

An Empirical Analysis of Subjectivity and Narrative Levels in Personal Weblog Storytelling Across Cultures

Reid Swanson

*USC Institute for Creative Technologies
Playa Vista, CA, 90094*

RSWANSON@ICT.USC.EDU

Andrew S. Gordon

*USC Institute for Creative Technologies
Playa Vista, CA, 90094*

GORDON@ICT.USC.EDU

Peter Khooshabeh

*The Army Research Lab West
Playa Vista, CA, 90094*

KHOOSHABEH@ICT.USC.EDU

Kenji Sagae

*UC Davis Department of Linguistics
Davis, CA 95616*

SAGAE@UCDAVIS.EDU

Richard Huskey

*The Ohio State University School of Communication
Columbus OH, 43210*

HUSKEY.29@OSU.EDU

Michael Mangus

*UC Santa Barbara Institute for Creative Biotechnologies
Santa Barbara, CA, 93106*

MANGUS@ICB.UCSB.EDU

Ori Amir

*UC Santa Barbara Institute for Creative Biotechnologies
Santa Barbara, CA, 93106*

ORIACADEM@GMAIL.COM

Rene Weber

*UC Santa Barbara Department of Communication
Santa Barbara, CA, 93106*

RENEW@COMM.UCSB.EDU

Editor: Barbara Di Eugenio

Submitted 12/2016; Accepted 11/2017; Published online 11/2017

Abstract

Storytelling is a universal activity, but the way in which discourse structure is used to persuasively convey ideas and emotions may depend on cultural factors. Because first-person accounts of life experiences can have a powerful impact in how a person is perceived, the storyteller may instinctively employ specific strategies to shape the audience's perception. Hypothesizing that some of the differences in storytelling can be captured by the use of narrative levels and subjectivity, we analyzed over one thousand narratives taken from personal weblogs. First, we compared stories from three different cultures written in their native languages: English, Chinese and Farsi. Second, we examined the impact of these two discourse properties on a reader's attitude and behavior toward the narrator. We found surprising similarities and differences in how stories are structured along these two

dimensions across cultures. These discourse properties have a small but significant impact on a reader's behavioral response toward the narrator.

Keywords: discourse analysis, narrative, storytelling, culture, persuasion

1. Introduction

Personal stories, the kind we tell about our daily lives and experiences, are deeply connected to the way we perceive and understand the world (Bruner, 1991; Gerrig, 1993) and are ubiquitous across cultures. They are a rich and complex form of discourse that not only allow us to communicate our experiences, but are also used to evoke emotional responses and change people's beliefs in targeted ways. One of the reasons they are so powerful is in their flexibility to adapt to the communicative and persuasive goals of the narrator. The same underlying set of events can be told as a story many different ways that preserves the basic semantic meaning, but could have substantially different content and make entirely different points. This is often referred to as the Rashomon effect, in reference to the 1950 Akira Kurosawa film where a sexual encounter and death are witnessed by four characters each with a unique, and dramatically different, view of the same fundamental events.

Stories are not merely a sequential retelling of events, but critically provide the *subjective* evaluations why we should care, especially with regard to the plans, goals and motivations of the characters (Labov and Waletzky, 1967). Therefore subjective language, which expresses opinions, emotions, thoughts, preferences, and other mental states of the narrator, is crucial for delivery of the intended interpretation of a personal story. Similar to how a soundtrack can set a specific mood in film to heighten the emotional impact of the sights and sounds of a story, skilled rhetoric can also serve to enhance the impact of events depicted in writing.

The discourse of a story is also characterized by *narrative levels*. A narrative is not only composed of discourse that takes place during the time of events, it also intertwines utterances that speak directly to the audience to provide necessary context as well as the narrated events taking place at the time of the experience. For example, a narrator might step outside the story for moment to provide additional information about the people in them to provide context for understanding their actions, especially if the narrator is aware that the audience has never met them.

The study of narratology has identified many devices that can be used in this way to emphasize aspects of the story the narrator finds important and to highlight a particular point. Our main goal for this research is to examine how different cultures construct personal stories using these devices and, if there are differences across cultures, test if stories that violate cultural norms impact the attitudes of the readers. There is some evidence that people embody at least some of their cultural values, which leads to different envisionments while reading and tell stories. For example, Leung and Cohen (2007) showed that American's typically take the perspective of the main character or 1st person point of view corresponding to a high value of the self in social situations, while Asian cultures more often take 3rd person point of view because of their high value in being seen by others as doing the right thing. These fundamental differences in how people across cultures envision and interpret a story may play a role in the use of narrative devices to construct stories and how a particular narrative is interpreted when reading or hearing one.

In this paper we focus on two of these types of narrative properties: *subjectivity*, which we define as portions of a story based on whether it expresses an unverifiable private mental state following the work of Wiebe et al. (2004), and *narrative levels*, which categorizes content based on whether a clause describes events at the time of the story or something external to the story world (Genette, 1980). We have chosen these properties because they are fundamental concepts central to many theories of narrative and they are simple enough that people can identify them in text.

Because the use of narrative levels and subjectivity is tied closely to how a storyteller shapes an audience’s perception of a set of events (Genette, 1980), analysis of narratives along these two dimensions may provide insight on the differences in narrative structure across cultures and whether these structural properties matter to the reader’s receptiveness to the narrator and her message. Narrative messages are often constructed with a particular cultural group in mind, either one’s own or another. When the message is targeted at another cultural group, for example public service announcements, translated literature and advertising, the subtle differences in the way subjective and narrative levels are used across these cultures could have unintended consequences in the way the target culture interprets the message. Understanding the role of these types of narrative discourse properties could help improve the receptiveness of targeted messages across cultures and guide future automated narrative generation systems in discourse planning for targeted audiences.

To examine these issues, we performed two exploratory experiments to address these questions based on the analysis of over one thousand narratives we collected for this study from three different cultures (Chinese, American, and Iranian). In the first, we found that there are both similarities in the way stories are structured according to these properties as well as significant differences. The most substantial differences occurred in the interactions between clause types rather than considering the distribution of each clause type independently. In the second, we examined the effects of these discourse properties on the attitudes and behavior of the participant toward the narrator for a subset of our cultures. In both cases we found that the subjective distribution of clauses has an impact in the attitudes of the reader toward the narrator. However, we were surprised that the attitudes of the Chinese participants decreased monotonically as the distribution of subjective clauses increased despite our expectation that their attitudes should peak at the cultural norm and decrease as the distribution diverged from this value.

2. Background

2.1 Personal Narratives

In this research our focus is on written forms of personal narrative (text documents) extracted from Internet Weblogs. A personal narrative is broadly defined as the non-fiction stories that people share with each other about their own life experiences. They are primarily about a specific event in the life of the narrator that lasts for minutes or hours (not generalizations over years) and more than a chronological sequence of events. The narrator must also provide some evaluative context about why the events were important enough to share and why the audience should care.

The phenomenal rise of personal weblogs has afforded new opportunities to collect and study electronic texts of personal narratives on a large scale. While blogging is popularly

associated with high-profile celebrities and political commentators, the typical weblog takes the form of a personal journal, read by a small number of friends and family (Munson and Resnick, 2011). As with the adoption of other forms of electronic communication, personal narratives in weblogs take on several new characteristics in adapting to a social media environment that is increasingly public and interconnected. Eisenlauer and Hoffman (Eisenlauer, 2010) argue that the on-going technological development of weblog software has led to an increase of collaborative narration, moving the form further toward Ochs and Capps (Ochs and Capps, 2001) conception of the hypernarrative, where discourse is best understood as a conversation among multiple participants. Langellier and Peterson (Langellier and Peterson, 2004) characterize this collaborative narration as a form of public performance, creating a productive paradox between the insincerity needed to craft a good story and the sincerity of the blogger as a character in the narrated events.

