Journal of Data Processing

DLINE JOURNALS

Print ISSN: 2278 – 6481
Online ISSN: 2278 – 649X
JDP 2025: 15 (3)
https://doi.org/10.6025/jdp/2025/15/3/124-140

# Computational Modeling of Subjectivity in First-Person Narratives for Identifying Diegetic and Extradiegetic Private States

Kenji Sagae[1], Andrew S. Gordon[1], Morteza Dehghani[1], Mike Metke[1], Jackie S. Kim[1], Sarah I. Gimbel[2], Christine Tipper[2], Jonas Kaplan[2] and Mary Helen Immordino-Yang[2]
[1]Institute for Creative Technologies, University of Southern California
Los Angeles, CA. USA
sagae,gordon,morteza,mmetke,skim@ict.usc.edu

[2]Brain and Creativity Institute
University of Southern California, Los Angeles, CA. USA
sgimbel,tipper,jtkaplan,immordin@usc.edu

**ABSTRACT**

*The paper explores the identification of subjective language in personal narratives, focusing on distinguishing between two narrative levels: diegetic (events within the story) and extradiegetic (the narrator's reflections). Subjective language, expressing emotions, opinions, and mental states, plays a crucial role in shaping the audience's interpretation. The study uses a dataset of 40 annotated personal weblog narratives, employing text classification techniques to automatically identify subjectivity at both levels. A multiclass classification model is trained using features like bag-of-words and part-of-speech tags. Results show a 58% accuracy for six-way classification, outperforming a baseline. Binary classifications for subjectivity and narrative level achieve 78% and 81% accuracy, respectively. Despite limitations due to a small dataset, the findings highlight the feasibility of computational modeling for analyzing narrative subjectivity, with potential applications in sentiment analysis, information retrieval, and commonsense knowledge extraction.*

**Keywords**: Subjectivity Modeling, Narratives, Part of Speech Tags, Weblogs

**Received:** 18 December 2024, Revised 28 February 2025, Accepted 27 March 2025



## 1. Introduction

Beyond communicating a simple description of a sequence of connected events, personal narratives are often crafted to evoke emotions or sway opinions by delivering a story through the point-of-view of one of its participants. Fully understanding a personal narrative therefore requires more than an accurate representation of the semantics of the story; the storyteller's intent is often expressed through subjective statements that may be used to frame specific events in ways that influence their interpretation by the audience. Similarly 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 serve to enhance the impact of events depicted in writing.

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. Despite significant efforts in research on identification of subjective language and detection of sentiment polarity for a handful of language genres, existing approaches fall short of the requirements for modeling subjectivity in personal narratives. Homodiegetic narrative, where the narrator is also a character in the story, presents an interesting challenge: subjective language may refer to mental or emotional states of the narrator as the storyteller, or of the narrator as a participant in the story. One way to characterize this distinction is to place specific instances of subjective language as referring to one of two *narrative levels*: the extradiegetic level, where the narration takes place, or the diegetic level, where the events of the story take place. In addition to its importance in interpreting narrative discourse, where it is important to distinguish emotional states occurring within the story from those that apply to the storyteller during the act of narration, automatic classification of diegetic level and subjectivity of narrative segments can be beneficial in a variety of practical applications involving narrative data. For example, when searching a large corpus of narratives by pecific activities, such as visits to the zoo or protest rallies, it may be desirable to focus on text at the diegetic level by appropriately weighing query terms. Isolating events at the diegetic level would also be desirable in automatic induction of schemas and acquisition of commonsense knowledge from narratives. On the other hand, when the information need targets the intellectual or emotional impact of an experience, without specific constraints on the activity described, focus should be on subjective statements and on passages referring to the extradiegetic level.

We present a data-driven modeling approach for identification of subjective language in each of these two narrative levels, showing how text classification techniques, combined with human annotation, can learn to classify subjectivity in personal narratives. Much of motivation in our work is shared by the line of research on computational approaches to subjectivity in narrative due to Wiebe [38], while our view of subjectivity is that defined by later work by Wiebe and colleagues (e.g., [37]). Unlike Wiebe's original investigation of subjectivity in third person narratives, we deal here with first person narratives, a genre that we describe in more detail in section 2.1. Additionally, while Wiebe's analysis focused on characterization of subjectivity in terms of linguistic elements, our approach focuses instead on the application of machine learning and text classification techniques to the task of identification of subjectivity, following recent research that we discuss in section 2.2. We conclude section 2 with a brief discussion about subjectivity and narrative levels. In section 3 we describe the narrative data used in our investigation, taken from personal Internet weblogs and selected from topics where we expect to find examples of subjective language referring to both the diegetic and extradiegetic levels. In sections 4 we present our computational modeling approach and experiments, in two parts. The first part of our approach involved the design of an annotation scheme for subjectivity and diegetic levels for first person narratives, and manual annotation of 40 narratives (section 4.1). The second part involved learning multiclass classification models from the annotated corpus (section 4.2). We present and discuss our results in section 5, and finally we present our conclusions and briefly discuss future work in section 6.

