)),$
$\{wants(Person, food\_for(Occ, Object), hunter(Person))\}),$
$\text{pro}(motive(walking(Person, Forest), hunt\_for(Object)),$
$\{intention(Person, hunt\_for(Object))\}),$

showing how the high-level desire for a type of food for a certain occasion becomes the support (in a sense a mental cause) for a hunter to have the motive for hunting for that type of food. By letting $Person = papa\_joe$, along with
$\text{pro}(wants(Person, food\_for(dinner, turkey), \{xmasDay\}),$
we have a supported inference for "hunting" as Papa Joe's motive for walking in the forest.

Such high-level desires and intentions are examples of *generalizations* that contribute to the coherence of the comprehension model, and to the creation of *expectations* in readers about the course of action that the story might follow. Expectations are formed according to the general *story knowledge* (Brewer and Lichtenstein 1982) that normally complications are bound to arise while fulfilling desires and achieving intentions: stories are expected to contain "surprise suspense moments" where such difficulties arise.

The latter two answers to question 6 would be rejected by any comprehension model when the WK contains stronger property arguments against these of the form:
$\text{pro}(\neg motive(walking(Person, Forest), catch\_birds),$
$\{fired\_at(Person, Birds)\})$
This argument activates its inference in any comprehension model that supports $fired\_at(papa\_joe, turkey)$, as shown in the previous section. Furthermore, there are several world knowledge counter arguments against "practice shooting" such as, that people do not go for practice shooting on Christmas Day, or that they do not walk for hours in the forest before shooting; the latter represented as:
$\text{pro}(\neg walk\_for\_hours(Person, Forest),$
$\{motive(walking(Person, Forest), practice\_shooting)\})$
Since $walk\_for\_hours(Person, Forest)$ is supported directly by the story, this argument would backward activate the negation of its body, so that no comprehension model with the $practice\_shooting$ motive would be true.

Hence the only answer that can be accepted for question 6 is that of "hunting", and indeed in our experiment there was no variability of answers, with all 8 students choosing "hunting". Clearly, one way for variability to arise is for different readers to use different parts of their WK in building their comprehension model. This was observed, for example, when students chose either "farm" or "village" as answers to Question 1: "Where did Papa Joe live?". Our framework accommodates this variability in terms of comprehension models as extended at the end of Section 2 to accommodate choices among allowed (but not accepted) answers. Variability arises by choosing to activate one of
$\text{pro}(lives(Person, village), \{lives(Person, country\_side)\}),$
$\text{pro}(lives(Person, farm), \{lives(Person, country\_side)\}),$
the inferences of both of which are activated in a comprehension model that includes $lives(papa\_joe, country\_side)$, which itself is supported since Papa Joe's home has a barn.

Another example of variability occurred in the answers for the group of questions 7,8,10,11, asking about the status of the turkeys. The majority of students followed a comprehension model as analyzed in the previous section. However, a (minority) group of students consistently answered that both turkeys were alive in all four questions. These readers had ignored or defeated the causal arguments that supported the inference that the first turkey was dead, perhaps based on the expectation that the desire of Papa Joe for turkey and his intention to hunt for it would not be met without complications. Such an expectation could be grounded in the story, from cues such as the reference to "pulled the trigger" instead of "fired". We believe that such expectations can be generated from standard *story knowledge* in the same way as we draw other elaborative inferences from WK.

4 Related Work

Automated story understanding has been an ongoing field of AI research for the last forty years; for a good overview see (Mueller 2002) and the website (Mueller 2013). Logic-related approaches have largely been concerned with the generation of appropriate representations, translations or annotations of narratives, with the implicit or explicit assumption that standard deduction or logical reasoning techniques can subsequently be applied to these (see e.g., the discussion in (Mueller 2003)). To our knowledge, little other work has been undertaken on developing psychology-inspired automated reasoning techniques specific to SC. Our treatment of events, causality and persistence arises from the Event Calculus (EC) (Kowalski and Sergot 1986) and its descendants, and the EC has also been used for narrative annotation in (Mueller 2003) and as a semantics for natural language in (van Lambalgen and Hamm 2005). Many other authors have emphasized the importance of commonsense knowledge and reasoning in SC, e.g., (Dahlgren, McDowell, and Stabler 1989; Mueller 2004), and how it can offer a basis for SC tasks beyond question answering (Michael 2013b). Finally, the potential synthesis of NTC and SC with the humanities field of *narratology* is discussed in (Mani 2013).

