ACS Poster Collection (2018) 1–19 Published 8/2018; Paper 2

Recognizing and Systematically Highlighting Intentions

Suri Bandler SCB@MIT.EDU
Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA, 02139 USA

Abstract

To communicate effectively, humans must analyze not only what is said, but also how it is said. I propose a cognitive theory of how humans glean information about a communicator's intentions from key characteristics of wording. The theory explains how we assemble textual evidence such as passive voice, instances of harm, and use of hedging words such as *alleged* to tell a coherent, compositional story of the communicator's rhetorical goals. I demonstrate this theory with a computational implementation. The implementation reads short news-like stories in simple English and identifies modulations in text that reveal the author's intent to influence three areas—sympathy, agency, and doubt. The system gathers objective evidence using a system of modular experts, interprets the evidence with culturally-specific subjectivity models, and distills the potentially-conflicting interpretations into a short, coherent argument about the author's intentions.

1. A Computational Approach to Critical Communication

A full model of human intelligence requires a meaningful understanding of human communication. In order for this understanding of communication to go beyond the superficial, such a model would need to be capable of understanding not just what is said, but how it is said. For instance, a news agency might report a police shooting either as "Adult kills Minor" or as "Terrorist is killed by Police." Although both can describe the same event, each statement might have a drastically different impact on a reader.

To advance this understanding, I developed a cognitive model of the process by which humans identify the ways an author's modulation of a story affects how it is understood. I then implemented the theory in a computational system called RASHI, which reads short news-like stories consisting of approximately 20 sentences in simple English such as the following characteristic example:

> **Regime slaughters the freedom fighter.** The Regime is a government. The regime is ruthless. Max is a boy and Max fights for freedom. Max helps our people. Max is a freedom fighter. Two barbarians were executed by the freedom fighter and the regime alleges the boy's actions to be terrorism. The boy had no choice. The regime has killed Max. The regime alleges the legitimacy of the boy's death. The regime prevents freedom. We wait for the boy's body. We want to honor him with a funeral. Tonight we will host a memorial service. This was a barbaric attack. We increase the resistance. Join us and fight with us for freedom.

I note that the same elements of this story could be used to describe a different perspective in which Max's actions are categorized as terrorism. In Section 4, I present one such story and describe the role of rhetorical modulation in identifying the contrasting authorial intentions.

Using a principled theory of human rhetorical techniques as described in this paper,[1] RASHI argues about how the author's word choices serve communicative goals such as affecting the reader's perceptions of sympathy, blameworthiness, and doubt. Given the above story, RASHI argues about how the author's word choice modulates blame:

> **Modulation of blame.** Overall, the author directs the majority of blame at the regime. Not only did the author blame the regime most frequently, but also the author casts favor on other agents. The author did so by excusing other agents' actions by casting their inequities as retribution, while not doing so for the regime. Similarly, the author casts favor on other agents as compared to the regime by referring to other agents as having done good deeds, while not doing so for the regime. Although the author uses passive voice for the regime, they do so only once and so it cannot be concluded that the author mitigates this blame by describing the regime's action as passive.

The cognitive theory and corresponding program are humanly plausible by design, namely no capabilities that go beyond the scope of human ability are assumed or required. In particular, the theory is systematic (using a system of specialized experts to identify cues in the text), structural (understanding text not as a bag-of-words, but as meaningful in relation to its components), principled (relying on explicitly articulated behavioral and knowledge hypotheses), and explanatory (able to form arguments and defend them with textual evidence and definite reasons). I demonstrate how the RASHI system, based on this theory and armed with story understanding (Winston, 2011), goes beyond generating binary classification or relying on simple keywords and human experts. The theory, and accordingly the implementation, are crucially based upon story understanding to facilitate a full, systematic interpretation of a communicator's modulation with explanation at every step.

In the sections that follow, I first describe my cognitive model of rhetorical intelligence. I argue that a story understanding approach lets both the theory and the implementation be humanly plausible. Then, by way of elaborating the theory, I present my detailed computational implementation. The system reads a simple news article, finds textual evidence indicating the author's rhetorical choices, evaluates those choices, and provides a summarized hypothesis of the author's intent. I compare my theory against prior work and conclude with a discussion of my contributions and exciting next steps and future applications.

2. We Choose How to Tell a Story: A Cognitive Model of Rhetorical Intelligence

To begin, I present a cognitive theory of human rhetorical intelligence and a concrete formulation of the commonsense knowledge upon which this intelligence relies. My cognitive theory consists of three claims about how humans deploy rhetorical intelligence. I claim that humans:

1. I note that this work focuses on *recognition* of author intent rather than on *generation* of text according to author intention, but the same principles discussed in this paper can be applied in either direction. For an example of text generation with respect to a communicator's intentions see Hovy (1987).

Approach a story as a composite. Rather than treating an article as a bag-of-words or relying on the presence or absence of individual keywords, human readers get a sense for the story by understanding the structure of sentences and paragraphs and by analyzing the story as a whole to determine a coherent, global argument about the author's intent. Even within the same sentence, the author's choices can interact in complex ways. Human readers identify not only words with negative sentiment, such as harm, but also strategic choices such as using passive voice or characterizing the victim as a child or a terrorist—factors that influence who gets blamed for the harm and how much. Similarly, when forming an overall argument about the author's intent, readers consider the full article, gathering evidence to support their argument and even distilling potentially conflicting evidence to form a coherent portrait of the author's intentions.

Draw on domain-specific expertise. While reading a story, humans draw upon commonsense knowledge not only to make inferences regarding events in the story itself—as is typically acknowledged or implemented in current systems—but also to identify rhetorical cues that signal authorial intent. If an author says "The bank was robbed" then a human reader will recognize that the author fails to implicate an agent and instead focuses on the event itself rather than the perpetrator. In contrast, if an author uses passive voice in a scientific article in a biology journal, a human reader will recognize that this rhetorical choice simply reflects industry standard and does not carry implications about authorial intent. Just as cognitive models rely on commonsense knowledge regarding causality in events, so too must a cognitive model capable of reasoning about an author's intentions rely on commonsense knowledge about rhetorical implications.

