

Automatic Identification of Narrative Diegesis and Point of View

Joshua D. Eisenberg and Mark A. Finlayson

11200 S.W. 8th Street, ECS Building, Miami, FL 33141

School of Computing and Information Sciences

Florida International University

{jeise003, markaf}@fiu.edu

Abstract

The style of narrative news affects how it is interpreted and received by readers. Two key stylistic characteristics of narrative text are *point of view* and *diegesis*: respectively, whether the narrative recounts events personally or impersonally, and whether the narrator is involved in the events of the story. Although central to the interpretation and reception of news, and of narratives more generally, there has been no prior work on automatically identifying these two characteristics in text. We develop automatic classifiers for point of view and diegesis, and compare the performance of different feature sets for both. We built a gold-standard corpus where we double-annotated to substantial agreement ($\kappa > 0.59$) 270 English novels for point of view and diegesis. As might be expected, personal pronouns comprise the best features for point of view classification, achieving an average F_1 of 0.928. For diegesis, the best features were personal pronouns and the occurrences of first person pronouns in the argument of verbs, achieving an average F_1 of 0.898. We apply the classifier to nearly 40,000 news texts across five different corpora comprising multiple genres (including newswire, opinion, blog posts, and scientific press releases), and show that the point of view and diegesis correlates largely as expected with the nominal genre of the texts. We release the training data and the classifier for use by the community.

1 Introduction

Interpreting a text's veridicality, correctly identifying the implications of its events, and properly de-

limiting the scope of its references are all challenging and important problems that are critical to achieving complete automatic understanding of news stories and, indeed, text generally. There has been significant progress on some of these problems for certain sorts of texts, for example, recognizing implications on short, impersonal, factual text in the long-running Recognizing Textual Entailment challenge (RTE¹). On the other hand, narrative text (including much news writing) presents additional complications, in that to accomplish the tasks above one must take into account the narrator's *point of view* (i.e., first person or third person), as well as the narrator's personal involvement in the story (a feature that narratologists call *diegesis*).

In news stories specifically writers are encouraged to use the third person point of view when they wish to emphasize their objectivity regarding the news they are reporting (Davison, 1983). In opinion pieces or blog posts, on the other hand, first person is more common and implies a more personal (and perhaps more subjective) view (Aufderheide, 1997). News writers are also often in the position of reporting on events which they themselves have not directly observed, and in these cases can use an uninvolved style (known as *heterodiegetic* narration) to communicate their relative remove from the action. When writers observe or participate in events directly, however, or are reporting on their own lives (such as in blog posts), they can use an involved narrative style (i.e., *homodiegetic* narration) to emphasize their personal knowledge and subjective, perhaps biased, orientation.

¹<http://aclweb.org/aclwiki/index.php?title=RTE>

Before we can integrate knowledge of point of view (POV) or diegesis into text understanding, we must be able to identify them, but there are no systems which enable automatic classification of these features. In this paper we develop reliable classifiers for both POV and diegesis, apply the classifiers to texts drawn from five different news genres, demonstrate the accuracy of the classifiers on these news texts, and show that the POV and diegesis correlates much as expected with the genre. We release the classifiers and the training data so the field may build on our work and integrate these features into other text processing systems.

Regarding the point of view of the narrator, narratologist Mieke Bal claimed “The different relationships of the narrative ‘I’ to the objects of narration are constant within each narrative text. This means that one can immediately, already on the first page, see which is the [point of view].” (Bal, 2009, p. 29) This assertion inspired the development of the classifiers presented here: we had annotators mark narrative POV and diegesis from the first 60 lines of each of 270 English novels, which is a generous simulation of “the first page”. This observation allowed us to transform the collection of data for supervised machine learning from an unmanageable burden (i.e., having annotators read every novel from start to finish) into a tractable task (reading only the first page). We chose novels for training, instead of news texts themselves, because of the novels’ greater diversity of language and style.