This productive paradox seen in weblog storytelling helps distinguish personal narrative from other narrative forms. Similar to most narratives personal narratives consist in descriptions of events that are causally related, but have the additional requirement that the narrator was a participant in events. The expectation is that the narrator describes events that actually took place, namely it is a non-fiction story, but that some license to stretch the truth may be taken for dramatic or persuasive purposes.

The genre of personal narrative also includes the stories told among family members while reviewing old photographs (Chalfen, 1987), the accounts shared among coworkers in office environments (Coopman and Meidlinger, 1988), the testimonials of people in interviews (Gordon, 2005), and the reflections of daily experiences of people written to private diaries (Steinitz, 1997). Personal stories found on Weblogs share many of the fundamental properties of the other types of discourse in this genre, but are widely available on the web across many different cultures.

2.2 Analysis of Subjective Language

A personal story, almost by definition, is a mixture of describing what actually happened, i.e. the plot, and the narrator's evaluations providing the reason why they are important. A large portion of these evaluations are subjective in nature that help to frame the story in terms of the narrator's opinions, emotions and preferences that can shape the reader's interpretation of events. Exactly defining what it means for an utterance to be subjective is a challenging task in its own right, which is one reason this area has received little attention in the computational linguistics community until fairly recently. However, over the last 10 years, primarily due to the rise of online shopping and review sites, automatic sentiment analysis and opinion mining has become a rapidly growing area of interest. Because of the challenging nature of this task, the problem is often greatly simplified and only a small subset of this overall space has been investigated. In this section we will describe several prominent approaches in the field and the definition we adopt for our work.

To help reduce the complexity, the types of subjective content under consideration are often highly constrained and restricted to very narrow definitions of what subjectivity means for the task. For example, one of the most common formulations of the problem is to frame it as binary classification task to determine if a text is expressing a positive or negative attitude toward a particular topic. More complex formulations that move beyond simple

polarity, for example categorizing texts based on a range emotional states, is still much less common. So far, the bulk of this kind of research has been focused on the automatic analysis of only a few genres of discourse, with the most prominent example being reviews of movies, products, restaurants, etc. (e.g. (Pang and Lee, 2005; Blitzer et al., 2007; Snyder and Barzilay, 2007)). Reviews are an attractive target of sentiment analysis, as they are abundantly available online, they restrict language processing tasks to well-defined domains, and they necessarily express opinions that can often be binned into negative or positive categories relatively easily. Within the context of automatically analyzing reviews, it is common to frame the task as sentiment polarity classification (“thumbs up” vs “thumbs down”), often aided by a preprocessing step that identifies *subjective* language, which Pang and Lee (2008) define simply as opinion-oriented language.

Another language genre where subjectivity and sentiment analysis has been studied extensively is news, where the identification of subjectivity is itself the target of analysis, rather than binary classification of sentiment polarity. In their work on subjectivity analysis, Wiebe et al. (Wiebe et al., 2004) take a broader view of subjective language, which they define as the expression of *private states* (Quirk et al., 1985), which includes emotions, opinions, evaluations and speculations. A third major area of application of sentiment and subjectivity analysis, which has been growing rapidly, is user-generated content, including Twitter, discussion boards, political weblogs, and YouTube video reviews (e.g. (Agarwal et al., 2011; Melville et al., 2009; Morency et al., 2011)).

Although far from exhaustive, the list of language genres mentioned above serves to illustrate how the goals of subjectivity analysis can vary widely when different types of content are considered. For example, in reviews it is more important to determine whether statements are positive or negative, while in news there is a greater focus on separating opinion from fact. Even though goals and even definitions may vary, the most common types of application are related to fulfilling information needs or estimating public interest and opinion regarding specific issues, products, etc. In the case of narrative, however, analysis of subjectivity and sentiment can play a different type of role. Characterizing the mental and emotional state of the narrator is crucial for analyzing the intent of a narrative beyond the facts and events of the story; narratives are often crafted with the explicit goal to have an emotional impact on the reader, sometimes more so than they are to convey a specific sequence of events. In contrast to the main role of subjectivity in reviews or editorial pieces, subjective language in narrative goes far beyond opinions. The expression of emotions, thoughts, preferences and other mental and emotional states is of primary importance.

In our work, we adopt Wiebe et al.’s notion of *subjective language* as the linguistic expression of *private states* (including opinions, evaluations, emotions, and speculations), which are experienced but are not open to external observation or verification by others. Our main focus is on private states of the narrator, since we are dealing with personal narratives, which express the narrator’s point of view. While it can be tempting to define subjective language as the statement of opinions, in contrast to objective statement of facts, this would be an imprecise definition. For example, while the text segment *I know her name* may be considered a statement of fact by the narrator, it is a case of subjective language. The key issue here is not whether a statement is true or made with certainty or privileged knowledge, or even whether it can be considered a fact, and rather whether it

expresses a private state and not something that can be observed or measured objectively and externally. For example, while *I felt sick* is a subjective statement, since it cannot be observed externally, the statement *I had a 102-degree fever* is objective. Similarly, *it was hot yesterday* is subjective (the narrator's opinion), while *it was 95 degrees yesterday* is objective. A statement such as *he was sad* could be subjective because it expresses a private state of a third person, or because it expresses the narrator's opinion or evaluation of a third person. This distinction is not necessary for our current work and introduces several complexities, such as decreasing inter-rater agreement. In either case the statement is considered subjective, which is sufficient for our work. On the other hand, *he said he was sad* is an objective statement, since it describes an event that can be observed externally (namely, the act of saying).

2.3 Narrative Levels

When telling a story, the utterances of the narrator are situated in a particular time and location. While the primary purpose of a personal narrative is to describe events that occurred in the past, every utterance is not necessarily referencing this time frame. For example, in some instances the narrator may step outside the story to speak directly to the audience in the present time and location.

Consider, for example, a narrative that recounts events that include the narrator being afraid of a puppy and disliking dogs as a child, but also expresses the now adult narrator's current embarrassment of this long abandoned fear and current fondness for dogs. The universe of the story, where the narrator is a child, is sometimes referred to as the DIEGETIC (or intradiegetic) level, and the act of narration is performed at the EXTRADIEGETIC level, where in this case the narrator is an adult addressing the reader. This example includes expression of several private states experienced by the narrator: as a character in the story (i.e. at the diegetic level), the narrator experiences fear at a specific moment, and holds a negative preference for dogs; in contrast, the narrator expresses the private states of embarrassment and positive preference for dogs at the time of storytelling (i.e. at the extradiegetic level).

In our discussion we adopt the terms proposed by Genette 1980 to speak of narrative levels in narrative. Although Genette proposes an additional narrative level, META-DIEGETIC, which concerns embedded narratives, this category is rarely used in personal narratives and was collapsed within the diegetic level. In other words, instead of performing a complete analysis of diegetic levels, we make only a binary distinction between the extradiegetic level and all other (intradiegetic) levels, with no distinction made in the levels of embedded narratives. An alternative way to characterize what we refer to as private states at the diegetic and extradiegetic levels is to use the notion of time points due to Reichenbach (1947): the narrator might refer to private states at speech time (at the time of narration), or at the event or reference time. This distinction reflects the narrator's exclusive advantage in framing the story to influence the audience's interpretation and reaction.

The impact of diegetic and extradiegetic material can be understood intuitively by considering the soundtrack in a movie. When watching a movie, we observe events taking place and a story unfolding, which may evoke emotion. External to the universe where the story takes place, however, we may also hear music (e.g. romantic music for a romantic

scene, or fast-paced music for a car chase), which sets a specific mood and serves to evoke or amplify emotional reactions. This music is at the extradiegetic level: it is audible to the audience only, and does not exist for the characters in the story.