Can justify their claims about rhetorical effect at every step. When asked about the implications of an author's word choice, humans can point to specific peices of evidence within the story, summarize the available evidence, present an argument, and identify how the evidence supports or weakens that argument.

In addition to these three claims, I present a conceptual formulation of the kind of knowledge that humans require in order to identify and interpret rhetorical choices (Table 1). This knowledge consists of the intentions that an author might possess—such as instilling credibility and evoking sympathy—and potential rhetorical mechanisms for achieving these intentions.

3. A Computational Approach to Recognizing and Highlighting Intentions

I developed the computational system RASHI to both demonstrate the viability and to refine the details of my cognitive theory. Armed with story understanding, my system reads a story, gathers explicit and objective textual evidence, interprets the evidence using models of the author's subjective frame of mind, and distills the potentially contradictory results into an argument about authorial intent. I discuss each of these capabilities in turn, using the previous *Regime slaughters the freedom fighter* story as a characteristic example.

Figure 1 depicts the flow of data within RASHI, summarizing this section.

Table 1. Examples of authorial intentions and the rhetorical mechanisms that authors can employ to achieve them. Underlined mechanisms indicate techniques implemented in my computational system.

Author intention	Example mechanisms
Cast doubt	Use terms such as "alleged" or "believed", use sarcasm or scare quotes
Evoke sympathy	Emphasize age/role in society, use colorful quotations or analogies, mention specific individuals, emphasize magnitude
Convey Importance	Repeat ideas, use passive versus active voice, change sentence order, emphasize impact, evoke feelings of unity/solidarity
Modulate blame	Use passive vs. active voice, justify via assigning role/title, claim expertise via location of writing, use idioms, make generalities
Instill credibility	Quote statistics, cite an expert, imply via section in paper published, demonstrate novelty
Align stories	Reference, make metaphor or allusion
Create intrigue	Omit information, use hyperbole, reference a specific or underdeveloped example
Cause surprise, humor	Break expectations (see for example (Taylor, 2018))
Make moral claim	Write from specific location, use first-hand accounts, assign epithets, make accusations, use idioms/generalities, use passive voice

3.1 Genesis Reads the Story

RASHI is built as a module on top of the Genesis Story Understanding System (Winston, 2014). Building RASHI on a story-understanding substrate ensures that not only the theory but also the components of the implementation realistically model what humans can actually do—they meet a computational imperative (Winston & Holmes, in press). Accordingly, RASHI's analysis consists of story-enabled operations such as pattern matching, rule-based inferences, and building complex nested representations.

I note that compositionality is a key feature of RASHI's story-enabled approach: RASHI does not merely haphazardly identify keywords or snippets of evidence but instead composes explanations in terms of the meaning of the constituents of sentences and of the story overall. This is particularly important as it is often the interaction between authorial choices that have a profound impact on gaining insight into an author's overall intentions. For example, an author may believe that harm is not always wrong, but justified in instances of self-defense. RASHI is capable of handling such concepts by, for instance, not just looking for helpful or harmful action keywords but for the relationship between subject, action, and object.

Additionally, composition allows for explainability, as RASHI can use constituent components of a given sentence or of a story as pieces of evidence supporting a final conclusion. An example of using constituent components in rhetorical analysis would be distilling an author's intention at

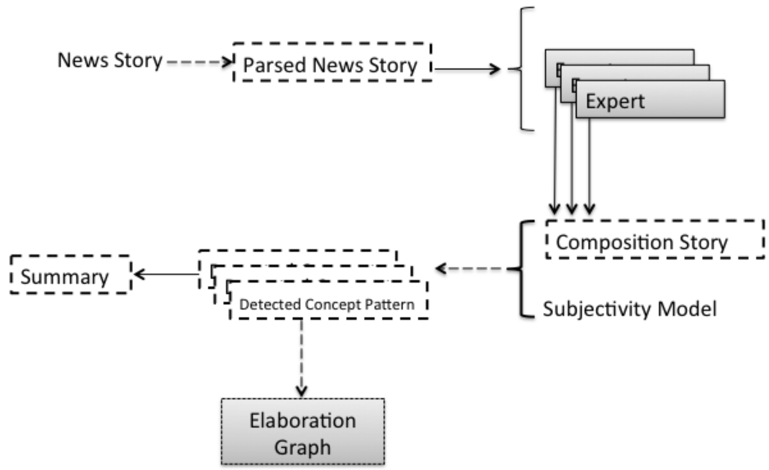

Figure 1. A depiction of the flow of data in RASHI. Plain text in the figure, such as News Story, indicate plain text input. Dotted arrows indicate that the processing is handled by the Genesis Story Understanding System while a solid arrow indicates that the processing is handled by RASHI. Dotted boxed text, such as Parsed News Story, indicates that data that is represented internally in the Genesis System. All such data can be externalized by Genesis into simple English. Grey boxes indicate a module of logic, where a solid border indicates that it is handled by RASHI and a dotted border indicates that it is handled by Genesis. In this figure, a News Story in simple, plain text is parsed into a Parsed News Story by Genesis. Then, RASHI applies its Experts, compiling the output of each into the composition story, which is represented internally in Genesis. Genesis then uses a provided subjectivity model, which is comprised of plain text descriptions of concept patterns, to filter the composition story, outputting a set of detected concept patterns. This set of concept patterns can be visualized as an elaboration graph by Genesis, or can be summarized by RASHI.

a story level: although an author might blame an agent in a specific sentence, a coherent global assessment might determine that an author mitigates this blame by demonstrating that the victim was deserving or that the agent performed many other good deeds, too. I give an example of one approach to such a process in Section 3.4. Story understanding, therefore, provides a robust foundation for modeling rhetorical analysis.