5 Conclusions and Future Work

This work has set up a conceptual framework for story comprehension using established AI techniques and knowhow

from temporal reasoning and argumentation. Unlike other formal reasoning frameworks, our work is strongly guided by Psychology research and challenges borne out of it, the messages from which we have taken very seriously when seeking to address some of the challenges that we have identified in Section 1. We have given a proof of concept of the applicability of our framework through an in-depth analysis of an example story following established methodologies from Psychology.

We aim to carry out a systematic (semi-automatic) evaluation of our framework through a series of tests taken from established standardized corpora. For this we would need to: *(i)* systemize the representation language for the background world knowledge (perhaps building on lexical databases such as WordNet (Miller 1995), FrameNet (Baker, Fillmore, and Lowe 1998), and PropBank (Palmer, Gildea, and Kingsbury 2005)), exploring the possibility of populating these theories using existing archives for common sense knowledge (such as Cyc (Lenat 1995)) or through the automated extraction of commonsense knowledge from text using natural language processing (Michael and Valiant 2008), and appealing to textual entailment for the semantics of the extracted knowledge (Michael 2009; 2013a); and *(ii)* address further computational issues, such as the challenge of cognitive economy and coherence. We will investigate whether the latter can be addressed by applying "computational heuristics" on top of (and without the need to reexamine) the solid semantic framework that we have developed thus far, and whether we will be able to draw from Psychology again to formulate such heuristics. In particular, we expect that the psychological studies will guide us in modularly introducing various computational operators such as *selection, dropping and generalization operators* for selecting which part of the world knowledge to use in forming the model, what earlier elaborative inferences to drop from the model as relatively unimportant and to summarize (parts of) the information in the story and model so far through generalizations that capture succinctly the message that the story wants to convey. These operators are instrumental in reducing the computation of arguments and attacking counterarguments in the (further) construction of the comprehension model and through which the comprehension model is focussed as the story progresses.

Our aim to test the approach systematically on a number of cases will require that we address and test our knowledge representation framework at two levels. At one level, story comprehension is based on knowledge about the structure, the content, and the function of the genre (Zwaan, Langston, and Graesser 1995). Readers of any story expect to read about characters (agents) that have and act according to goals and motivations, but who are also frustrated in their attempts to fulfil them. Thus we are challenged to have an effective representation of general story knowledge that would include (meta) knowledge about *reader expectations*. On the other hand, as we have seen in this paper, story comprehension is also based on the activation of content-related conceptual knowledge which does not necessarily generalize across stories with different content. Therefore, the evaluation of our framework requires a two-pronged approach:

evaluating the extent to which our framework can account for the comprehension of two different sets of short stories: a set of stories with a similar underlying theme and context as the target story employed in this study (i.e., life in the countryside) and a second set of short stories with different themes and contexts.

References

Albrecht, J. E., and O'Brien, E. J. 1993. Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19:1061–1070.

Baker, C. F.; Fillmore, C. J.; and Lowe, J. B. 1998. The Berkeley FrameNet Project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics - Volume 1*, ACL '98, 86–90.

Brewer, W., and Lichtenstein, E. 1982. Stories are to entertain: A structural-affect theory of stories. *Journal of Pragmatics* 6:473–486.

Crain-Thoreson, C.; Lippman, M. Z.; and McClendon-Magnuson, D. 1997. Windows on comprehension: reading comprehension processes as revealed by two think-aloud procedures. *Journal of Educational Psychology* 89:579–591.

Dahlgren, K.; McDowell, J.; and Stabler, E. 1989. Knowledge representation for commonsense reasoning with text. *Computational Linguistics* 15(3):149–170. http://acl.ldc.upenn.edu/J/J89/J89-3002.pdf.

Gernsbacher, M. A. 1990. *Language comprehension as structure building*. Hillsdale, NJ: Erlbaum.

Gerrig, R. J., and O'Brien, E. J. 2005. The scope of memory-based processing. *Discourse Processes*.

Graesser, A. C.; Millis, K. K.; and Zwaan, R. A. 1997. Discourse comprehension. *Annual Review of Psychology* 48.

Kakas, A. C.; Mancarella, P.; Sadri, F.; Stathis, K.; and Toni, F. 2008. Computational Logic Foundations of KGP Agents. *J. Artif. Intell. Res. (JAIR)* 33:285–348.