## 2. Background

### 2.1 Personal Narratives

The genre of the personal narrative is broadly defined as the non-fiction stories that people share with each other about their own life experiences. This genre of discourse includes the stories told among family members while reviewing old photographs [4], the accounts shared among coworkers in office environments [6], the testimonials of people in interviews [10], and the reflections of daily experiences of people written to private diaries [35]. In this research our focus is the written forms of personal narrative (text documents), which are more amiable to automated analysis than other forms. Specifically, we develop and evaluate our methods on personal narratives extracted from Internet weblogs.

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 [25]. 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 [7] 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 [26] conception of the hypernarrative, where discourse is best understood as a conversation among multiple participants. Langellier and Peterson [18] 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. As in all narrative, personal narrative consists of descriptions of multiple events that are causally related, but requires further that the narrative perspective is the author’s own. The expectation of the reader is that the narration reflects the truthful interpretation of events actually experienced by the author, but the truth of the narration is constrained by the demands of good storytelling.

### 2.2 Analysis of Subjective Language

Because personal narratives feature a storyteller who is also a character in the story, it is common for the events of the story to be framed by the narrator in terms of opinions, emotions, preferences, and other commentary that influences the reader’s interpretation of the events. While it is possible for a narrator to be objective in recounting first-hand participation in a story, our analysis is focused on personal narratives that are framed by subjective language employed by the narrator, and more precisely on computational models for identification of subjectivity in personal narratives.

Although there has been remarkable interest in analysis of sentiment and subjectivity in text in the past decade, the bulk of the research has been focused on a few language genres, with the most prominent example being reviews of movies, products, restaurants, etc. (e.g., [27, 3, 33]). Reviews are attractive as the 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. In analysis of 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 [28] 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. [37] take a broader view of subjective language, which they define as the expression of *private states* [29], 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., [1, 23, 24]).

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. Correctly assessing the mental and emotional state of the narrator is crucial for understanding 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, speculations and other mental processes), 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 mea sured 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. It is not important at this point to distinguish whether a statement such as *he was sad* is 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, since in either case the statement is subjective. 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 Private States and Narrative Levels

The genre of the personal narrative is particularly interesting from the point-of-view of analysis of subjectivity in that the narrator experiences emotions and holds opinions both within the story, as a character along with other story participants, and also outside of story. Accordingly, the narrator may employ subjective language that applies to at least two different narrative levels. 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).

Although in our discussion we adopt the terms proposed by Genette [9] to speak of diegetic levels in narrative, we do so only to determine whether private states are either internal or external to the universe of the story, leaving aside the more complex issues of matching private states to more levels in embedded narratives. 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 [30]: the narrator might refer to private states at speech time (at the time of narration), or at the event or reference time. However, our main concern is not necessarily one of time; the distinction we make in the present work is between private states experienced by the narrator as a character in the story, and private states experienced by the narrator as the storyteller. 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. Data

In developing and evaluating a data-driven approach to our classification task, we required a corpus of personal stories containing substantial amounts of subjective statements describing private states belonging to either the diegetic or extradiegetic level, meaning that the narrator experiences the private state either within the universe of the story, or outside, at the time of the act of narration, respectively. Although weblog content is abundant and readily available, selecting and annotating random weblog posts would be inefficient. Gordon and Swanson [13] estimated that only 4.8% of randomly sampled non-spam English-language weblog posts can be characterized as personal stories, defined by them as non-fiction narrative discourse describing a specific series of events in the past, spanning minutes, hours, or days, where the storyteller or close associate is a participant. Even within this small subset of posts, our expectation is that the balance between description of private states held at the diegetic and extradiegetic levels will vary widely. For example, the play-by-play narrative of a baseball game might focus entirely on the diegetic universe, with descriptions of excitement, happiness or apprehension applying only to the diegetic level, while an account of cherished childhood memories might move the narrator to describe an emotional state triggered by the events in the diegetic universe but experienced during narration, at the extradiegetic level.

To curate a well-balanced corpus for analysis, we focused our efforts on finding weblog posts about situations and activities that would lend themselves to a mix of these two types of expression of subjectivity, or private states. We specifically targeted narratives of socially-questionable behavior, e.g., stories of stealing, quitting a team, giving a child up for adoption, or getting into a physical fight. We expected that bloggers who shared personal narratives about socially-questionable behavior would feel the need to be descriptive of the events that occurred, including opinions, thoughts and emotions held at the time, and to provide some rationale or justification for their behavior, leading to expression of their current feelings about the past events of the narrative. Collectively, we brainstormed a list of such situations that could potentially be found in public weblogs (Figure 1).