The Genesis Story Understanding System takes as input a textual story and parses it using the START parser (Katz, 1997). Due to current limitations of the parser, which is not the focus of this work, input stories must be composed in simple English. Commonsense knowledge in the form of rules supplies missing logical and causal connections, and knowledge in the form of concept patterns helps identify overarching themes such as revenge. If-then rules (e.g. `If xx refers to yy as a victim, then xx evokes sympathy for yy.`) can be deductive or inductive. Concept patterns are small stories that Genesis can identify within a larger story. Some concept patterns contain a "consequently" clause that is inserted into the original story whenever the concept pattern matches. For example, the "Victimhood" concept pattern is defined as follows:

`Start description of "Victimhood". xx is an entity. yy is a person. xx refers to yy as a victim. Consequently, xx evokes sympathy for yy. The end.`

In particular, RASHI uses the Genesis Story Understanding System for parsing stories, analyzing Genesis's internal, hierarchical representation of sentences rather than the text directly. Because of this structure it is possible to identify roles in a sentence like subject and object and parts of speech, used for identifying the action of a sentence. The START parser also exposes markers such as when a sentence is passive. Finally, RASHI uses Genesis to recognize the concept patterns that comprise an author's subjectivity model, as discussed in Section 3.3.

To begin the analysis of authorial intent, RASHI takes as input a textual story consisting of about 20 simple English sentences. Throughout this work, the process for composing input stories was as follows: I would find example articles in the news or compose a sample news story from multiple perspectives. Then, I would feed it through the parser, adapting the level of language to allow for parsing. Importantly, I gathered and composed these stories independently of implementing RASHI. Instead, I gathered my corpus of stories in conjunctions with delineating the potential rhetorical intentions and example mechanisms outlined in Table 1. The corpus of generated stories included but was not limited to depictions of police shootings, terrorist attacks, children playing together, students winning awards, and teachers helping students. Then, independently of the stories and based upon Table 1, I designed and implemented RASHI. Finally, I used the stories to garner the success of my system, in particular in order to better *generalize* its capabilities and to handle edge cases. As a result, RASHI can handle any story that START can parse, producing more interesting results from stories that present elements of the rhetorical devices explored in Table 2 and discussed below.

3.2 RASHI's Experts Gather Explicit Evidence

After reading the story, RASHI looks for textual evidence of the author's rhetorical choices to begin the process of determining the author's intentions. Table 1, shown previously, lists a number of *possible* intentions and their corresponding mechanisms. Of course, humans are more sophisticated than this table indicates, possessing many more possible motivations and means of achieving them. This table, however, gives insight into the breadth of what humans are capable of and identifies ways to get concrete value from such a huge space of possibilities. The present computational system can detect a subset of these mechanisms. Using a system of Experts, where each Expert is a module responsible for identifying one specific rhetorical indicator, RASHI finds textual evidence that could indicate authorial intent. Implementing these Experts required making specific implementation choices; I summarize my Experts and implementation choices in Table 2. As shown in the rightmost column of Table 2, each Expert detects rhetorical indicators using one of a few related techniques. The most common technique is to search for word meaning. *Property finding*, as used by the Evasiveness and Enemy Experts, consists of finding properties in a predefined set. These properties

Table 2. RASHI gathers objective rhetorical evidence using these eleven experts.

Expert name	Knowledge domain	Detection technique
ABUSE EXPERT	Recognizes perpetrator of harm or violence	**Property Finding:** Action Valence (subject of negative action)
VICTIM EXPERT	Recognizes the recipient of harm or violence	Action Valence (recipient of negative action)
BENEFACTOR EXPERT	Recognizes the agent of a positive action	Action Valence (subject of positive action)
BENEFICIARY EXPERT	Recognizes the recipient of a positive action	Action Valence (recipient of positive action)
EVASIVENESS EXPERT	Recognizes hedging tactics	**Property Finding:** Presence of Qualifier (e.g., alleged, believed)
ENEMY EXPERT	Recognizes possible enemies	**Synonym Matching:** Wordnet (e.g., bad-person, wrongdoer, terrorist)
YOUTH EXPERT	Recognizes young entities	Wordnet (e.g., Juvenile, minor, etc.)
KARMA EXPERT	Recognizes when help and harm enemies (i.e., deserving / undeserving figures).	**Misc.** Victim + Beneficiary + Enemy Expert
PASSIVE-VOICE EXPERT	Recognizes use of passive verb and lack of agent.	Presence of parser tag
ALIAS EXPERT	Consolidates different names that refer to the same individual (anaphora resolution)	Anaphora Resolution (e.g., via classifications: Mark is a terrorist. The terrorist...).
EMPHASIS EXPERT	Detects emphasis	Repetition Detection

can either be directly stated by the author in a particular sentence, such as in the case of "alleged terrorist", or inferred from a previous characterization. For example, if Max is previously referred to as a child, then RASHI will consider this property in future instances such as when Max is referred to as a victim. Alternatively, *Synonym matching*, as used by the Youth and Abuse Experts, uses WordNet (Fellbaum, 1998) to determine which words are close in meaning. For example, RASHI's Youth Expert will match any type of juvenile, including minor, child, and kid.

Two Experts employ special-purpose patterns that aren't about word meaning. The Passive Voice Expert simply checks the relevant lexical property found during parsing. The Repetition Expert, which activates after all other Experts, looks for repeated use of rhetorical strategies as found by other Experts, such as the repeated characterization of the same individual as a victim. Its search is supported by an Alias Expert that consolidates different names for the same individual to ensure accurate counting (i.e. which performs simple anaphora resolution). Finally, some Experts search not only the story itself, but the output of other Experts. The Karma Expert identifies harm dealt by and to enemies using information provided by the Victim Expert and the Enemy Expert. In this way, the Karma Expert provides a nuanced compositional understanding in terms of more basic Experts: the author may render a harm more forgivable by casting its victim as an enemy.