Kintsch, W. 1988. The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review* 95:163–182.

Kintsch, W. 2005. An overview of top-down and bottom-up effects in comprehension: The C-I perspective. *Discourse Processes*.

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

Lenat, D. B. 1995. CYC: A Large-Scale Investment in Knowledge Infrastructure. *Communications of the ACM* 38(11):32–38.

Levesque, H. J.; Davis, E.; and Morgenstern, L. 2011. The Winograd Schema Challenge. In *Proceedings of the Thirteenth International Conference on Principles of Knowledge Representation and Reasoning*.

Mandler, J. M., and Johnson, N. S. 1977. Remembrance of things parsed: Story structure and recall. *Cognitive Psychology* 9:11–151.

Mani, I. 2013. *Computational Modeling of Narrative*. Morgan and Claypool.

McNamara, D. S., and Magliano, J. 2009. Toward a comprehensive model of comprehension. *The Psychology of Learning and Motivation* 51:297–384.

Michael, L., and Valiant, L. G. 2008. A First Experimental Demonstration of Massive Knowledge Infusion. In *Proc. of 11th International Conference on Principles of Knowledge Representation and Reasoning (KR'08)*, 378–389.

Michael, L. 2009. Reading Between the Lines. In *Proc. of 21st International Joint Conference on Artificial Intelligence (IJCAI'09)*, 1525–1530.

Michael, L. 2013a. Machines with Websense. In *Proc. of 11th International Symposium on Logical Formalizations of Commonsense Reasoning (Commonsense'13)*.

Michael, L. 2013b. Story Understanding... Calculemus! In *Proc. of 11th International Symposium on Logical Formalizations of Commonsense Reasoning (Commonsense'13)*.

Miller, G. A. 1995. WordNet: A Lexical Database for English. *Communications of the ACM* 38(11):39–41.

Modgil, S., and Prakken, H. 2012. A general account of argumentation with preferences. *Artificial Intelligence* 195:361–397.

Mueller, E. 2002. Story understanding. In Nadel, L., ed., *Encyclopedia of Cognitive Science*, volume 4, 238–246. London: Nature Publishing Group.

Mueller, E. 2003. Story understanding through multi-representation model construction. In Hirst, G., and Nirenburg, S., eds., *Text Meaning: Proceedings of the HLT-NAACL 2003 Workshop*, 46–53. East Stroudsburg, PA: Association for Computational Linguistics.

Mueller, E. 2004. Understanding script-based stories using commonsense reasoning. *Cognitive Systems Research* 5(4):307–340.

Mueller, E. 2013. Story understanding resources. http://xenia.media.mit.edu/ mueller/storyund/storyres.html. Accessed February 28, 2013.

Palmer, M.; Gildea, D.; and Kingsbury, P. 2005. The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics* 31(1):71–106.

Rao, A. S., and Georgeff, M. P. 1995. BDI Agents: From Theory to Practice. In *First International Conference On Multi-Agent Systems (ICMAS-95)*, 312–319.

Rapp, D. N., and Taylor, H. A. 2004. Interactive dimensions in the construction of mental representations for text. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30:988–1001.

Rumelhart. 1975. Notes on a schema for stories. In Bobrow, D. G., and Collins, A., eds., *Representation and understanding: Studies in cognitive science*, 211–236. New York: Academic Press.

van den Broek, P.; Lorch, R. F.; Linderholm, T.; and Gustafson, M. 2001. The effects of readers' goals on inference generation and memory for texts. *Memory and Cognition* 29:1081–1087.

van den Broek, P. 1994. Comprehension and memory of narrative texts: Inferences and coherence. In Gernsbacher, M. A., ed., *Handbook of psycholinguistics*, 539–588. London, UK: Academic Press.

van Harmelen, F.; Lifschitz, V.; and Porter, B. 2008. *Handbook of Knowledge Representation*. Elsevier Science. ISBN: 978-0-444-52211-5.

van Lambalgen, M., and Hamm, F. 2005. *The Proper Treatment of Events*. Blackwell.

Zwaan, R. A., and Radvansky, G. A. 1998. Situation models in language comprehension and memory. *Psychological Bulletin* 123:162–185.

Zwaan, R. A.; Langston, M. C.; and Graesser, A. C. 1995. The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science* 6:292–297.