3. Experiment 1: Narrative Discourse Structure Across Cultures

It is clear that the content of stories will vary across cultures, reflecting the differences in daily activities and values each culture emphasizes. It is not as obvious if, or in what ways, the narrative levels and subjectivity vary across cultures. For our experiments we examined the narrative levels and subjectivity as they are expressed at the granularity of a clause and will subsequently refer to these two types of discourse features as narrative clauses. Each of these clause types can completely partition the discourse and together are fundamental to the characterization of narrative as a genre. For brevity, we subsequently refer to these clause types as narrative clauses. In our first experiment we investigate this question through an empirical study that compares the distribution of narrative clauses across three different cultures over a large number of personal stories found on the Web.

In our first experiment we investigated the usage of narrative clauses across three different cultures and how the distributions of clauses differed among them. The basic approach proceeded in four stages. First, we trained an automated classifier to identify personal stories from Internet Weblogs in three different languages. Second, we applied our classifier to automatically collect a large sample of personal stories across our three target cultures. Third, we randomly sampled a subset of these documents and manually verified the documents were actually personal stories, which could be used for further annotation and analysis. Fourth, we manually annotated the stories along the diegetic and subjective dimensions described above. Fifth, we analyzed the distribution of clauses across the cultures to identify similarities and differences between them.

3.1 Large Scale Personal Narrative Data Collection

The first step of our process required a large set of personal stories from the web across our three target cultures: American, Chinese and Iranian. There are several criteria that such a collection should possess for our experiments. First, the collection should be large enough to contain a representative sample of the types of activities people in that culture are likely to engage in. Second, the process from which these stories are drawn should be similar across cultures, so that we do not introduce any unnecessary biases. Since there is no existing repository that exclusively consists of personal narratives meeting these criteria, we started by collecting candidate Weblog posts using an automated system to identify personal narratives from arbitrary Weblog posts following the approach of Gordon and Swanson (2009).

In this section we describe the automated approach for identifying our large, but noisy, collection of personal narratives. In Section 4 we will describe our process for sampling a subset of this corpus for narrative clause annotation and, because the automated set is noisy, manually curating the documents to ensure only personal narratives were included.

In Gordon and Swanson’s approach the problem of personal narrative identification is formulated as a standard document classification task where a Weblog post is either labeled STORY if the majority of discourse on the page is a personal story adhering to our

definition in section 2.1 or NOT STORY if it is not. After developing a set of annotation guidelines based on the definition of a personal story, the agreement between annotators on this task for English narratives was 0.68 using Cohen’s κ to adjust for chance agreement. 4,985 unique posts were labeled by 2 annotators with 5.4% of the posts being identified as personal stories. The labeled data was then used to train a statistical classifier. We used the standard classification measures precision, recall and F_1 -score to evaluate the performance of the system resulting in a precision of 0.66, recall of 0.48, and F_1 -score of 0.55. These measures attempt to provide more information than the accuracy of the system, which can be highly misleading when the class labels are unevenly distributed. Precision measures how often the system gets the correct answer based on how many guesses it made and is defined as $\frac{\text{true positives}}{\text{true positives} + \text{false positives}}$. Recall measures how many positive examples the system identified out of all the possible positive examples and is defined as $\frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$. The F_1 -score is the harmonic mean between the two, which has a high value when both precision and recall are high and penalizes them when they are different from each other. Despite the somewhat low performance, our results are near state-of-the art for this type of documentation classification problem and suitable for tasks tolerant to some noise or can incorporate manual filtering like this one.

For the two experiments described in this work we used a new corpus of English stories following the same approach as Gordon and Swanson. This corpus was created by applying the classifier to every English weblog post identified by Spinn3r.com from January 2010 to September 2014, resulting in 25 million unique stories.

Chinese stories were identified by following a similar procedure previously validated by Gordon et al. 2013. Instead of using Spinn3r, we extracted posts from the weblog hosting provider Sina (blog.sina.cn) identified by member lists and a search over the CommonCrawl.org index. This identified 57,496 unique blogs and 995,661 total posts. Topics discussed by users on this hosting service span the full breadth of Chinese society, with increased attention on Chinese popular culture and the entertainment industry. Most posts appearing on Sina’s weblog hosting service are written in Standard Chinese (Mandarin) using the simplified Chinese character set. We chose Sina Blogs primarily because they provide a public directory listing to the blogs of over ten thousand of their users. By crawling this directory, we gathered the URLs for each user’s weblog. Unfortunately, the XML RSS feeds provided by Sina for each of these weblogs contained only a subset (the latest five) of users’ posts. Accordingly, we downloaded the textual content of all publicly accessible posts of each user as HTML files, and extracted the user-authored content of each page automatically. We applied a similar annotation process as described above on 6,999 unique posts with 11.5% being identified as stories. The resulting inter-rater agreement was 0.46 adjusting for chance using Cohen’s κ . We believe the κ is low for several reasons. Below are the three reasons we believe are causing most of the difficulty. First, posts often contain a mix of content, which can include at least one personal story. Deciding exactly how much narrative content is sufficient to warrant labeling the post a story can be difficult. Second, some posts have narrative like qualities, i.e., sequences of events, but have little or no causal or evaluative structure. These can be difficult for annotators especially for short posts that are under 5 or 6 sentences. Third, our definition requires a story to be about a particular event spanning a well specified time frame and is exclusionary of generalizations

Culture	# Labeled	+ Story	κ	Precision	Recall	F ₁	Corpus
American	4,985	270	0.68	0.66	0.48	0.55	25M
Chinese	6,999	805	0.46	0.60	0.46	0.52	64K
Iranian	18,490	462	-	0.49	0.15	0.23	19K

Table 1: Summary of the automated classifier results for each culture. **# Labeled** indicates the total number of posts that were manually annotated. **+ Story** indicates the number of posts that were positively labeled as a personal story. **Corpus** is the total number of stories identified by our automated classifier and used as the corpus to draw from in this experiment.

over many episodes. However, these cases are not always easy to distinguish nor are they always mutually exclusive.

An automated classifier was trained on the labeled data (precision=0.60, recall=0.46, and F₁-score=0.52) and applied to the entire Sina dataset. Where there were discrepancies between automated labels and manually provided labels the manually provided labels were used. A total of 64,231 stories were identified.

We also collected Iranian stories in a similar fashion to the Chinese using 12 popular Iranian weblog hosting services. We applied a similar annotation process on 18,490 unique posts, which indicated 2.5% as personal stories. Inter-rater agreement was assumed to be similar for the Iranian population. These annotations were also used to create an automated story classifier for stories written in Farsi (precision=0.49, recall=0.151, and F₁-score=0.23)¹. Using this classifier, we identified 132,641 unique blogs and extracted 778,745 posts. The automated classifiers used to create our base corpus of English, Mandarin and Farsi stories are available online².

A summary of the annotation and performance statistics for each culture is presented in Table 1.

The differences in narrative content and performance of the classifiers between the three cultures is substantially different and it would be interesting to explore the source of these differences in future work. We suspect that some of the differences in the distribution of stories across languages may not be cultural, but could simply be due to the different types of users each Weblog hosting platform attracts. Our classifiers rely heavily on simple lexical features, which were shown to work well for English where features such as past tense verbs and personal pronouns were highly indicative of narrative content. However, for example, Chinese lacks verb tense, which could make recognizing past tense actions more difficult because multi-word or more complex constructions might be needed to detect them. This might also require more training data to achieve a similar level of performance.

In all three languages the precision is higher than the recall indicating that our classifiers are often correct when it classifies a post as a story, however it fails to identify a large number

1. We are aware of the low recall of the Farsi classifier, however, we have been unable to perform a thorough error analysis of the classifier at this time and we do not currently wish to speculate why the performance is lower than the others. Because of the design of the experiments in part 2 the low performance will not be a concern for this work.