Note that, as mentioned in the previous section, these matching techniques are more than just keyword matching or bag-of-word routines as they are both word and context dependent. When finding objective textual evidence, RASHI takes into account sentence structure, linking the textual evidence to the original source as well as noting any information regarding position in the sentence or parse. As an example, consider a step-by-step view into RASHI processing the following sample sentence: "Police kill attacker." First, RASHI parses this sentence using Genesis and the START parser. Then, each Expert analyzes the parsed version of this sentence or analyzes the output of other Experts. Running this sentence through RASHI's system of Experts, the Abuse Expert recognizes that kill is a verb that induces harm and uses the parsed form of the sentence to identify the *subject*, namely the Police, as the perpetrator. The Victim Expert also searches for instances of harm but instead makes note of the *object* of the sentence in question, here noting that the attacker is the recipient of harm. Already, RASHI's Experts rely upon the structure of a sentence rather than simply recognizing specific words. Going a step further, the Karma Expert turns on after the Victim Expert and uses the Enemy Expert to determine and make note if the author casts the victim in question as a potential enemy, here using WordNet to determine that an attacker is a form of wrong-doer. The Karma Expert, therefore, relies upon the context of the action, namely that harm was perpetrated against a potential enemy. Note no conclusions have yet to be reached about what these rhetorical choices indicate about the author's intentions!

Importantly, RASHI's system of Experts is both impartial (in that it strictly identifies textual evidence of rhetorical mechanisms without drawing conclusions) as well as modular (in that it is flexible to the addition of Experts and to using a multiplicity of methods). We can see this in the previous example, as no conclusions are reached by the Karma expert about the implications of the author's choices and the Karma Expert uses the output of both the Victim Expert and the Enemy Expert.

RASHI's Experts are impartial In this evidence-gathering stage, RASHI's Experts exhaustively search for all instances of a rhetorical device without drawing any conclusions regarding relevance or coherence. Intentionally separating evidence gathering and analysis allows RASHI to analyze authorial intent on a global level, taking into account both the author and the set of observations in its entirety. When processing the story as a composite, certain observations may turn out to be inconsequential, while others may be contradictory. For example, consider the sentence *"Child is killed by alleged terrorist."* Here, the available evidence simultaneously suggests that the terrorist should be both blamed, because they hurt a child, as well as exonerated because their role as a terrorist is alleged and the killing is described in passive voice. The implications of such contradictions are not yet noted, handled instead at a later stage. At this point, all possible matches are added, allowing RASHI to remain impartial in the evidence gathering stage.

RASHI's Experts are modular The cognitive theory itself remains uncommitted to particular implementation details. To reflect this, RASHI is modular; future Experts can easily be added and existing Experts can be changed to use various underlying implementations. For example, a different implementation of the Victim Expert could use an alternative knowledge base such as ConceptNet (Liu & Singh, 2004) rather than WordNet synonyms to classify an action as causing harm or benefit. The modular implementation allows for the simultaneous use of a multiplicity of humanly plausible methods.

The textual evidence collected in this stage describe the author's rhetorical decisions. Consisting of a sequence of sentences in Genesis's inner language, the collection of evidence comprises a *composition story.* For example, the following is the composition story translated to English by Genesis for the characteristic example:

Composition Story. The author refers to our people as a beneficiary. The author says that Max carries out a kindness. The author uses passivity for the regime. The author labels Max as passivity's object. The author refers to Max as a victim. The author says that the regime commits an inequity. The author uses passivity for Max. The author labels two barbarians as passivity's object. The author refers to two barbarians as a victim. The author says that Max commits an inequity against a malefactor. The author refers to Max as a victim. The author says that the regime commits an inequity. The author says that the regime alleges something. The author says classification to be alleged. The author says that the regime alleges something. The author says death's legitimacy is alleged. The author refers to Max as a youth. The author refers to Max as a victim repeatedly. The author says the regime commits an inequity repeatedly. The author says the regime alleges something repeatedly.

Note that the above composition story is an *externalization* of Genesis's inner language representation into English which is why the phrasing is somewhat stilted.

RASHI's implementation demonstrates that designing within a specific domain can have a large impact: with just a few mechanisms, RASHI displays a powerful ability to identify a wide variety potential modulating factors.

3.3 RASHI Interprets Textual Evidence and Focuses Intent Using Subjectivity Models

After RASHI gathers the composition story, it must evaluate and interpret the evidence as indicative of the author's intent to modulate meaning. The significance of a given piece of evidence depends on the author's own values or priorities and on their cultural background, including nationality, age, profession or the forum for which they are writing. For example, some cultures may view harm against a child as especially egregious; for such cultures, describing a victim as a child is a significant rhetorical move. As another example, passive voice may be a communal expectation. Consider scientific writing in the biology community—community standards dictate the use of passive voice, especially in a methodology section. In that case, the use of passive voice does not carry the same connotations that it might in a newspaper article describing a crime. If we are to not only describe but interpret rhetorical choices, we must be able to model the author's world view and goals.

To this end, I developed *subjectivity models*, modular representations of an author's values, cultural connotations, and priorities. Subjectivity models consist of Genesis rules and concept patterns that connect the explicit textual information gathered in the composition story to a set of highly personal connotations and implications. Subjectivity models let RASHI interpret textual choices as

Table 3. A subjectivity model for a sympathy-focused author. This model contains concept patterns for an author who comes from a culture in which children are more highly protected.

```
If xx refers to yy as a victim, then xx elicits sympathy for yy.
If xx evokes yy's childhood and xx elicits sympathy for yy, then xx increases
sympathy for yy.
If xx refers to yy as a victim and xx labels yy as passivity's object, then xx
increases sympathy for yy.
Start description of "Childhood". xx and yy are entities. xx refers to yy as a youth.
Consequently, xx evokes yy's childhood. The end.
Start description of "Victimhood". xx is an entity. yy is a person. xx refers to yy as
a victim. Consequently, xx evokes sympathy for yy. The end.
```

rhetorically meaningful or meaningless. They let RASHI interpret textual choices as delivering a particular subjective effect and therefore serving a particular rhetorical goal.