To conduct these situation-specific searches, we used the technologies and methodologies described by Gordon et al. [14], which were used by them to find hundreds of personal narratives in weblogs related to health emergencies. The approach begins with the automatic filtering of personal narratives from streams of weblog posts, applying supervised story classifier to three years of non-spam English-language weblog posts provided by a weblog aggregator (Spinn3r.com). The filtered story collection (over 20 million posts) was then indexed using a text retrieval engine (Apache Lucene), which could be queried with a large array of weighted terms. Initial queries were authored for each of the socially-questionable behavior following Gordon and Swanson [12], where paragraph-sized fictional prototypes were used to retrieve similar instances. Retrieved posts were then annotated as to their relevance to the query, and this feedback was used to further refine the query and weight query terms using the Rocchio relevance feedback algorithm [32]. We identified 460 posts containing narratives of socially-questionable behavior using this approach, from which we selected 40 posts that we judged to be most compelling as personal narratives. These 40 posts include 22 of the 26 topics in Figure 1, with no topic appearing in more than three stories. Topics appearing in multiple posts include lying, divorce, protest rallies, breaking the law, and quitting a team, abortion, disobeying a superior, murdering someone, getting into a physical fight, killing an animal, prostitution and physically punishing a child. The four topics that are not represented in our selection of posts are: stealing, taking an unfair advantage, putting your own interests above others, and neglecting to care for children.

## 4. Approach

Following previous data-driven efforts on subjectivity and sentiment analysis, exemplified by the work of Wiebe et al. [37] and Pang and Lee [27], we use a machine learning approach typically associated with text

- Participating in a protest rally
- Quitting a job
- Telling a lie
- Getting into a physical fight
- Converting from one religion to another
- Having an abortion
- Stealing something
- Crossing a picket line
- Changing one's political party

- Changing the country of your citizenship
- Making a large personal sacrifice
- Cheating in a romantic relationship
- Getting a divorce
- Quitting a team
- Cheating on a test
- Killing an animal
- Prostitution
- Taking an unfair advantage
- Physically punishing a child
- Violating a religious practice
- Breaking the law
- Murdering someone
- Disobeying a superior
- Putting a child up for adoption
- Putting your own interest above others
- Neglecting to care for children

Figure 1. List of topics involving socially questionable behavior for personal narrative selection

classification. While Pang and Lee leverage "found-data" to train a classifier for subjectivity detection in reviews without the need for manual annotation, our approach has in common with Wiebe et al.'s that it involves the definition of an annotation scheme, training of human annotators, and manual annotation of training data for subjectivity classification. As an initial attempt to address subjectivity modeling in personal narratives, we use a simple discriminative text classification approach, with a relatively small training dataset consisting of 40 narratives. These narratives were found originally in personal weblogs, and for reasons unrelated to the work described here were edited for length prior to annotation[1]. In most cases, editing consisted largely of removing sentences from the original weblog post, with occasional addition of a few words to restore coherence to the edited text. The edited narratives retained the vocabulary and much of the narrative structure of the original posts, and contained 160 to 185 tokens (words and punctuation) each. The average length for these narratives is 169 tokens.

### 4.1 Text Annotation

After selection of narratives from personal weblogs, the first step towards creation of the dataset necessary for training a data-driven model and subsequent empirical validation of our overall approach was the definition of an annotation scheme. The annotation scheme described here and used in our experiments is the product of

[1]These narratives were used as stimuli in a separate series of experiments that examine the emotional reactions of readers of personal narratives. These experiments required that human subjects read each narrative in 36 seconds, and for this reason each of the 40 narratives was shortened from its original weblog post version. In future work we plan to investigate the relationship between subjectivity in narrative, as annotated in the work describe here, and emotional impact on readers.

iterative refinement involving a computational linguist and two annotators. The annotators, whose back grounds are in Linguistics and Psychology, first acquired familiarity with basic concepts in narratology and computational analysis of narrative by reading the background chapter of Indejeet Mani's book *Computational Modeling of Narrative* [19]. 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, and was followed by annotation of the 40 narratives in our dataset by each of the two annotators individually. The annotation task consists of tagging segments in the narrative with one of six labels, described below. Text segments were determined automatically and consist of one or more clauses. The use of clauses as the granularity for annotation was motivated by concerns both principled 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 [20], as applied to entire texts by Marcu [22]. 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, taking 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. 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 [16], and largely follows the EDU segmentation approach described by Tofiloski et al. [36], but without the full set of rules and lexical patterns necessary for complete EDU segmentation according to the RST guidelines. Because our segments are sometimes too fine-grained, and because narratives sometimes include passages that do not fall within one of the categories defined by our scheme, the annotation scheme provides the option of tagging specific clauses as Other/None (see below).