2. <https://github.com/asgordon/StoryNonstory>

of posts that do actually contain a personal story. While the performance of the classifiers, especially for Chinese and Farsi, is less than ideal, it is important to note that this is not a problem for this research because all the stories used in both our experiments go through an additional screening by our human annotators.

4. Text Annotation

After collecting a large sample of candidate personal narratives using our classifiers we randomly selected documents from this corpus to create a smaller curated dataset of personal narratives which we annotated with the narrative level and subjective clause labels. We randomly selected documents from each of these corpora and followed the procedure described later in this section to produce segmented narratives annotated along these two dimensions.

During the course of our annotation process we defined a detailed annotation scheme. The annotation scheme described here is the product of an iterative refinement involving a computational linguist and two annotators. The annotators, whose backgrounds are in Linguistics and Psychology, first acquired familiarity with basic concepts in narratology and computational analysis of narrative by reading the background chapter of *Computational Modeling of Narrative* (Mani, 2012). They then annotated a practice set of about 30 narratives, individually, but in frequent consultation. This process resulted in refinements to the annotation scheme and guidelines for dealing with borderline cases, which were recorded in an annotation manual.

The final set of annotations were produced by applying our process to personal stories drawn from our corpus following a three-step process: *filtering*, *segmentation*, and *narrative level and subjectivity annotation*. Our goal was to annotate 600 English, 300 Mandarin, and 300 Farsi stories for a total of 1,200. We collected fewer stories for Mandarin and Farsi because of the difficulty in finding suitable annotators in those native languages. After the collection effort a number of annotations were filtered for missing data and quality reasons, which left us with an odd number of stories and in the case of Chinese, fewer than 300.

4.1 Filtering

Because the precision of automatic story identification is not perfect, some posts selected automatically are not in fact stories. The first step in our annotation pipeline was to manually discard posts that were not personal stories. The same definition and guidelines for annotating personal stories from our prior work (Gordon and Swanson, 2009) was used to filter out non-stories. Posts that contained multiple personal stories were also discarded by our annotators.

4.2 Segmentation

The next step was to break the text into the segments that will be labeled according to narrative level and subjectivity. Text segments were determined automatically using a set of heuristics described below and consist of one or more clauses. Once the documents were segmented, the annotators labeled each segment for subjectivity and narrative level. The use of clauses as the granularity for annotation was motivated by concerns both princi-

pled and practical in nature. Perhaps the easiest segmentation strategy for identification of subjective passages in narrative is to consider each sentence as a target for labeling. Sentences, however, are clearly too coarse grained, since a single sentence may express an unbounded number of objective and subjective statements through coordination. A more suitable strategy is to define segments in the spirit of the Elementary Discourse Units (EDUs) in Rhetorical Structure Theory (Mann and Thompson, 1988), as applied to entire texts by Marcu (1999). Instead of addressing the challenges of adapting full EDU segmentation to the needs of our task, we opted to use a simplified segmentation scheme inspired by EDUs. We used clauses as the target of annotation with the application of rules and simple heuristics to prevent segmentation of certain types of subordinate clauses that tend not to be relevant to our annotation. Practically, this means that discourse units must be continuous and non overlapping spans of text, which is not a requirement for the annotation guidelines used for Rhetorical Structure Theory segmentation and parsing by Marcu. For example, the non-finite subordinate clauses in *he told her not to go* and *I like going to the movies* are not split into segments separate from their matrix clauses. Other examples of subordinated language that results in multi-clause segments include *going to the movies is what I like to do on weekends* (one segment with four clauses) and *he said he would return* (one segment with two clauses). Our segmentation approach is based on identification of syntactic patterns in parse trees produced automatically by the Stanford parser (Klein and Manning, 2003), and largely follows the EDU segmentation approach described by Tofiloski et al. (Tofiloski et al., 2009), but without the full set of rules and lexical patterns necessary for complete EDU segmentation according to the RST guidelines. The source code implementing our rules for segmentation and automatic classification is available online³.

4.3 Narrative Level and Subjectivity Annotation

The last step was to assign two labels to each segment, one indicating narrative level (extradiegetic vs. diegetic), and the other a binary indicator of subjectivity.

Although grammatical tense is often indicative of narrative level, since events in the story are usually expressed in past tense, this is certainly not always the case (e.g. in *A man walks into a bar*, the event described is in the diegetic level, but described using the present tense).

The notions of *emotion* and *sentiment* are certainly important aspects of narratives that are relevant to our overall goals, but we focused our efforts on the related notion of *subjectivity* as the expression of private states. This simplifies labeling of cases, such as reported speech and reported emotions. For example, in *he said he was sad*, we do not treat *he was sad* as an independent segment, since it is subordinated language. The single segment is labeled as objective, reflecting the *saying* event, even though it involves reporting of a private state. However, in *I knew he was sad*, there is a single segment and it is labeled as subjective, not because of the emotion reported, but because *knowing* is a private state.

3. The automated segmenter based on these heuristics is available at <https://github.com/asgordon/NarrativeAnalysis>

4.4 Annotated Narrative Dataset

We annotated these stories along our two dimensions following a similar set of guidelines developed by Sagae et al. 2013 and Rahimtoroghi et al. 2014, with the definitions of the clause types given below:

1. Subjectivity

- (a) **subjective** clauses express private states, which include emotions, opinions, evaluations and speculations that are not open to external observation or verification by others.
- (b) **objective** clauses express states that can be externally observed and verified by others.

2. Diegesis

- (a) **diegetic** clauses give information about events as they occurred in the world of the story.
- (b) **extradiegetic** clauses give information about the world in which the narrator is addressing the reader.

Upon completion of the process we had 617 English, 261 Mandarin, and 323 Farsi stories annotated along these 2 dimensions. Most of the stories were labeled by a single annotator who matched the language and culture of the target narrative. A subset of the English language stories were annotated by raters from all three cultures to assess the inter-rater reliability treating it as a 4 category labeling task. Raw pairwise agreement over 571 segments (40 stories, with 12 to 18 segments per story) on this four-way labeling task was 84%. We measured chance-corrected agreement over all annotations using Krippendorff's α and obtained a value of 0.73, which is generally considered acceptable agreement for these types of annotation tasks. Because the agreement between annotators across all three cultures was high, our expectation was that agreement on labels on stories in their native languages would also be acceptable.

To produce the final annotations, cases where the annotators disagreed were discussed between themselves and a final label was chosen through their discussion. Table 2 presents a personal narrative annotated following the guidelines presented above⁴.

All annotations discussed in this work were performed manually by our raters, however a tool was developed to segment and automatically label clauses along these two dimensions for these three languages. This tool is publicly available on github⁵.

Overall, these stories are more diegetic than extradiegetic, but the distribution is more equitable than one might expect (58% vs. 42%). They are also highly subjective expressing a private state 68% of the time. Additionally, each clause type has an impact on the distribution of the other. For example, a clause is much more likely to be diegetic given the clause is not subjective $p(\text{Diegetic} = \text{yes} | \text{Subjective} = \text{no}) = 0.77$, whereas the likelihood of a clause to be diegetic given it is subjective $p(\text{Diegetic} = \text{yes} | \text{Subjective} = \text{yes}) = 0.49$. A summary of the clause distributions across all stories is provided in Table 3.

4. Because of issues regarding the expectation of privacy from bloggers (Hayden, 2013) and the nature of the material in our narratives, we do not use examples taken from our corpus.