Subjectivity models can encode not only cultural values but also areas of importance upon which to focus. For example, an individual author might prioritize modulating agency in a story. What it means to modulate agency will differ between authors, and so a subjectivity model can serve to focus the analysis on a specific topic while also clearly delineating an author's values. Table 3 shows a sympathy-focused subjectivity model for an author who comes from a culture in which children are highly protected. Table 4 shows an agency-focused subjectivity model, including concept patters of Passivity, Blame, Retribution, Misguided Kindness, and Kindness and two rules. RASHI uses the Genesis Story Understanding System to detect instances of these concept patterns within the composition story, which depicts the full set of the author's rhetorical choices. Currently, subjectivity models are hand-crafted to represent a cohesive set of priorities. In the future, I hope to provide RASHI with the capability of hypothesizing over a universe of subjectivity models about which most likely represents an author's priorities.

In this way, subjectivity models implement knowledge of *authorial purpose* as described in the table of author intentions and example mechanisms (Table 1). By describing an author's focus or cultural perspective, subjectivity models provide tools to systematically uncover an author's intentions from rhetorical choices. Effectively, concept patterns in a subjectivity model provide the mapping backwards from rhetorical technique to underlying authorial intention. When RASHI applies the agency-focused subjectivity model to the example article's composition story, RASHI generates the elaboration graph in Figure 2 detailing its conclusions.

As an aside, a simplifying assumption I make with these subjectivity models is that an author's own values will dictate how the author will attempt to modulate the story to affect an audience. This assumes that the author's values align perfectly with the audience's, eliminating some theory-of-mind difficulties. An interesting future direction would be to enable RASHI to reason about an author's mental model of the audience.

Table 4. An example subjectivity model for an agency-focused author. Five concept patterns and two rules model how agency changes when the passive voice is used and how agency translates into blame, retribution, kindness, and misguided kindness and both filters and highlights relevant pieces of a composition story.

```
If xx decreases yy’s agency, then xx mitigates yy’s blame.
If xx says that yy commits an inequity, then xx blames yy.
Start description of "Passivity". xx uses passivity for yy. Consequently, xx decreases yy’s agency. The end.
Start description of "Blame". xx and yy are entities. xx says that yy commits an inequity. Consequently, xx blames yy. The end.
Start description of "Retribution". xx and yy are entities. xx says that yy commits an inequity against a malefactor. Consequently, xx mitigates blame against yy. The end.
Start description of "Misguided Kindness". xx and yy are entities. xx says that yy carries out a kindness against a malefactor. Consequently, xx casts doubt on yy’s intentions. The end.
Start description of "Kindness". xx and yy are entities. xx says that yy carries out a kindness. Consequently, xx casts admiration on yy’s intentions. The end.
```

3.4 RASHI Distills Evidence into a Coherent Argument

In the previous stage, RASHI used a subjectivity model to determine what textual evidence is relevant to the given author and to generate conclusions about the implications of the author’s rhetorical choices. This procedure is exhaustive, identifying many, potentially conflicting, pieces of evidence. Thus, in the next stage, RASHI tries to reconcile the available evidence, consolidating it into a coherent picture of the author’s intentions.

For the example article and the agency-focused subjectivity model (Table 4), RASHI generates the following summary:

> **Modulation of blame.** Overall, the author directs the majority of blame at the regime. Not only did the author blame the regime most frequently, but also the author casts favor on other agents. The author did so by excusing other agents’ actions by casting their inequities as retribution, while not doing so for the regime. Similarly, the author casts favor on other agents as compared to the regime by referring to other agents as having done good deeds, while not doing so for the regime. Although the author uses passive voice for the regime, they do so only once and so it can not be concluded that the author mitigates this blame by describing the regime’s action as passive.

RASHI uses a series of hand-coded heuristics that comprise a decision tree to summarize the author’s intentions (see similar discourse production techniques, e.g. (Davey & Longuet-Higgins, 1978)). The summary process forms an argument regarding a specific goal stemming from a central concept pattern. In the example summary, the central concept pattern is “Blame" from the agency-focused subjectivity model found in Table 4. The process is the same regardless of the topic of focus and is described below.