The tags in our annotation scheme are:

Story Event Denotes a clause that corresponds to an event in the story.

Story Private State Denotes a clause that corresponds to an expression of a private state of the narrator that applies within the diegetic level.

Story Other Denotes other material that applies mainly to the story, such as descriptions, direct quotes, etc.
Subjective Statement Denotes an expression of a private state at the time of narration (at the extradiegetic level), rather than within the story.

Objective Statement Denotes an expression of fact that applies at the time of narration (at the extradiegetic level), rather than within the story.

Other/None Used for tagging clauses that do not fall within one of the categories above.

The annotation scheme is intended to distinguish discourse segments along two dimensions:

(1) subjective vs. objective language; and (2) language that refers to the diegetic vs. the extradiegetic level. Although the notions of emotion and sentiment are certainly important aspects of narratives that are relevant to our overall goals, 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 a Story Event, 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 a Story Private State, not because of the emotion reported, but because knowing is a private state.

The manual annotation process was done with the aid of the Story Workbench [8], a flexible environment for linguistic annotation, customized specifically for our annotation task.

Figure 2 shows a narrative being annotated using the Story Workbench. The top middle section shows the text of the narrative, and the top right shows the segments to be annotated.

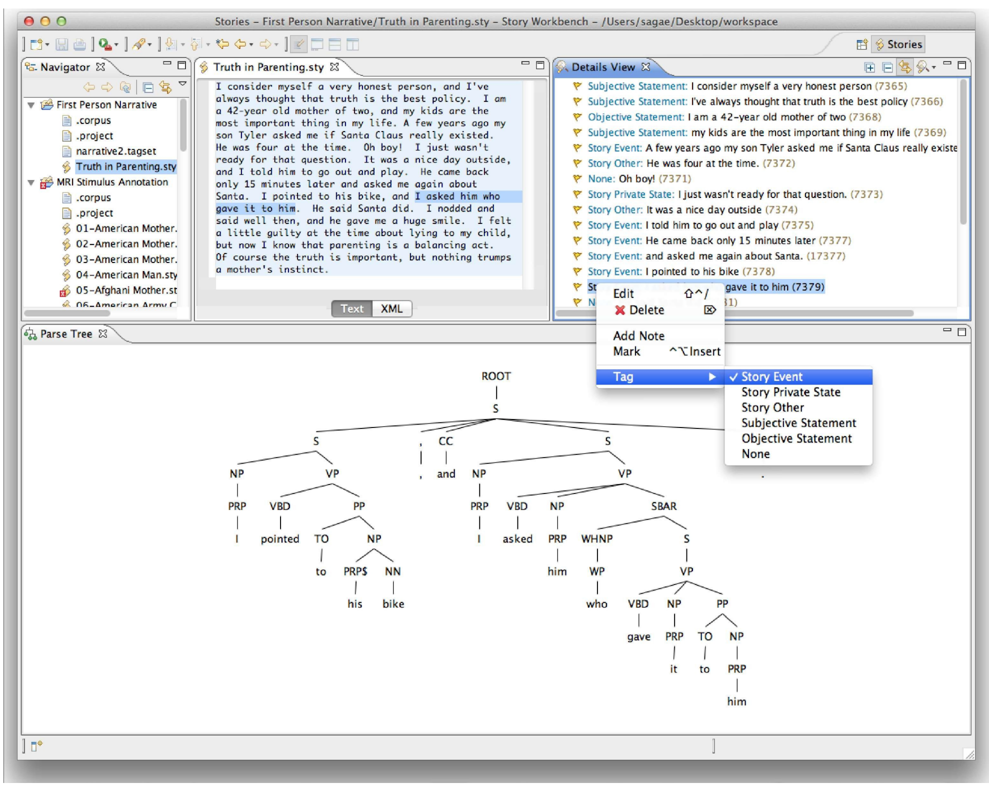


Figure 2. The Story Workbench tool for linguistic annotation [8] customized for annotation of private states in their narrative levels. The top three sections of the interface show a list of narratives to annotate, the text of a narrative, and a list of segments to be labeled. The bottom area shows an automatic syntactic analysis of the sentence being annotated

Segmentation is performed automatically using clause boundaries determined by the Stanford parser [16] integrated in the Story Workbench. From the point-of-view of the annotators, this segmentation and population of the top right area with segments happen automatically and nearly instantly once text is loaded or entered in the text area. Annotators simply go down the list of segments, choosing one of the six labels for each of the segments. Raw pairwise agreement over 571 segments (40 stories, with 12 to 18 segments per story) on the six-way labeling task was 84%. We measured chance-corrected agreement using Krippendorf's and obtained a value of 73%. To produce the final annotations, cases where the annotators disagreed were discussed and a final label was chosen. The most frequent label is the first in the list above, Story Event, which accounts for 34% of the segments. Story Private State, the second most frequent label, accounts for 29% of the segments. Table 1 shows several characteristics of the final annotated narrative corpus used in our experiment.