5. Available at <https://github.com/asgordon/NarrativeAnalysis>

LVL SBJ Narrative Clause		
E	S	I consider myself a very honest person,
E	S	and I've always thought that truth is the best policy.
E	O	I am a 42-year old mother of two,
E	S	and my kids are the most important thing in my life.
D	O	A few years ago my son Tyler asked me if Santa Claus really existed.
D	O	He was four at the time.
E	S	Oh boy!
D	S	I just wasn't ready for that question.
D	S	It was a nice day outside,
D	O	and I told him to go out and play.
D	O	He came back only 15 minutes later
D	O	and asked me again about Santa.
D	O	I pointed to his bike,
D	O	and I asked him who gave it to him.
D	O	He said Santa did.
D	O	I nodded
D	O	and said well then,
D	S	and he gave me a huge smile.
D	S	I felt a little guilty at the time about lying to my child,
E	S	but now I know that parenting is a balancing act.
E	S	Of course the truth is important,
E	S	but nothing trumps a mother's instinct.

Table 2: An example story segmented and annotated using the narrative clause labels. *Narrative levels* (LVL): **D**iegetic, **E**xtradiegetic and *S*ubjectivity (SBJ): **S**ubjective, **O**bjective.

Diegetic	Subjective		Sum	$p(D)$	$p(S = y D)$
	Yes	No			
Yes	18700	13932	32632	0.58	0.83
No	19250	4074	23324	0.42	0.74
Sum	37950	18006	55956		
$p(S)$	0.68	0.32			
$p(D = y S)$	0.49	0.77			Stories = 1,201

Table 3: The frequencies of clauses for each of the possible combinations of types and their overall distributions. The column for $p(D)$ indicates the marginal probability of the diegesis variable. $p(S)$ is similarly displayed for subjectivity. The column $p(D = y|S)$ contains the conditional probability that a clause is diegetic (Yes) given the value of subjectivity either Yes or No corresponding to the value in the appropriate row. The conditional probability $p(S = y|D)$ is also provided. The total number of stories used to create this table is shown in the bottom right.

Language	Diegetic	Subjective		Sum	$p(D)$	$p(S = y D)$
		Yes	No			
Mandarin	Yes	2414	2173	4587	0.50 ^{†‡}	0.53 ^{†‡}
	No	3400	1204	4604	0.50 ^{†‡}	0.74 ^{†‡}
	Sum	5814	3377	9191		
	$p(S)$	0.63 ^{†‡}	0.37 ^{†‡}			
	$p(D = y S)$	0.42 ^{†‡}	0.64 ^{†‡}			Stories = 261
English	Yes	12144	8764	20908	0.58	0.58
	No	12414	2553	14967	0.42	0.83
	Sum	24558	11317	35875		
	$p(S)$	0.68	0.32			
	$p(D = y S)$	0.49	0.77			Stories = 617
Farsi	Yes	4142	2995	7137	0.66 ^{†‡}	0.58 [†]
	No	3436	317	3753	0.34 ^{†‡}	0.92 ^{†‡}
	Sum	7578	3312	10890		
	$p(S)$	0.70 ^{†‡}	0.30 ^{†‡}			
	$p(D = y S)$	0.55 ^{†‡}	0.90 ^{†‡}			Stories = 323

Table 4: A comparison of the narrative clause frequencies and distributions across cultures using a format similar to Table 3. A [†] indicates a significant difference ($p \ll 0.01$) from English and a [‡] indicates a significant difference ($p \ll 0.01$) from Farsi.

4.5 Cross-Cultural Comparison Results

We also performed an analysis of the clause distributions between cultures. We used a binomial test between each pair of cultures to assess significance between the distributions. The distributions were all significantly different ($p \ll 0.01$) for all distributions except $p(S = y|D)$ between English and Farsi. Mandarin stories had the least amount of diegetic material and subjective material, whereas Farsi stories were the most diegetic and subjective. There were also large differences in the conditional distributions. The likelihood that a clause is subjective given that it is extradiegetic was greater than 50% for all languages, however it was near certainty (92%) for Farsi stories and only 74% for Mandarin stories with English stories in between. In other words extra-diegetic material is less likely in Farsi stories, but when there is an extra-diegetic clause it is almost certainly expressing an opinion or private mental state. In contrast, compared to the other languages, Mandarin stories are more likely to contain extra-diegetic material that is objective and English stories split the difference between the two. A summary of the frequencies and distributions over all stories in each corpus is provided in Table 4.

5. Experiment 2: Impact on Attitudes and Behavior

Experiment 1 established that each culture structures its stories differently with regard to the distribution of narrative clause types. In this section we investigate whether these structural differences are embedded strongly enough in a cultural group that they could play a role in influencing the attitudes and behaviors of the reader. A narrator can craft a story for many communicative and persuasive purposes (Schank, 1991) that might shape the reader’s attitude toward the narrator. One of the persuasive goals often found in personal narratives is to reaffirm the narrator’s identity and to project a desired self image to the audience. When a personal story is about a morally charged event, e.g., having an abortion, cheating on a test, or feeling guilty about stealing, the narrator will also often frame the story in a way that tries to convince the audience they acted appropriately.

Based on similarity-liking theory (Byrne, 1997), which posits that people gravitate toward others they perceive as similar, our hypothesis is that stories whose narrative clause distribution matches the reader’s cultural norm will receive the most positive attitudes and decrease as the distribution diverges. Our hypothesis maintains that the structural distribution of a story is not inherent to the genre of narrative, but that members of a cultural group adopt these types of features from the stories of others in their group. This adoption will also lead to a preference for stories that share structural properties that are similar to those of their group. This is similar to the work by (Niederhoffer and Pennebaker, 2002) who first demonstrated that readers can subconsciously pick up on certain stylistic properties of a discourse, such as word choice, and will alter their behavior by entraining their writing style to match the other party.

Here, we investigate whether the narrative discourse properties described above can also impact the attitudes and behaviors of the reader of a story. For example a story written in English, but structured similarly to a Chinese story may not be as well received because it does not “feel” right to an American audience, i.e., is intuitively less similar. Specifically, when narrative properties match cultural expectations we believe a reader will be more likely to:

- perceive that the narrator belongs to the same culture.
- have a positive attitude toward the narrator.
- have a positive attitude toward the story.
- find the narrator more trustworthy.
- agree with the decisions or actions of the narrator.

5.1 Design

To test our hypotheses we designed an experiment to assess how the distribution of narrative clauses impacts the attitudes and behavior of readers as those distributions diverge from the cultural norm of the reader. The basic design of the experiment was to have a participant read a story from our annotated corpus in their native language, answer several follow up questions and play a 1-shot version of the Trust Game (Berg et al., 1995). This was setup as a web-based survey across four pages: *Prequestionnaire*, *Story*, *Survey Questions* and *Trust Game*. Although participants read stories drawn from their own native language and culture, the distribution of clauses varies widely within our dataset and is unknown to the reader. This allows us to test if the aggregate responses of the readers in a culture

differ based on changes in the underlying narrative clause distribution. The Trust Game, described below, is designed to more directly assess a person's attitude toward another than self reported measures because it induces a behavioral response that requires a real valued concession on the part of the participant.

There are many confounding factors that could also influence the responses of the readers, for example, the writing complexity, writing quality, word choice, length of the story and content. We deal with these issues by making certain assumptions about the data. Primarily, we assume that these factors are independent from the narrative discourse style and therefore over a large sample of stories these factors will not substantially affect the aggregate results. While we recognize this assumption may not completely hold for our sample, we believe alternative experimental designs are at least equally problematic and much more difficult to conduct.