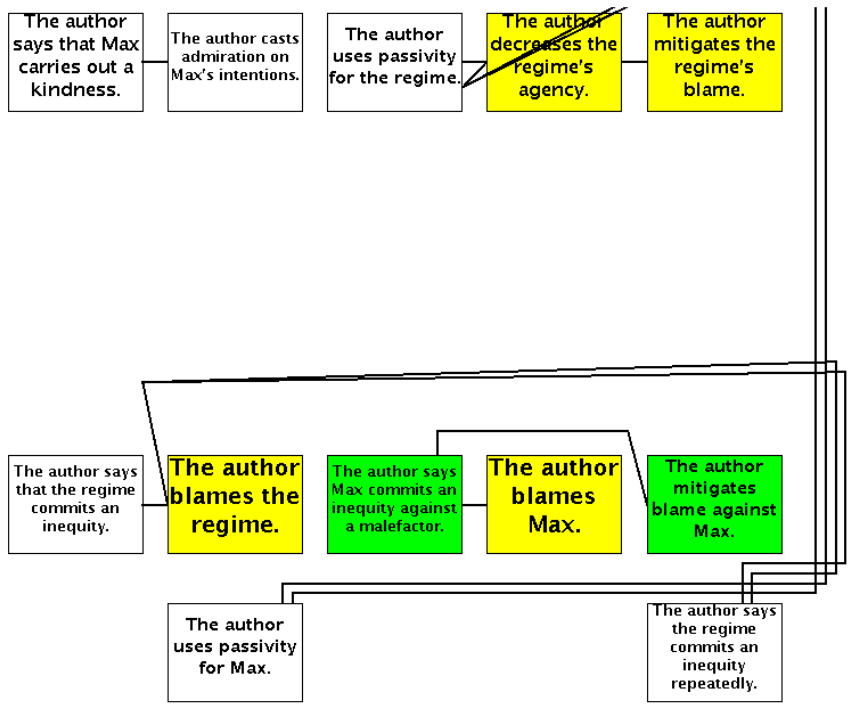

Figure 2. This elaboration graph depicts the textual evidence that RASHI found relevant to the author's intention to modulate the text regarding agency. RASHI uses the agency-focused subjectivity model (Table 4) to determine which evidence is relevant in the composition story and to draw immediate conclusions. Here, RASHI finds evidence of Agency and Passivity, Blame, Retribution, and Kindness but does not find evidence of Misguided Kindness.

The summarizer first tallies the different participants in the "Blame" concept patterns occurring throughout the story. The idea is that repetition can help resolve conflicting or ambiguous authorial cues. For example, repetition can serve as a heuristic to identify an author's primary target, as a consistently blamed character, presumably, indicates that the author deliberately intends to place blame.

Next, having found a character who stands out as a primary target, the summarizer searches for secondary effects that may mitigate or exacerbate the blame placed on the target. The summarizer checks to see whether blame is exclusively placed on the target or whether the blame is spread across others. Using hand-coded knowledge about the relationship between concept patterns (such as blame, retribution, and kindness), the summarizer also checks which characters are portrayed as having done kindnesses and whether any blameworthy actions constitute acts of retribution. When the target character is excused in this way, the summarizer describes the blame as comparatively less; when other characters are excused in this way, the summarizer describes the target's blame as comparatively greater (an approach taken by, for example, Sayan, 2014).

To generate natural-sounding discourse, the summarizer effectively walks down a tree, determining at each stage if the latest evidence strengthens or contradicts previous observations, and filling in the appropriate built-in text template accordingly. For example, RASHI first "branches" according to if the most blamed character is the *only* blamed character (choosing between "Overall, the author directs the majority of the blame at xx" or "Overall, the author unilaterally blames xx"). Then, if the author exacerbated the character's guilt in any way, RASHI adds "Not only did the author blame xx most frequently, but also...". If RASHI finds multiple exacerbating factors, such as casting others' harm as retribution and describing others as having done good deeds, RASHI bridges the explanations with "Similarly,...". The opposite effect occurs when there is evidence of the author mitigating the blame rather than exacerbating it.

Of course, at present, these summarizations techniques are hand-coded and are meant to be a systematic means of capturing the domain of knowledge currently embodied in RASHI. The summarizer is build on top of an extendable framework and exciting next steps include exploring additional domains and building the decision tree directly from concept patterns.

4. RASHI Recognizes Many Different Author Motivations

To show the breadth of RASHI's capabilities, in this section I demonstrate RASHI's results on two example stories and three subjectivity models. The first example story is *Regime slaughters freedom fighter*, as described above. The second story is an alternative telling of the first story from a different perspective:

> **A mother and child were murdered by a terrorist.** The killing is a terror attack. Terror attacks hurt our city. The mother and child were going to the doctor. The mother and child waited at the bus stop. The terrorist killed the mother and child with kitchen knives. The mother was killed by the terrorist while protecting the child. The terrorist stabbed the mother and child in the back. The funerals are tomorrow and the city encourages the community to come. The city will increase security because of the terrorist attack. People need to remain alert.

Below, I juxtapose RASHI's conclusions regarding agency (Table 4), sympathy (Table 3), and doubt (Table 5) in these two stories, where each concepts is defined according to the subjectivity model referenced by figure.

Table 5. A subjectivity model for a doubt-focused author.

```
If xx alleges yy, then xx casts doubt on yy.
Start description of "Object of Doubt". The author says that yy is alleged. Consequently,
the author casts doubt on yy's label. The end.
Start description of "Decrease Credibility". xx and yy are entities. xx says that yy
alleges something. Consequently, xx decreases yy's credibility. The end.
```

Table 6. A comparison of RASHI analyzing two different stories describing the same event along three dimensions: sympathy, agency, and doubt.

Regime slaughters freedom fighter.	*Mother and child murdered by terrorist.*
AGENCY. Overall, the author directs the majority of blame at the regime. Not only did the author blame the regime most frequently, but also the author casts favor on other agents. The author did so by excusing other agents' actions by casting their inequities as retribution, while not doing so for the regime. Similarly, the author casts favor on other agents as compared to the regime by referring to other agents as having done good deeds, while not doing so for the regime. Although the author uses passive voice for the regime, they do so only once and so it cannot be concluded that the author mitigates this blame by describing the regime's action as passive.	**AGENCY**. Overall, the author unilaterally blames the terrorist. The author's use of passive voice for the terrorist does not substantially mitigate the placement of blame because the author unilaterally and repeatedly blames the terrorist.
SYMPATHY. Overall, the author evokes the most sympathy for Max. Not only did the author refer to Max as a victim most frequently, but also the author emphasizes this sympathy by referring to Max as a child. Similarly, but to a lower degree, the author emphasizes sympathy for Max by using the passive voice when describing Max as receiving harm. By doing so, the author increases attention on Max's victimhood.	**SYMPATHY**. Overall, the author evokes the most sympathy for the mother. Not only did the author refer to the mother as a victim most frequently, but also the author emphasizes this sympathy by using the passive voice when describing the mother as receiving harm. By doing so, the author increases attention on the mother's victimhood.
DOUBT. Overall, the author solely questions the credibility of the regime. The author does so by stating that the regime made qualified claims, such as by 'alleging', 'suspecting', or 'believing' rather than directly claiming or stating. At the same time, because the regime qualified its remarks, the author casts doubt on the regime's claims as well. More specifically, the author casts doubt on the boy's terrorism and the boy's death's legitimacy.	[No modulation of doubt.]