Appendix A shows an example of how a narrative is annotated according to our annotation scheme[2].

## 4.2 Classification of Subjective Language and Narrative Level

We now turn to automatic classification of narrative segments according to our annotation scheme. As in the manual annotation process, segmentation is performed as a pre-processing step using the clause boundaries produced by the Stanford parser. The next step is then

| Corpus attribute | Value |
|---|---|
| Number of narratives | 40 |
| Average number of tokens per narrative | 169 |
| Average number of segments per narrative | 14 |
| Frequency of Story Event label | 194 (34%) |
| Frequency of Story Private State label | 166 (29%) |
| Frequency of Story Other label | 57 (10%) |
| Frequency of Subjective Statement label | 85 (15%) |
| Frequency of Objective Statement label | 63 (11%) |
| Frequency of None label | 5 (1%) |

Table 1. Characteristics of the annotated corpus of narratives

[2]Because of issues regarding the expectation of privacy from bloggers [15] and the nature of the material in our narratives, we do not use examples taken from our corpus.

to tag each resulting segment with one of the six categories in the annotation scheme. We first approached this step as a straightforward text classification task at the segment level, treating each segment as independent. We use multiclass classification with Maximum Entropy models [2][3], and for each segment we extract features of the following types:

Bag of words (unigram features, or $w_i$);

Part-of-speech tags for each word in the segment, as assigned by the Stanford parser ($t_i$);

Word bigrams ($w_i, w_{i+1}$);

Part-of-speech tag bigrams ($t_i, t_{i+1}$);

Part-of-speech tag trigrams ($t_i, t_{i+1}, t_{i+2}$);

Word/part-of-speech tag bigrams ($w_i, t_{i+1}$ ; $t_i, w_{i+1}$);

Word/part-of-speech tag trigrams ($w_i, t_{i+1}, t_{i+2}$ ; $t_i, w_{i+1}, t_{i+2}$ ; $t_i, t_{i+1}, w_{i+2}$);

These features are intended to capture the bag-of-words representation widely used in text classification, augmented with n-grams to provide some context, and backing off to part-of-speech tags to reduce the sparsity of n-grams. Since our annotation scheme includes two separate dimensions of the narrative segments, the classification task could be performed in two steps (subjectivity detection, and narrative level classification), or a single step of six-way classification. In both cases the same set of feature types is used. Missing entirely from our classification approach is any notion that each segment is in fact not independent from its context. A possible extension to our current model is to add dynamic features for neighboring segments, making the overall model a conditional random field [17] that optimizes the entire set of segment labels for the entire narrative jointly. This is left as future work.

## 5. Results and Discussion

Because single-step six-way classification and the two-step classification approaches discussed in the previous section produced very similar results, we focus our discussion on the simpler approach of single-step classification. To perform an empirical evaluation of this approach, we performed a "leave one narrative out" cross-validation using our annotated set of 40 narratives: to label each of the 40 narratives, we used the remaining 39 to train a classification model. The overall accuracy using the six labels was 58%, a substantial improvement over a simple majority baseline (34%). While far below the level of human annotation in this task, our results are encouraging given our simple text classification approach. A more informative evaluation is to consider the precision and recall of each category individually. Precision for a category c is defined as the number of correct assignments of the c label divided by the total number of times the classifier assigned the c label to a segment.

[3] Our classifier was implemented with Yoshimasa Tsuruoka's Maximum Entropy library available at http://www.logos.t.u-tokyo.ac.jp/~tsuruoka/maxent/

Recall for a category c is defined as the number of correct assignments of the c label, divided by the total number of instances of the c label in our answer key, or manual annotation. Intuitively, precision corresponds to how often the classifier is correct when it assigns a certain label, and recall corresponds to what portion of the items with a certain label the classifier can find. By labeling every segment as c, we would obtain perfect recall, but poor precision. Conversely, by assigning the c label very conservatively and only in cases of very high confidence, we could obtain high precision, at the cost of low recall. Table 2 show the precision and recall values for each of the categories in our annotation scheme.