Once we developed reliable classifiers trained and tested with this annotated data, we applied the classifiers to 39,653 news-related texts across five news genres, including: the Reuter’s corpus containing standard newswire reporting; a corpus of scientific press releases scraped from EurekaAlerts; the CSC Islamist Extremist corpus containing ideological story telling, propaganda, and wartime press releases; a selection of opinion and editorial articles scraped from LexisNexis, the Spinn3r web blog corpus, and . We checked a sample of the results, confirming that the classifiers performed highly accurately over these genres. The classifiers allowed us to quickly assess the POV and diegesis of the texts and show how expectations of objectivity or involvement differ across genres.

The paper proceeds as follows. In §2 we define

point of view and diegesis, and discuss their different attributes. In §3 we describe the annotation of the training and testing corpus, and then in §4 describe the development of the classifiers. In §5 we detail the results of applying the classifiers to the news texts. In §6 we outline related work, and in §7 we discuss how shortcomings of the work and how it might be improved. We summarize the contributions in §8. In short, this paper asks the question: can point of view and diegesis be automatically classified? The experimental results in this paper show that it can be done.

2 Definitions

2.1 Point of View

The point of view (POV) of a narrative is whether the narrator describes events in a personal or impersonal manner. There are, in theory, three possible points of view, corresponding to grammatical person: first, second, and third person. First person point of view involves a narrator referring to themselves, and implies a direct, personal observation of events. In a third person narrative, by contrast, the narrator is outside the story’s course of action, looking in. The narrator tells the reader what happens to the characters of the story without ever referring to the narrator’s own thoughts or feelings.

In theory second person POV is also possible, although exceedingly rare. In a second person narrative, the narrator tells the *reader* what he or she is feeling or doing, giving the impression that the narrator is speaking specifically to the reader themselves and perhaps even controlling their actions. This is a relatively rare point of view (in our training corpus of English novels it occurred only once), and because of this we exclude it from consideration.

Knowing the point of view (first or third person) is important for understanding the implied veridicality as well as the scope of references within the text. Consider the following example:

- (1) *John made everyone feel bad. He is a jerk.*

With regard to reference, if this is part of a first person narrative, the narrator is included in the scope of the pronoun *everyone*, implying that the narrator himself has been made to feel bad. In this case we might discount the objectivity of the second sen-

tence if we know that the narrator himself feels bad on account of John. A third person narrator, by contrast, is excluded from the reference set, one can make no inference about his internal state and, thus, it does not affect our judgment of the implications of the accuracy or objectivity of later statements.

With regard to veridicality, if the narration is third person, statements of fact can be taken at face value with a higher default assumption of truthfulness. A first person narrator, in contrast, is experiencing the events not from an external, objective point of view but from a personal point of view, and so assessment of the truth or accuracy of their statements is subject to the same questions as a second-hand report.

2.2 Diegesis

Diegesis is whether the narrator is involved (homodiegetic) or not involved (heterodiegetic) in the story. In a homodiegetic narrative, the narrator is not just the narrator but a character as well, performing actions that drive the plot forward. In a heterodiegetic narrative, the narrator is observing the action but not influencing its course. As reflected in Table 1, third person narrators are almost exclusively heterodiegetic, but first person narrators can be either. Like point of view, diegesis provides information to the reader on how to discount statements of fact, and so to judge the veridicality of the text.

3 Corpus

To train and test our classifiers we chose a corpus of diverse texts and had it annotated for point of view and diegesis. We used the Corpus of English Novels (De Smet, 2008), which contains 292 English novels published between 1881 and 1922, and was assembled to represent approximately a generation of writers from turn-of-the-century English literature. Novels were included in the corpus if they were available freely from Project Gutenberg (Hart, 1971) when the corpus was assembled in 2007. There are twenty-five authors represented in the corpus, including, for example, Arthur Conan Doyle, Edith Wharton, and Robert Louis Stevenson. Genres represented span a wide range including drama, fantasy, adventure, historical fiction, and romance.

To simulate “the first page” of each novel, we manually trimmed each text file so that they started

with the beginning of the first chapter. This was done by hand since automating this process was not a trivial task. Then, we automatically trimmed each file down to the first 60 lines, as defined by line breaks in the original files (which reflect the Gutenberg project’s typesetting). These shortened texts were used by our annotators, and were the data on which the classifiers were trained and tested.