For example, in a preliminary study prior this design we also considered an alternative approach based on a controlled experiment. In this preliminary design, a small number of stories were selected and then each was modified by an author to skew the distribution of clauses toward other cultural norms. However, we were immediately faced with several challenges that could not be reconciled. Using a small sample of stories posed the first challenge with this approach. If the stories were not at least minimally engaging to the target audience, the responses to any follow on surveys were likely to be extremely noisy and uninformative. An extensive amount of pretesting would have been required to make this work and it was unclear how consensus would be reached to move beyond this phase. It was also challenging to adjust the distribution of an existing story without significantly altering the confounding factors, for example the content, document length and word choice. Simply adding clauses to increase the relative frequency of one type resulted in stories of different length. It also often adds additional information not contained in the other versions, which would confound the results. By replacing one clause type with another, e.g., extradiegetic to diegetic, it was also possible to significantly alter the content of the story through omission. There were also difficulties in maintaining the same lexical difficulty, reading level and naturalness of the story. For these reasons we found it impractical to perform a small controlled study that we felt overcame these issues satisfactorily.

Modeling subjectivity and narrative levels: Our primary hypothesis in this study is that as the distribution of narrative clauses diverges from a culture's norm, the attitudes and behaviors of readers will be less positive toward the narrator. To model this situation we represent our independent variable (diegesis) as the percentage of clauses that are diegetic and the independent variable (subjectivity) as the percentage of clauses that are subjective. Since we predict that the attitudes of the reader will become less positive as the diegesis and subjectivity diverge from the cultural norm, we model our independent variables **D** and **S** as the distance from the cultural average. For example, we see from Table 4 the cultural norm for diegesis is 58%. So if a particular story in English is 30% diegetic then its distance from the cultural average would be -28%. If our hypothesis is correct we would expect to see a non-linear parabolic relationship where our dependent variables reach a maximum at the cultural norm and lower on either side. We will describe how we handle this expected non-linear relationship in the Results section.

Prequestionnaire: The participant was first asked to complete a prequestionnaire that gathered several pieces of information about them. First, we gathered some basic

demographic information, such as their age, gender, home country, and level of education. To ensure our Chinese participants were not overly Westernized, we also included a subset of the Stephenson acculturation inventory (Stephenson, 2000) that was slightly modified for our American participants to keep the semantics consistent across both cultures⁶.

Story: After completing the prequestionnaire a participant was shown a story drawn from the collection of their native language. They were given as much time as they wanted, but once they progressed they were unable to revisit the story again.

Survey questions: On the next page we asked seven survey questions using a 7-point bipolar scale.

- Q1. I enjoyed reading this story.
- Q2. The narrator did the right thing in this situation.
- Q3. Given the situation described I would have done the same (or similar) thing as the author.
- Q4. This story was written by someone like me or from my background.
- Q5. Given the opportunity I would be friends with the author.
- Q6. I found this story interesting.
- Q7. I have been in a situation exactly like this before.

Q1 and Q6 were asked to gauge the reader's general interest in the story. We used simple intuitive questions for this assessment because we believed that these direct questions would be nearly as reliable as other measures of interest or engagement used for narrative while keeping the survey short. In at least some of our stories, the reader will perceive that the narrator was faced with a difficult decision. Q2 and Q3 were intended to assess whether the narrator has convinced the reader that she chose reasonable actions under the circumstances. In instances in which readers provide strong opinions either for or against the narrator, these questions could enable future work based on morally charged narratives. One of our primary research questions was to determine if the discourse structure that varied substantially from the cultural norm of the reader would have a significant impact on the reader's cultural affinity toward the narrator. We used Q5 and especially Q4 to measure this response. We also believed a reader's response to the Trust Game or one of our other response variables might be moderated by how familiar the reader was with a given situation. For example, a reader intimately familiar with a difficult situation might empathize with the problem the narrator was describing, or alternatively a reader familiar with a boring situation might not pay attention to the details of the story making the judgments less reliable. We asked Q7 for this purpose, although it was not used in the experiments described here.

Trust game: After answering the survey questions we had the participants play a 1-shot version of the Trust Game (Berg et al., 1995), which is designed to measure a participant's trust of an anonymous (or partially anonymous) partner. The game was designed to address a problem with self-reported measures of attitudes related to trust where participants will often report values that do not correspond to their actual behavior, because there are no actual consequences to their self reported decisions. In the game, attitudes are measured

6. We also asked the participants to take a standard need for cognition scale (Cacioppo and Petty, 1984), but did not use the results for this work.

not by what participants say they would do but how they actually behave by requiring them to give up (or believe they are giving up) part of their monetary compensation.

For our version of the game game, we told the participants that we were in contact with the author and they could offer any portion of their participation compensation, from \$0.00 to \$1.50, to them. We would triple the amount specified and offer this new augmented value to the author. The author would then have a chance to reciprocate and give back any portion of this total value to the participant. The participant (in theory) has an opportunity to make more money than originally promised, however, standard economic models assuming rational behavior, based on the Nash equilibrium, predict a participant should give \$0.00. The amount given above \$0.00 is stipulated to correlate with how much the participant trusts the other player. In our experiment we were not actually in contact with the author and no monetary transfer took place, although the participant was unaware of this until after the entire survey was complete.

Experiment: Questions 1-5 of our survey and the monetary value specified in the Trust Game are our dependent variables that address our research questions and hypotheses. Question 1-3 are intended to provide assessments of the reader's attitude toward the message of the story, whereas question 4-5 address the reader's attitude toward the narrator. We also believe that a reader's level of engagement could moderate effects. We see this happening in two possible ways. First, if a reader is highly interested with the story then a discerning reader may be more attuned to the discourse properties of a text. Alternatively, we could imagine a disengaged reader may be less influenced by the content of the story, which they are not engaged with, but discourse properties may still be subconsciously impacting their attitudes. We address this concern with a simple self reported assessment of engagement in question 6.

The monetary value specified in the Trust Game is our final dependent variable that not only assess the reader's attitude toward the narrator, but how willing they are to alter their behavior based on that attitude.

5.2 Participant Recruitment

We recruited participants from two of our cultures (American's and Chinese) and hosted the survey using the crowdsourcing service CrowdFlower⁷. Due to the difficulty of recruiting Iranian nationals living in the United States online they were excluded from this portion of the study. Participants were paid \$1.50 for participating in the experiment, which took about 10-15 minutes to complete. For the Chinese participants, we targeted native Chinese nationals who are currently in the United States, but who are just visiting or have lived here less than 5 years. We randomly sampled 500 stories from our English corpus and 250 from our Mandarin corpus. Our target was to have each story read by 2 participants with each participant reading a single story. Because of the difficulty in recruiting participants, especially Chinese, our final design only included one rating per story as described in the next section.

We were able to recruit a total of 742 American participants and 374 Mandarin speaking participants. To help remove unreliable participants, we filtered participants using a number of criteria.

7. www.crowdfunder.com

We first removed anyone who did not answer all of the required questions in the main part of our survey. This removed 76 English and 13 Mandarin participants.

CrowdFlower is typically used as a crowdsourcing platform for performing microtasks with objective verifiable outcomes. In a prototypical CrowdFlower task (unlike ours) a number of hidden “test” questions are included on the page, which the answer is known to the author but not the worker. These test questions are then used to create a quality profile of workers over all the tasks they perform. This score ranges from 0 (has answered no test questions correctly) to 1 (has answered all test questions correctly). Although we did not have any of these test questions on our survey, we could still rely on the quality score of the participant based on the other tasks they have performed on CrowdFlower. To try to balance our concerns about obtaining highly reliable participants, while not filtering too many qualified subjects, we selected a threshold quality score of 0.8. Applying this filter removed 322 English and 178 Mandarin participants.

The final criteria we used was the length of time the participant had been in the United States as determined by our prequestionnaire. The American participants were required to have been in the US for longer than 5 years, while the Mandarin participants were required to have been in the US less than 5 years. This removed 22 American participants and 96 Mandarin participants.