The imbalance of high recall and lower precision for Story Event reflects that our classifier tends to prefer the assignment of the Story Event label over other labels. In particular, a substantial number of segments that should be labeled Story Private State or Objective Statement are labeled by the classifier as Story Event. In one case, the error appears to be in the more general dimension of subjectivity, and in the other, the error is related to distinguishing between narrative levels. This is also reflected in higher precision than recall in identifying the Story Private State and Objective Statement categories. The confusion between Story Event and Story Private State reflects that, even though the model often correctly identifies that the segment is referring to the diegetic level (which is likely due to part-of-speech features that reflect verb tense), it is less accurate in distinguishing between events and private states. In those cases, the error is in subjectivity classification. The confusion between Story Event and Objective Statement, conversely, shows that the classifier sometimes distinguishes subjective segments correctly, but fails to assign them the correct diegetic level. Not surprisingly the Story Private State category is also often confused with the Subjective Statement category (segments corresponding to expressions of private states of the narrator as the storyteller, outside of the story). This highlights the challenge of classifying correctly along both dimensions in our annotation scheme, which is necessary for analysis of subjectivity specifically at the extradiegetic and diegetic levels. Our results for identification of subjective statements that apply to the extradiegetic level are more balanced: we correctly identify almost half of the narrator's expressions of private states, with a relatively low rate of false alarm at about 50%. This is a particularly important category, since it corresponds to the narrator's reflections about the events in the story.

| **Classification task** | **Majority baseline accuracy** | **Accuracy** |
|---|---|---|
| Six-way classification | 34% | 58% |
| Binary subjectivity classification | 56% | 78% |
| Binary diegetic level classification | 69% | 81% |

Table 3. Accuracy results for our main classification task using our six-category scheme (Section 4.1), and for two binary classification tasks, each focusing only on subjectivity or diegetic level. Accuracy of a majority baseline classifier is also shown for comparison

When we consider each dimension separately, we observe substantially higher accuracy, corresponding to easier classification tasks. On the binary task of identifying segments that refer to the diegetic level vs. to the extradiegetic level, which we evaluate simply by grouping the the first three labels of our scheme in to one category (diegetic), and the remaining three labels into another category (extradiegetic), we obtain 81% accuracy. For comparison, a majority baseline for this task would assign the diegetic label to all segments and obtain 69% accuracy, since segments that refer to the extradiegetic level are substantially outnumbered by segments referring to the diegetic level. In the binary subjectivity classification task (grouping the Story Private State and Subjective Statement categories into one subjective category), where segments are simply classified as subjective or not, as is common in natural language processing, our approach does well, with 78% accuracy, compared to 56% for a majority baseline. These results, summarized in Table 3, highlight that the combined task of finding subjective language within the appropriate narrative level is predictably more challenging than either subjectivity classification or narrative level classification in isolation.

## 6. Conclusion and Future Work

We have described a methodology for analysis of subjective language in narrative that involves manual annotation to produce training material that can be used to learn computational models for automatic identification of subjectivity at the diegetic and extradiegetic levels. Although our classification accuracy still needs improvement, it shows promise given the small number of narratives in our training data, and it highlights some of the challenges in this type of classification. Our next step is to annotate a larger set of personal narratives to generate a larger training set and separate development and test sets using unedited text from weblogs. We believe a larger training set will improve the accuracy of our simple classification framework, and that further accuracy improvements may be obtained by going beyond our current framework where each segment is classified independently. In future work, we plan to abandon the assumption that segments are independent, and apply a structured classification approach (e.g., conditional random fields [17]). Additional annotated data will be important for exploring the use of features beyond unigrams and part-of-speech tags (e.g., features extracted from syntactic trees) using development data. In addition, although our current set of 40 narratives similar in length allows us to see how well our classification approach performs across a variety of topics, we plan to confirm that our models generalize to personal narratives from weblogs in their original forms.

With our current text classification model, subjectivity classification accuracy (78%) is at a level where automatic identification of subjective language in personal narratives could be of practical use. For example, Riloff et al. [31] have shown that subjectivity classification at this level of accuracy is useful for improving the precision of information extraction systems. Similarly, our approach to the classification of the aggregated diegetic and extradiegetic categories performs well enough (81% accuracy) for potential use in a range of other natural language processing technologies. In many cases it would be desirable to filter out passages that refer to the diegetic or extradiegetic level in order to improve performance or precision. For example, information retrieval system that support searches for narratives of specific activities, such as protest rallies or automobile crashes, may garner improvements by indexing only the diegetic material in the document collection. Where relevance feedback is used to refine queries [14], diegetic material could be weighted more heavily when selecting and weighting query terms. Gains should also be expected in language processing systems that aim to generalize over events described in narratives, as in schema induction [5] and commonsense knowledge extraction [21, 11]. Similarly, some systems may benefit by ignoring the events of narratives, particularly where the emotional or intellectual impact of an experience defines the retrieval criteria [34].

## Acknowledgements

We are grateful to Mark Finlayson for providing us with the Story Workbench and extending it with customization options specific to our annotation task, including automatic segmentation. We also thank David Traum and Louis-Philippe Morency for insightful discussions about time points in discourse and sentiment analysis of usergenerated content, respectively. Finally we thank the anonymous reviewers for insightful comments and suggestions that helped us strengthen and clarify the paper.