I note that, here, the subjectivity models act like filters, serving to focus RASHI on a specific area of relevance to the author. It is possible that an author might care about all of these things and nothing prohibits RASHI from using a combined subjectivity model. To demonstrate the wide variety of conclusions reached, I include RASHI's summarizer's output for each combination of area of focus and example story.

While contradictions are reconciled via the summarizer, it is still interesting and important to see the apparent contradictions in an author's rhetorical choices. Beyond striving for a coherent argument about the author's intentions, RASHI brings to light the many nuances that exist within the same piece and can potentially be used to help authors ensure that their rhetorical choices match their implicit intentions during the editing process. Additionally, RASHI allows readers to note possible concessions made by the author, which can be helpful during political negotiation in terms of finding a potential avenue for compromise.

5. Related Work

As a computational theory of modulated communication, RASHI connects to work in related fields such as sentiment analysis, framing theory, and cognitive systems.

Sentiment Analysis RASHI resembles other computational approaches to understanding authorial intent, such as machine learning techniques for sentiment analysis. But whereas sentiment analysis aims to classify text fragments as belonging to one of a number of predetermined sentiments (such as "very negative", "negative", "neutral", "positive" or "very positive,") (see for example (Pang et al., 2002; Radford et al., 2017; Socher et al., 2013)), RASHI gathers evidence using principled expert knowledge to produce coherent arguments, explains its decisions, and provides a full analysis of authorial intent rather than classification across discrete categories.

Even approaches that are technically compositional in design lack the same level of story-based explainability. For example, Socher et al. (2013) trained a neural network to use a parse tree of a text fragment as input and to determine its sentiment through a recursive processing of nodes in the parse tree, updating its understanding as it climbs. Although this treats a sentence as a composition of nodes, the system is neither capable of explaining its conclusions for a given example nor of providing a interpretable, detailed analysis, generating a numerical result and a discrete classification. RASHI furthermore differs from most computational approaches to sentiment analysis (Radford et al., 2017; Socher et al., 2013) in that it is built on a humanly-plausible story-enabled model.

Framing Theory RASHI fits naturally within the field of framing theory, which studies how communicators elicit specific emotions or mental states in their audience by invoking a particular subset of the audience's beliefs (Entman, 1993). Indeed, RASHI's subjectivity models can be viewed as the collection of framing techniques at the author's disposal.

My theory distinguishes itself through a set of additional commitments. First, I argue for the importance of a systematic and computational approach. Previous work in framing theory has revolved around case studies, with researchers and the field itself accumulating observations over time. In contrast, story understanding provides RASHI with the tools to *integrate* fundamental observations into compositional knowledge modules in order to accumulate evidence resulting in a systematic framework of author intentions. Similarly, whereas framing theory lacks a coherent definition (Hallahan, 1999) and often uses vague, "casually-defined" categories (Entman, 1993), I developed a concrete framework of authorial motivations, which I demonstrated through a precise, principled computational implementation. Combined, story understanding serves as a cohesive, cognitive, and computational foundation for studying framing theory. Hallahan (1999) posits that storytelling is the most complex form of framing, but I claim that it is the exact opposite: framing is a complex form of story understanding and so it is best modeled using the conceptual constituents of stories.

Second, whereas framing theory tends to focus on the audience (Chong & Druckman, 2007; Hallahan, 1999; Touri & Koteyko, 2015), RASHI focuses on a goal-directed author. As a result, RASHI goes beyond other computational implementations such as those of Touri and Koteyko (2015)—which uses statistical keywords to determine salient excerpts and then use human experts to explain the frames that they might evoke. Rather than rely on simple keywords determined to be important due to their relative frequency, my approach models the writing process as an author choosing from a set of universal rhetorical mechanisms. In my system, subjectivity is made explicit via subjectivity models, which can be applied consistently across multiple stories.

Cognitive Systems I present a cognitive model of how modulation of communication signals the mental state of a communicator, similar to Langley's (2017) cognitive systems analysis of personality and conversational style. Our approaches are congruent in that they both characterize domain knowledge, design a principled architecture, and attribute communication cues to individual cognitive differences. The main difference in our work is essentially one of domain, or more specifically, what factor is modulating our communication: while Langley focuses personality, my focus is instead on authorial intent.

6. Contributions

In this paper, I have defined a cognitive theory of how humans understand not just what is said, but how it is said. I posited that reasoning about rhetorical intent must be compositional, knowledge-based, and explainable—features that are well suited to a story-understanding substrate. To flesh out the theory, I articulated nine explicit motivations that categorize an author's intentions.

By way of demonstration and elaboration of the theory, I developed RASHI, a computational framework that identifies authorial intent using a subset of these motivations and corresponding mechanisms—modulating doubt, sympathy, and agency. With its story-based substrate, RASHI is able to go beyond mere keyword matching, analyzing the structure of sentences and how sentences relate in a story to form a coherent picture of an author's intentions. RASHI not only collects evidence, but distills it into a coherent argument, identifying mitigating factors and reconciling contradictions. In this way, RASHI represents a step toward a rich, cognitive and computational understanding of human communication. In the process, I made the following contributions:

I proposed a cognitive theory of critical analysis. I claimed that humans conduct their analysis on the sentence level as compared to simply relying on bag-of-words techniques, draw on previous knowledge about authorial intentions and the mechanisms that are available to them, and can explain their reasoning at any step. I argued both that a cognitive model of critical analysis must satisfy these three claims and that story-understanding provides an ideal substrate for doing so.