After filtering we had 311 American stories rated by 322 participants (11 stories annotated by 2 raters) and 85 Chinese stories rated by 89 participants (4 stories annotated by 2 raters). For the stories annotated by more than one participant we randomly selected an annotation from one of the raters, so that our final dataset consisted of 311 American stories and 85 Chinese stories.

5.3 Results

We used the results of our questionnaire to build several linear models to test our hypotheses that stories whose clause distributions more closely matches the cultural norm of the participant will engender more trust toward the narrator. For these models we treated Q6 and the narrative clauses distance from the norm as our independent variables. Our dependent variables, each modeled with a separate linear model, were the results of the Trust Game and Q3-Q5. The results of other questions were not used in these experiments.

Our primary model described here is based on using the results of the Trust Game as our dependent variable. Our model consists of 4 main effects: **culture**, **D**iegetic Distance from Norm, **S**ubjective Distance from Norm, and **Q6**. **Culture** is a categorical variable with two values. **D** and **S** are continuous prediction variables representing the distance from the cultural norm as described above. **Q6** is an ordinal variable with 7 values ranging from Strongly Disagree (-3) to Strongly Agree (+3). Our response variable is the continuous amount of money offered in the **Trust Game**.

To account for the non-linear relationship expected between the narrative clause distances, we apply a second order polynomial transformation on the subjective and diegetic distance variables, which could both approximate a linear relationship or the type of non-linear relationship we expect (i.e., a peak value of trust at the cultural norm and lower values for both high and lower distributions). We also include all two way interactions between the variables and the three-way interaction between **culture** \times **D** \times **S**. Our model

Variable	DF	SS	F	p-value
(Intercept)	1	17.811	67.524	4.35e-15
culture	2	0.035	0.132	0.7169
D ²	2	0.301	0.570	0.5660
S ²	2	1.121	2.125	0.1210
Q6	6	1.748	1.105	0.3592
culture:D ²	2	0.639	1.211	0.2993
culture:S²	2	3.263	6.185	0.0023
D ² :S ²	4	2.202	2.087	0.0821
D ² :Q6	12	3.897	1.231	0.2598
S²:Q6	12	6.535	2.065	0.0187
culture:Q6	6	2.290	1.447	0.1958
culture:D ² :S ²	4	1.455	1.379	0.2407
<hr/>				
r^2	=	0.204		
r^2_{adj}	=	0.081	$p = 0.0041$	

Table 5: The ANOVA results of our model with the degrees of freedom (DF), sum of squares (SS), F-statistic (F) and p-value for each factor in the model. Significant effects are highlighted in bold.

is expressed in a slightly modified version of Wilkinson notation (Wilkinson and Rogers, 1973) in equation (1) below.

$$\text{Trust} \sim \text{culture} * \text{D}^2 * \text{S}^2 + \text{Q6} + \text{culture:Q6} + \text{D}^2:\text{Q6} + \text{S}^2:\text{Q6} \quad (1)$$

A:B indicates the interaction between all factors/levels/categories of the variables A and B . $A * B$ is a shorthand for the linear combination of A and B along with the interactions between them (i.e., $A + B + A:B$). In equation (1) we use the notation D^2 and S^2 to mean that a second order polynomial transformation has been applied to these terms, which is not standard, but helps the readability of the model and the terms in Table 5. The polynomial is added in order to have sufficient power to model the expected non-linear parabolic relationship between our independent and response variables.

We used a type III sum of squares analysis of variance with divergence contrasts for the categorical variables and polynomial contrasts for the ordinal variable. Table 5 summarizes the results of this analysis with the degrees of freedom, sum of squared errors and F statistic used to calculate the p-value for each term in the model. The table shows that none of the main effects of this model are significant (p-value < 0.05), but the interaction between culture and subjective distance (culture:S²) and the subjective distance and the reader's interest (S²:Q6) in the story were. This indicates that the distribution of subjective material has an effect on the level of trust in the narrator based on the culture and that effect of subjective material is moderated by the interest level of the reader. The model as a whole explains a relatively small overall amount of the variance (adjusted $r^2 = 0.081$), but it is significant and non-trivial considering the model does not rely on the content of the stories at all.

We also applied models with the same independent variables to our other dependent variables Q3-Q5. In all cases we did not find any significant main effects or interactions that involved the narrative discourse properties, although all of the models were able to explain more of the variance than for Trust (Q3: adjusted $r^2 > 0.13$, $p < 0.0001$; Q4: adjusted $r^2 > 0.14$, $p \ll 0.0001$; Q5: adjusted $r^2 > 0.23$, $p \ll 0.0001$).

6. Discussion and Conclusion

Stories are a rich method of discourse that both communicate informational content, but also subjective evaluative remarks that provide additional meaning to the audience. Narrators implicitly and explicitly intertwine these types of discourse to compose stories that resonate with their intended audience. Although these clause types could be combined arbitrarily, we see clear preferences for particular distributions and that these distributions vary across the three cultures we investigated.

Not only are the distributions different across cultures, but we also found that the clause distribution has some impact on a reader's attitude and behavior toward the narrator. Our results suggest that it is important to consider the discourse structure when authoring or translating a story for an audience from a specific culture/language. The attitudes expressed towards the narrator by American's primarily fit with what we would expect from similarity-liking theory and previous types of linguistic style matching. For this audience, a reader's attitudes are most positive when the stories narrative clause distribution is near the cultural norm. However, the Chinese results were not what we would expect. Although the results were not significant, the trend for diegetic clauses is the opposite of what we would expect and the relationship for subjective clauses is linear, which was not predicted. We currently do not have a good explanation or hypothesized mechanism that could account for these distributions and a better understanding will require deep knowledge of Chinese cultural factors related to expressions of opinion and trust.

We believe there are a number of additional experiments that could provide additional insight. First, we would be interested in looking at additional culture's, such as Iranian, which were excluded due to difficulties in recruiting. In addition to providing more data, which would help verify the impact of narrative structure between cultures, but also help us determine if Chinese is an exception to our similarity-liking hypothesis or another mechanism is more likely. While our web survey allowed us to increase the number of participants, it had two significant drawbacks. One, it introduced noise in into the responses from unreliable participants, and two it limited the types of behavioral interactions that we could perform (i.e., restricted to a 1-shot version of the Trust Game). Moving to a more controlled environment would limit our ability to recruit participants, but would reduce the noise and allow us to study more complex behavioral responses. While, we believe it is infeasible to control for all the possible textual variables (e.g., length, content, difficulty, etc.), there are some steps that could be taken to control for some of them. For example, we could control for the content by selecting stories that are all based on a similar event, such as attending a wedding or going to a funeral. By controlling for the content we could potentially be more discriminative in isolating the narrative discourse structure. Comparing stories about different topics might also shed light on what kinds of stories are most likely to be affected by discourse structure, for example stories about traumatic events versus mundane activities.

7. Acknowledgments

The projects or efforts depicted were or are sponsored by the U. S. Army. The content or information presented does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred. We would also like to thank Heejung Kim for her comments and suggestions.