## References

[1] Agarwal, Apoorv., Xie, Boyi., Vovsha, Ilia., Rambow, Owen., Passonneau, Rebecca. (2011). Sentiment analysis of Twitter data. *In:* Proceedings of the Workshop on Languages in Social Media (pp. 30–38). Stroudsburg, PA: Association for Computational Linguistics.

[2] Berger, Adam L., Della Pietra., Vincent J, Della Pietra., Stephen A. (1996). A maximum entropy approach to natural language processing. *Computational Linguistics,* 22(1), 39–71.

[3] Blitzer, John., Dredze, Mark., Pereira, Fernando. (2007). Biographies, Bollywood, boom boxes and blenders: Domain adaptation for sentiment classification. *In:* Carroll, John A., van den Bosch, Antal., Zaenen, Annie. (Eds.), Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (pp. 440–447). Prague, Czech Republic.

[4] Chalfen, Richard. (1987). *Snapshot version of life*. Bowling Green, OH: Bowling Green State University Popular Press.

[5] Chambers, Nathanael., Jurafsky, Dan. (2009). Unsupervised learning of narrative schemas and their participants. *In:* Su, Keh-Yih., Su, Jian., Wiebe, Janyce. (Eds.), *Proceedings of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the AFNLP* (pp. 297–305). Singapore: Association for Computational Linguistics.

[6] Coopman, Stephanie J., Meidlinger, Katherine B. (1988). Interpersonal stories told by a Catholic parish staff. *American Communication Journal,* 1(3).

[7] Eisenlauer, Volker. (2010). Once upon a blog... Storytelling in weblogs. *In:* Hoffmann, Christian R. (Ed.), *Narrative Revisited: Telling a story in the new age of media* (pp. 79–108). Amsterdam: John Benjamins Publishing Company.

[8] Finlayson, Mark Alan. (2011). The Story Workbench: An extensible semi-automatic text annotation tool. *In:* Intelligent Narrative Technologies IV, Papers from the 2011 AIIDE Workshop, Stanford, California, USA, October 10–11, 2011 (pp. 21–24). Stanford, CA: AAAI Press.

[9] Genette, Gerard. (1980). *Narrative discourse: An essay in method*. Ithaca, NY: Cornell University Press.

[10] Gordon, Andrew S. (2005). The fictionalization of lessons learned. *IEEE Multimedia,* 12(4), 12–14.

[11] Gordon, Andrew S., Bejan, Cosmin A., Sagae, Kenji. (2011). Commonsense causal reasoning using millions of personal stories. *In:* Burgard, Wolfram, & Roth, Dan (Eds.), *Proceedings of the Twenty-Fifth Conference*

*on Artificial Intelligence*. San Francisco, CA: Association for the Advancement of Artificial Intelligence.

[12] Gordon, Andrew S., Swanson, Reid. (2008). Storyupgrade: Finding stories in internet weblogs. *In:* Adar, Eytan, Hurst, Matthew, Finin, Tim, Glance, Natalie S., Nicolov, Nicolas., Tseng, Belle L. (Eds.), Proceedings of the Second International Conference on Weblogs and Social Media (ICWSM 2008) (pp. 188–189). Seattle, WA: Association for the Advancement of Artificial Intelligence.

[13] Gordon, Andrew S., Swanson, Reid. (2009). Identifying personal stories in millions of weblog entries. *In:* Soboroff, Ian, Java, Akshay (Eds.), Data Challenge Workshop, Papers from the 2009 ICWSM Workshop (AAAI Technical Report WS-09-01) (pp. 16–23). Association for the Advancement of Artificial Intelligence.

[14] Gordon, Andrew S., Wienberg, Christopher., Sood, Sara Owsley. (2012). Different strokes of different folks: Searching for health narratives in weblogs. *In:* 2012 International Conference on Privacy, Security, Risk and Trust (PASSAT) and 2012 International Conference on Social Computing (SocialCom) (pp. 490–495). Amsterdam, Netherlands: *IEEE Computer Society*.

[15] Hayden, Erika Check. (2013). Guidance issued for US internet research. *Nature News, 496*, 411.

[16] Klein, Dan., Manning, Christopher D. (2003). Fast exact inference with a factored model for natural language parsing. *In:* Becker, Suzanna, Thrun, Sebastian, Obermayer, Klaus (Eds.), *Advances in Neural Information Processing Systems 15* (pp. 3–10). Cambridge, MA: MIT Press.

[17] Lafferty, John D., McCallum, Andrew., Pereira, Fernando C. N. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. *In:* Brodley, Carla E., Danyluk, Andrea Pohoreckyj (Eds.), Proceedings of the Eighteenth International Conference on Machine Learning (pp. 282–289). San Francisco, CA: Morgan Kaufmann Publishers Inc.