I identified nine author intentions and corresponding mechanisms. I highlighted a conceptual subdivision between objective and subjective components of rhetorical analysis. Additionally, the formalization of intentions and associated mechanisms lays strong groundwork for future study and expansion of RASHI. By subdividing the problem of identifying author intentions into finding objective, explicit evidence and attributing motive, I separated concerns of collecting data from concerns of subjectively interpreting it, allowing for transparency and explainability. I broke down the problem of what an author means into an analysis of rhetorical choices and their implications.

I implemented Experts that find evidence of rhetorical purpose through a subset of these mechanisms. The Experts can independently analyze a story while allowing for an integrated and then summarized analysis of their conclusions. I demonstrated through the combined result the impact that a Society of Mind (Minsky, 1986, Prologue) approach can have in a specific domain by allowing for a modular while compatible collaboration of subcomponents, each of which can explain the reasoning behind its conclusions. Although RASHI depends on being provided with a subjectivity model, I envision RASHI instead using the meta-text of an article, such as the date and location it was written and basic biographical information of the author, to search for the most plausible subjectivity model and resulting conclusions.

I developed a summarizer that distills evidence into a coherent picture of authorial intent. RASHI analyzes the interplay between the author's rhetorical choices. RASHI determines the amount of weight that should be given to contradictory pieces of evidence by recognizing that the author might use techniques to emphasize or mitigate contradictory factors and by using a hierarchy of rhetorical intensity to decide between effects. The summarizer is just one way in which RASHI explains itself and the result is a natural language, coherent summary of its argument describing authorial intentions.

I demonstrated the critical role of story understanding Story understanding techniques such as concept recognition, knowledge patterns of if-then rules for common sense, and WordNet integration allowed for RASHI to explain its decisions and be flexible across subjectivity models. Story understanding leads to a natural implementation of compiling the composition story of textual evidence and then interpreting the composition story according to a subjectivity model, removing the need to rely on human experts. Comparing a story understanding approach and my architectural design to other techniques demonstrated the crucial role structured text, as compared to bag-of-word approaches, plays in taking systematic bias analysis to the next level.

Overall, through both my theoretical and computational contributions, I demonstrated the power of modeling rhetorical analysis as the process of interpreting an author's writing as goal directed. I showed that by working within a specific domain it is possible to define a concrete, impactful set building blocks and the tools to use them. I demonstrated the value in modeling rhetorical analysis as the process of gathering objective textual evidence and then of analyzing it with respect to an author's subjective frame of mind. Most importantly, I argued that in order to have a rich, human level understanding of communication, we must be able to reason not just about what we say, but how we say it.

7. Acknowledgments

This research was made possible in part by the discretionary fellowship that I received from the Tau Beta Pi Fellowship Program. I would like to thank Professor Patrick Winston and Dylan Holmes for their help, guidance and good advice—without them this work would not have been possible.

References

Chong, D., & Druckman, J. N. (2007). Framing theory. *Annual Review of Political Science*, *10*, 103–126.

Davey, A. C., & Longuet-Higgins, H. C. (1978). A computational model of discourse production. In R. N. Campbell & P. T. Smith (Eds.), *Recent advances in the psychology of language: Formal and experimental approaches*, 125–136. Boston, MA: Springer.

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, *43*, 51–58.

Fellbaum, C. (Ed.). (1998). *WordNet: An electronic lexical database*. Cambridge, MA: MIT Press.

Hallahan, K. (1999). Seven models of framing: Implications for public relations. *Journal of Public Relations Research*, *11*, 205–242.

Hovy, E. (1987). Generating natural language under pragmatic constraints. *Journal of Pragmatics*, *11*, 689–719.

Katz, B. (1997). Annotating the World Wide Web using natural language. *Proceedings of the Fifth International Conference on Computer-Assisted Information Retrieval* (pp. 136–155). Montreal, Canada: Le Centre De Hautes Etudes Internationales D'informatique Documentaire.

Langley, P. (2017). A cognitive systems analysis of personality and conversational style. *Proceedings of the Fifth Annual Conference on Advances in Cognitive Systems*. Troy, New York: Cognitive Systems Foundation.

Liu, H., & Singh, P. (2004). ConceptNet – A practical commonsense reasoning tool-kit. *BT Technology Journal*, *22*, 211–226.

Minsky, M. L. (1986). *The society of mind*. New York: Simon & Schuster.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing* (pp. 79–86). Philadelphia, PA: Association for Computational Linguistics.

Radford, A., Józefowicz, R., & Sutskever, I. (2017). Learning to generate reviews and discovering sentiment. *arXiv preprint*, *abs/1704.01444*.

Sayan, S. (2014). *Audience aware computational discourse generation for instruction and persuasion*. Master's thesis, Electrical Engineering and Computer Science Department, MIT, Cambridge, MA.

Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A. Y., & Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing* (pp. 1631–1642). Seattle, WA: Association for Computational Linguistics.

Taylor, A. (2018). *Computational recognition and comprehension of humor in the context of a general error investigation system*. Master's thesis.

Touri, M., & Koteyko, N. (2015). Using corpus linguistic software in the extraction of news frames: Towards a dynamic process of frame analysis in journalistic texts. *International Journal of Social*

Research Methodology, *18*, 601–616.

Winston, P. H. (2011). The strong story hypothesis and the directed perception hypothesis. *Proceedings of the AAAI Fall Symposium on Advances in Cognitive Systems*.

Winston, P. H. (2014). *The Genesis story understanding and story telling system: A Twenty-First Century step toward artificial intelligence*. Memo 019, Center for Brains, Minds, and Machines, MIT, Cambridge, MA.

Winston, P. H., & Holmes, D. (in press). The Genesis manifesto: Story understanding and human intelligence.