We wrote an annotation guide for point of view and diegesis, and trained two undergraduate students to perform the annotations. The first 20 books from the corpus were used to train the annotators, and the remaining 272 texts were annotated by both annotators. After annotation was complete we realized that two of the files erroneously contained text from the preface instead of the first chapter, so we removed them from our study. Minus the training and removed texts, we produced a gold-standard corpus of 270 novels annotated for point of view and diegesis.

3.1 Inter-annotator Agreement

We evaluated the inter-annotator agreement using Cohen’s kappa coefficient (κ). For point of view κ was 0.635, which is considered substantial (Landis and Koch, 1977). The κ for diegesis is 0.592, almost substantial. Out of 270 markings, there were 36 and 33 conflicts between the annotators for POV and diegesis respectively. The first author resolved the conflicts in the POV and diegesis annotations by reading the text and determined the correct characteristic according to the annotation guide. We release this gold-standard corpus, including the annotation guide, for use by the community.²

3.2 Interaction of POV with Diegesis

Table 1 shows the distribution of the texts in the corpus across the various categories. Of the 270 texts in the corpus, 74 had first person narrators, only 1 had second person, and 195 were third person. For diegesis, 55 were homodiegetic and 215 were heterodiegetic. There was only one second person narrator; this type of narrator is atypical in narrative texts in general, and we excluded this text from training and testing.

²We have archived the code, annotated data, and annotation guide in the CSAIL Work Products section of the CSAIL Digital Archive, stored in the MIT DSpace online repository at <https://dspace.mit.edu/handle/1721.1/29808>.

	First	Second	Third
Homodiegetic	54 (20%)	1 (0.4%)	-
Heterodiegetic	20 (7.4%)	-	195 (72.2%)

Table 1: Distribution of POV and Diegesis. Each non-zero entry lists the number of texts in the category as well as the percentage of the total corpus.

As we expected, there are no third person homodiegetic texts in the training corpus. Although in principle this is possible, it is narratively awkward, requires the narrator to be involved in the action of the story (homodiegetic), but report the events from a dispassionate, third-person point of view, never referring to themselves directly. Our data imply that this type of narrator is, at the very least, rare in turn of the century English literature. More generally, from our own incidental experience of narrative, we would expect this be quite rare across narrative in general.

4 Developing the Classifiers

We implemented the preprocessing (§4.1), SVM training, cross-validation testing (§4.2), and feature extraction for the classifiers (§4.3 and §4.4) in Java³.

4.1 Preprocessing

The preprocessing was the same for both classifiers. The full text of the first 60 lines of the first chapter was loaded into a string, then all text within quotes was deleted using a regular expression. For both POV and diegesis it is important to focus on language that is uttered by the narrator, whereas quoted text represents words uttered by the characters of the narrative. The benefits of removing the quoted text is shown in Tables 3 and 4. After we removed the quoted text, we used the Stanford CoreNLP suite to tokenize and detect sentence boundaries (Manning et al., 2014). Finally, we removed all punctuation⁴. This produced an array of tokenized sentences, ready for feature extraction.

³We have archived a snapshot of the code, plus all the additional supplementary material, in the CSAIL Work Products section of the CSAIL Digital Archive, stored in the MIT DSpace online repository at <https://dspace.mit.edu/handle/1721.1/29808>.

⁴Specifically, the six characters [. ? ! , ; :].

4.2 Experimental Procedure

To determine the best sets of features for classification, we conducted two experiments, one each for POV and diegesis. In each case, texts were preprocessed as described above (§4.1), and various features were extracted as described below. Then we partitioned the corpus training and testing sets using ten-fold cross-validation. Precisely, this was done as follows: for POV, the texts annotated as first person were divided into ten sets containing nearly equal numbers of texts, and we did the same for the third person texts. Then the first set of both the first person and third person texts were designated as the test sets and the classifier was trained on the remaining nine sets from each class. This was repeated with each set (second, third, fourth, etc. . . .), designating each set in order as the test set, with the remaining sets used for training. There are more third person narrators in the corpus; hence, each training fold has more examples of third person narrators than first person narrators. We performed cross-validation for diegesis in exactly the same manner.