[18] Langellier, Kristin., Peterson, Eric E. (2004). *Storytelling in daily life*. Philadelphia, PA: Temple University Press.

[19] Mani, Inderjeet. (2012). *Computational modeling of narrative* (Synthesis Lectures on Human Language Technologies, No. 5). Morgan Claypool Publishers.

[20] Mann, William C., Thompson, Sandra A. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text,* 8, 243–281.

[21] Manshadi, Mehdi., Swanson, Reid., Gordon, Andrew S. (2008). Learning a probabilistic model of event sequences from internet weblog stories. In Wilson, David, Lane, H. Chad (Eds.), Proceedings of the Twenty-First International Florida Artificial Intelligence Research Society Conference (pp. 159–164). Coconut Grove, FL.

[22] Marcu, Daniel. (1999). A decision-based approach to rhetorical parsing. In Dale, Robert, & Church, Kenneth Ward (Eds.), Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics (pp. 365–372). College Park, MD: Association for Computational Linguistics.

[23] Melville, Prem., Gryc, Wojciech., Lawrence, Richard D. (2009). Sentiment analysis of blogs by combining lexical knowledge with text classification. *In:* Elder IV, John F., Fogelman-Soulié, Françoise, Flach, Peter A., Zaki, Mohammed Javeed (Eds.), Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1275–1284). New York, NY: Association for Computing Machinery.

[24] Morency, Louis-Philippe., Mihalcea, Rada., Doshi, Payal. (2011). Towards multimodal sentiment analysis: Harvesting opinions from the web. *In:* Bourlard, Hervé, Huang, Thomas S., Vidal, Enrique., Gatica-Perez, Daniel., Morency, Louis-Philippe (Eds.), *In: Proceedings of the 13th International Conference on Multimodal Interfaces* (pp. 169–176). New York, NY: Association for Computing Machinery.

[25] Munson, Sean A., Resnick, Paul. (2011). The prevalence of political discourse in non-political blogs. In Adamic, Lada A., Baeza-Yates, Ricardo A., Counts, Scott (Eds.), Proceedings of the Fifth International Conference on Weblogs and Social Media (pp. 233–240). Barcelona, Catalonia, Spain.

[26] Ochs, Elinor., Capps, Lisa. (2001). *Living narrative: Creating lives in everyday storytelling*. Cambridge, MA: Harvard University Press.

[27] Pang, Bo., Lee, Lillian. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In Scott, Donia., Daelemans, Walter., Walker, Marilyn A. (Eds.), *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics* (pp. 271–278). Stroudsburg, PA: Association for Computational Linguistics.

[28] Pang, Bo., Lee, Lillian. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval,* 2(1–2), 1–135.

[29] Quirk, Randolph., Greenbaum, Sidney., Leech, Geoffrey., Svartvik, Jan. (1985). *A comprehensive grammar of the English language*. New York, NY: Longman.

[30] Reichenbach, Hans. (1947). *Elements of symbolic logic*. New York, NY: Macmillan Co.

[31] Riloff, Ellen., Wiebe, Janyce., Phillips, William. (2005). Exploiting subjectivity classification to improve information extraction. In Manuela M. Veloso & Subbarao Kambhampati (Eds.), Proceedings of the Twentieth National Conference on Artificial Intelligence and the Seventeenth Innovative Applications of Artificial Intelligence Conference (pp. 1106–1111). AAAI Press.

[32] Rocchio, J. J. (1971). Relevance feedback in information retrieval. *In:* Gerard Salton (Ed.), The SMART retrieval system: Experiments in automatic document processing (pp. 313–323). Prentice Hall.

[33] Snyder, Benjamin., Barzilay, Regina. (2007). Multiple aspect ranking using the good grief algorithm in human language technologies. *In:* Candace L. Sidner, Tanja Schultz, Matthew Stone., Cheng Xiang Zhai (Eds.), Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics (pp. 300–307).

[34] Sood, Sara Owsley. (2008). Buzz: Mining and presenting interesting stories. International Journal of Art

*and Technology,* 1(1).

[35] Steinitz, Rebecca. (1997). Writing diaries, reading diaries: The mechanics of memory. *Communication Review*, 2, 43–58.

[36] Tofiloski, Milan., Brooke, Julian., Taboada, Maite. (2009). A syntactic and lexical-based discourse segmenter. In *ACL 2009: In:* Proceedings of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the AFNLP, Short Papers (pp. 77–80). Association for Computational Linguistics.

[37] Wiebe, Janyce M., Wilson, Theresa., Bruce, Rebecca., Bell, Matthew., Martin, Melanie. (2004). Learning subjective language. *Computational Linguistics,* 30(3), 277–308.

[38] Wiebe, Janyce Marbury. (1990). *Recognizing subjective sentences: A computational investigation of narrative text* (Doctoral dissertation). State University of New York at Buffalo.