References

- Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow, and Rebecca Passonneau. Sentiment analysis of twitter data. In *Proceedings of the Workshop on Languages in Social Media*, pages 30–38, Stroudsburg, PA, 2011. Association for Computational Linguistics.
- Joyce Berg, John Dickhaut, and Kevin McCabe. Trust, Reciprocity, and Social History. *Games and Economic Behavior*, 10(1):122–142, July 1995. ISSN 0899-8256. doi: 10.1006/game.1995.1027. URL <http://www.sciencedirect.com/science/article/pii/S0899825685710275>.
- John Blitzer, Mark Dredze, and Fernando Pereira. Biographies, bollywood, boom boxes and blenders: Domain adaptation for sentiment classification. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, 2007.
- Jerome Bruner. The Narrative Construction of Reality. *Critical Inquiry*, 18(1):1–21, 1991. The University of Chicago Press.
- Dom Byrne. An Overview (and Underview) of Research and Theory within the Attraction Paradigm. *Journal of Social and Personal Relationships*, 14(3):417–431, June 1997. ISSN 0265-4075, 1460-3608. doi: 10.1177/0265407597143008. URL <http://spr.sagepub.com/content/14/3/417>.
- John T. Cacioppo and Richard E. Petty. The need for cognition: relationships to attitudinal processess. In J. H. Harvey, R. P. McGlynn, and J. E. Maddux, editors, *Social Perception in Clinical and Counseling Psychology*. Texas Tech University Press, Lubbock, Tex., January 1984. ISBN 978-0-89672-127-2.
- Richard Chalfen. *Snapshot Version of Life*. Bowling Green State University Popular Press, Bowling Green, OH, 1987.
- Stephanie J. Coopman and Katherine B. Meidlinger. Interpersonal stories told by a catholic parish staff. *American Communication Journal*, 1(3), 1988.
- Volker Eisenlauer. *Narrative Revisited: Telling a story in the new age of media*, chapter Once upon a blog... Storytelling in weblogs, pages 79–108. John Benjamins Publishing Company, 2010.
- Gerard Genette. *Narrative Discourse: An Essay in Method*. Cornell University Press, 1980.
- Richard J. Gerrig. *Experiencing Narrative Worlds: On the Psychological Activities of Reading*. Yale University Press, New Haven, June 1993. ISBN 978-0-300-05434-7.

- Andrew S. Gordon. The fictionalization of lessons learned. *IEEE Multimedia*, 12(4):12–14, 2005.
- Andrew S. Gordon and Reid Swanson. Identifying Personal Stories in Millions of Weblog Entries. In *Third International Conference on Weblogs and Social Media, Data Challenge Workshop*, San Jose, CA, May 2009.
- Andrew S. Gordon, Luwen Huangfu, Kenji Sagae, Wenji Mao, and Wen Chen. Identifying Personal Narratives in Chinese Weblog Posts. In *Intelligent Narrative Technologies (INT-6). Papers from the 2013 AIIDE Workshop*, page 23, 2013. URL <http://www.aaai.org/ocs/index.php/AIIDE/AIIDE13/paper/view/7419/7647>.
- Erika Check Hayden. Guidance issued for us internet research. *Nature News*, 496:411, 2013.
- Dan Klein and Christopher D. Manning. Fast exact inference with a factored model for natural language parsing. In *In Advances in Neural Information Processing Systems 15 (NIPS)*, pages 3–10, Cambridge, MA, 2003. MIT Press.
- William Labov and Joshua Waletzky. Narrative Analysis: Oral Versions of Personal Experience. *Journal of Narrative and Life History*, 7(1-4):3–38, 1967.
- Kristin Langellier and Eric E. Peterson. *Storytelling in Daily Life*. Temple University Press, Philadelphia, PA, 2004.
- Angela K.-y Leung and Dov Cohen. The soft embodiment of culture: camera angles and motion through time and space. *Psychological Science*, 18(9):824–830, September 2007. ISSN 0956-7976. doi: 10.1111/j.1467-9280.2007.01986.x.
- Inderjeet Mani. *Computational Modeling of Narrative*, volume 5 of *Synthesis Lectures on Human Language Technologies*. Morgan & Claypool Publishers, 2012.
- William C. Mann and Sandra A. Thompson. Rhetorical structure theory: Toward a functional theory of text organization. *Text*, 8:243–281, 1988.
- Daniel Marcu. A decision-based approach to rhetorical parsing. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, ACL '99, pages 365–372, Stroudsburg, PA, USA, 1999. Association for Computational Linguistics. ISBN 1-55860-609-3. doi: 10.3115/1034678.1034736.
- Prem Melville, Wojciech Gryc, and Richard D Lawrence. Sentiment analysis of blogs by combining lexical knowledge with text classification. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, 2009. Association for Computing Machinery.
- Louis-Philippe Morency, Rada Mihalcea, and Payal Doshi. Towards multimodal sentiment analysis: Harvesting opinions from the web. In *Proceedings of the 13th International Conference on Multimodal Interfaces*, pages 169–176, New York, NY, 2011. Association for Computing Machinery.

- Sean A. Munson and Paul Resnick. The prevalence of political discourse in non-political blogs. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, Barcelona, Spain, 2011.
- Kate G. Niederhoffer and James W. Pennebaker. Linguistic Style Matching in Social Interaction. *Journal of Language and Social Psychology*, 21(4):337–360, December 2002. ISSN 0261-927X, 1552-6526. doi: 10.1177/026192702237953. URL <http://jls.sagepub.com/content/21/4/337>.
- Elinor Ochs and Lisa Capps. *Living Narrative: Creating Lives in Everyday Storytelling*. Harvard University Press, Cambridge, MA, 2001.
- Bo Pang and Lillian Lee. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, Stroudsburg, PA, 2005. Association for Computational Linguistics.
- Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135, 2008.
- Randolph Quirk, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. *A Comprehensive Grammar of the English Language*. Longman, New York, 1985.
- Elahe Rahimtoroghi, Thomas Corcoran, Reid Swanson, Marilyn A. Walker, Kenji Sagae, and Andrew S. Gordon. Minimal Narrative Annotation Schemes and Their Applications. In *7th Workshop on Intelligent Narrative Technologies*, Milwaukee, WI, June 2014.
- Hans Reichenbach. *Elements of Symbolic Logic*. Macmillan & Co., New York, 1947.
- Kenji Sagae, Andrew S. Gordon, Morteza Dehghani, Mike Metke, Jackie S. Kim, Sarah I. Gimbél, Christine Tipper, Jonas Kaplan, and Mary Helen Immordino-Yang. A Data-Driven Approach for Classification of Subjectivity in Personal Narratives. In Mark A. Finlayson, Bernhard Fisseni, Benedikt Löwe, and Jan Christoph Meister, editors, *2013 Workshop on Computational Models of Narrative*, volume 32 of *OpenAccess Series in Informatics (OASICs)*, pages 198–213, Dagstuhl, Germany, 2013. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. ISBN 978-3-939897-57-6. doi: <http://dx.doi.org/10.4230/OASICs.CMN.2013.198>. URL <http://drops.dagstuhl.de/opus/volltexte/2013/4145>.
- Roger C. Schank. *Tell Me a Story: A New Look at Real and Artificial Memory*. Atheneum, January 1991. ISBN 0-684-19049-4.
- Benjamin Snyder and Regina Barzilay. Multiple aspect ranking using the good grief algorithm in human language technologies. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, pages 300–307, 2007.
- Rebecca Steinitz. Writing diaries, reading diaries: The mechanics of memory. *Communication Review*, 2:43–58, 1997.

- M. Stephenson. Development and validation of the Stephenson Multigroup Acculturation Scale (SMAS). *Psychological Assessment*, 12(1):77–88, March 2000. ISSN 1040-3590.
- Milan Tofiloski, Julian Brooke, and Maite Taboada. A syntactic and lexical-based discourse segmenter. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, ACLShort '09, pages 77–80, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics.
- Janyce M. Wiebe, Theresa Wilson, Rebecca Bruce, Matthew Bell, and Melanie Martin. Learning subjective language. *Computational Linguistics*, 30(3):277–308, 2004.
- G. N. Wilkinson and C. E. Rogers. Symbolic Description of Factorial Models for Analysis of Variance. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 22(3): 392–399, 1973. ISSN 0035-9254. doi: 10.2307/2346786. URL <http://www.jstor.org/stable/2346786>.