We then trained an SVM classifier on the training fold using specific features as described below (Chang and Lin, 2011). To evaluate performance of the classifiers we report macro-averaged precision, recall, and F_1 measure. This is done by averaging, without any weighting, the precision, recall, and F_1 from each fold. We also report the average of F_1 for overall performance (weighted by number of texts).

4.3 Determining the Best POV Feature Set

The best set of features for point of view should be straightforward: narrators either refers to themselves (first person) or they don't (third person). Naturally, a first-person narrators will refer to themselves with first person pronouns, and so the presence of first person pronouns in non-quoted text should be a clear indicator of a first person point of view. Importantly, as soon as a narrator uses a first person pronoun they become a first person narrator, regardless of how long they were impersonally narrating. A list of the sets of first, second, and third person pronouns that we used as features can be found in Table 2.

We investigated eight different features sets for POV classification. The classifier with the best per-

1st	I, me, my, mine, myself, we, us, our, ours
2nd	you, your, yours
3rd	he, him, his, she, her hers, they, them, theirs

Table 2: Pronouns used for classification.

formance uses counts of the first, second, and third person pronouns as the feature set. Six of the remaining experiments use different subsets of the pronouns: we test the performance of on each individual set of pronoun as well as each combination of two pronouns sets. Features sets that did not consider first person pronouns were unable to classify first person narrations, but, importantly, first person pronouns alone were not the best for classifying first person narratives. The classifier that considers all three types of pronouns has an F_1 almost six percentage points higher than the classifier that only considers first person pronouns.

Previously we discussed that it is important to remove quoted text before the features are extracted. To test this we ran an experiment where we did not remove quoted text in preprocessing, and then used all pronouns as in the best performing classifier. This negatively impacted F_1 for first person narrators by 13 percentage points and the F_1 for third person narrators by about 3 percentage points. This shows that it is important to remove quoted text before extracting features for POV classification. The only feature sets that did worse than the feature set with quoted text removed were those feature sets that did not include first person pronouns.

4.4 Determining the Best Diegesis Feature Set

Pronouns are also a prominent feature of diegesis, but it is not as simple as counting which pronouns are used: diegesis captures the relationship of the narrator to the story. On the one hand, if the narrator never refers to themselves (i.e., a third person narrator), then it is extremely unlikely that they are participating in the story they are telling, and so they are, by default, a heterodiegetic narrator. On the other hand, first person narrators may be either homo- or heterodiegetic. In this case one cannot merely count the number and type of pronouns that occur, but must pay attention to when first person pronouns, which represent the narrator, are used as arguments of verbs that represent events in the story. Event de-

tection is a difficult task (Verhagen et al., 2007), so we focus on finding when first person pronouns are used as arguments of any verb. While in reality not all verbs represent events, a large fraction do, and as the performance of the classifier shows this feature correlates well with the category. To find the arguments of verbs, we use our in-house semantic role labeler (SRL) that is integrated into the Story Workbench (Finlayson, 2008; Finlayson, 2011).

We tested four different sets of features for diegesis classification. The simplest counts how many times each first person pronoun appears in an argument of a verb. Although this classifier is somewhat successful, it is somewhat weak identifying homodiegetic narrators.

The best performing diegesis classifier uses occurrences of the first, second, and third person pronouns in addition to the features from the simple diegesis classifier as features. We hypothesized that we could further improve the performance of this classifier by including a feature that counted the occurrences of second and third person pronouns as arguments of verbs that also have a first person pronoun as an argument (this is listed as the “co-occurrence” feature in Table 4). Our reasoning was that this feature would encode where the narrator and another character were connected by the same event, which is indicative of homodiegesis. Contrary to our expectations, however, this feature undermined homodiegetic classification: this classifier could not train an SVM model that could recognize homodiegetic narrators. This was the weakest of all of the diegesis classifiers.

Above we claimed that removal of quoted text is useful for diegesis classification. To show this, we took the feature set from our best diegesis classifier (with first person pronouns as arguments to a verb, and the occurrences of all pronouns), and took out the quoted text removal from the pipeline. This caused the F_1 measure to drop over 13 percentage points for homodiegetic and approximately 2 percentage points for heterodiegetic. These drops in performance indicate that the classifier performs better when quoted text is removed.

Feature Set	Quoted Text Removed ↓	First Person			Third Person			Avg. F_1
		Precision	Recall	F_1	Precision	Recall	F_1	
Majority class baseline		0	0	0	0.724	1	0.839	0.607
3rd person pronouns only	✓	0	0	0	0.73	0.994	0.842	0.61
2nd person pronouns only	✓	0	0	0	0.745	0.984	0.848	0.615
2nd & 3rd person pronouns	✓	0	0	0	0.735	0.979	0.839	0.608
All pronouns		0.911	0.671	0.743	0.893	0.963	0.924	0.874
1st person pronouns	✓	0.969	0.7	0.793	0.903	0.989	0.943	0.902
1st & 3rd person pronouns	✓	0.955	0.729	0.808	0.911	0.984	0.945	0.907
1st & 2nd person pronouns	✓	0.94	0.757	0.814	0.921	0.974	0.944	0.908
All pronouns	✓	0.944	0.814	0.859	0.938	0.973	0.954	0.928

Table 3: Performance of point of view classification for different feature sets. The left hand column describes different sets of features used to train the SVM classifier. These features are extracted from the novels in the CEN. The columns to the right show the performance of each classifier when tested on CEN novels. Each data point is macro-averaged across the 10-folds of cross validation.

Feature Set	Quoted Text Removed All Pronouns → ↓	Homodiegetic			Heterodiegetic			Avg. F_1
		Precision	Recall	F_1	Precision	Recall	F_1	
Majority class baseline		0	0	0	0.796	1	0.886	0.706
1st pers. pronoun as verb arg.	✓	0.805	0.5	0.586	0.892	0.962	0.924	0.852
1st pers. as arg. + co-occurrence	✓ ✓	0.847	0.480	0.589	0.889	0.976	0.93	0.858
1st pers. pronoun as verb arg.	✓	0.907	0.58	0.677	0.91	0.981	0.943	0.886
1st pers. pronoun as verb arg.	✓ ✓	0.931	0.62	0.721	0.917	0.981	0.947	0.898

Table 4: Performance of diegesis classification for different feature sets. The “co-occurrence” feature is explained in the text. The left hand column describes different feature sets used to train the SVM classifier. These features are extracted from the novels in the CEN. The columns to the right show the performance of each classifier when tested on CEN novels. Each data point is macro-averaged across the 10-folds of cross validation.

5 Application of the Classifiers to News

To reveal the relationship of POV and diegesis to news story genres, we applied both classifiers a diverse set of news corpora. The classifiers for these experiments were trained on all 269 first and third person texts from the CEN,⁵ using the best performing sets of features. We applied the classifiers to texts drawn from five corpora: the Reuters-21578 newswire corpus,⁶ a corpus of scientific press releases scraped from EurekAlerts, a selection of opinion and editorial articles scraped from LexisNexis, the Spinn3r web blog corpus (Burton et al., 2009), and the CSC Islamist Extremist corpus containing ideological story telling, propaganda, and wartime press releases (Ceran et al., 2012). The stories from the Spinn3r web blog corpus were found by Gordon

⁵The number of texts was 269 because one text in the corpus of 270 texts was second person.

⁶<http://www.daviddlewis.com/resources/testcollections/reuters21578/>

and Swanson (2009) and the CSC Islamist extremist stories were found by Ceran *et al.* (2012). These five corpora are used for testing the POV and diegesis classifiers; these corpora are not used for training the classifiers. For each experiment in this section, the best set of POV and diegesis features from §4.3 and §4.4, were used to train a classifier, these classifiers were trained on the first page of each novel from the CEN. For each corpora, after running the classifiers we randomly sampled texts and checked their classification to produce an estimate of the true accuracy of the classifiers. Sample sizes were determined by calculating the number of samples required to achieve a 99% confidence for a point estimate of proportion, using the proportion estimated by the classifier (Devore, 2011). In all cases the ratio of first person to third person texts (and homo- to hetero-diegetic texts) was chosen to be equal to the ratio in the classification.

5.1 Reuters-21578 Newswire

This corpus contains 19,043 texts, and all but one were marked by the classifiers as third person and heterodiegetic. We expected this, as journalists typically use the third person POV and heterodiegetic narration to communicate objectivity.

The erroneous classification of one text as first person was the result of a type of language we did not anticipate. The article in question uses direct speech to quote a letter written by Paul Volcker, Federal Reserve Board chair, to President Ronald Reagan. The majority of the article is the text of the letter, where Volcker repeatedly refers to himself, using the pronoun “I”. The POV classifier interpreted this document as 1st person because the text of Volcker’s letter was not removed in the quotation removal phase. The letter is quoted using direct speech, which our simple, regular-expression-based quotation detection system cannot recognize.

To estimate the true accuracy of the POV classifier over the Reuters corpus we randomly sampled and checked the POV of 200 texts (including the single first person text). All of the classifications were correct except the single first person text, resulting in an accuracy estimate of 99.5% over the newswire text for the POV classifier (1.3% margin of error at 99% confidence).

To estimate the true accuracy of the diegesis classifier over this corpus we randomly sampled and checked the diegesis of 200 texts (including the single homodiegetic text). Of the 199 heterodiegetic texts, all were correct, while the single homodiegetic text was incorrect, resulting in an accuracy estimate of 99% for the diegesis classifier over the newswire text (1.81% margin of error at 99% confidence).

5.2 EurekAlert Press Releases

This corpus contains 12,135 texts scraped from EurekAlert⁷, dated between June 1st and December 31st, 2009. The distribution of this corpus is similar to the Reuters corpus, and over 99% of the texts were classified as third person and heterodiegetic narrations. Press offices write press releases to entice journalists to write newswire articles, and so it makes sense that they will attempt to mimic the desired narrative distance in the press release, seeking

⁷<http://www.eurekalert.org/>

to present themselves as unbiased narrators.

To estimate the true accuracy of the POV classifier over the press releases we randomly sampled and checked the diegesis of 120 texts, including two first person and 118 third person. Of the two first person texts, one was correct, and of the 118 third person texts, 115 were correct, resulting in an accuracy estimate for the POV classifier of 97% over the press release text (4.03% margin of error at 99% confidence).

To estimate the true accuracy of the diegesis classifier over this corpus we randomly sampled and checked the diegesis of 120 texts, including 2 homodiegetic and 118 heterodiegetic. Of the two homodiegetic texts, neither were correct, and of the 118 heterodiegetic texts, 111 were correct, resulting in an accuracy estimate for the diegesis classifier of 94% over the press release text (5.6% margin of error at 99% confidence).

5.3 LexisNexis Opinions and Editorials

This corpus comprises 4,974 texts labeled opinion or editorial scraped from the LexisNexis website⁸, dated between January 2012 and August 2016. Texts were included if they contained more than 100 words and appeared in one of a set of major world publications including, for example, the New York Times, the Washington Post, and the Wall Street Journal. About one-quarter of these texts are first person, and more than half of the first person narrators were homodiegetic. We expected this increased abundance of first person and homodiegetic texts, as the purpose of these types of articles is often to express individual opinions or the writer’s personal experience of events.

To estimate the true accuracy of the POV classifier over the LexisNexis articles, we randomly sampled and checked the POV of 200 texts, 50 from those classified as first person and 150 from those classified as third person. Of the 50 texts classified as first person all were confirmed correct, while of the 150 texts classified as third person only 90 were confirmed correct. This suggests that our classifier is not properly identifying all of the first person narrators in the LexisNexis corpus, and results in an accuracy estimate of 70% for the POV classifier over

⁸<http://www.lexisnexis.com/hottopics/lnacademic/>

the LexisNexis texts (2.7% margin of error at 99% confidence).

To estimate the true accuracy of the diegesis classifier over this corpus we randomly sampled and checked the diegesis of 200 texts, including 24 homodiegetic and 126 heterodiegetic texts. Of the 24 homodiegetic texts, all were correct, and of the 126 heterodiegetic texts, 51 were correct, allowing us to estimate that the diegesis classifier has an accuracy of 40% over the press release text (11% margin of error at 99% confidence).

5.4 Spinn3r Web Blogs

This corpus comprises 201 stories extracted by Gordon and Swanson (2009) from the Spinn3r 2009 Web Blog corpus (Burton et al., 2009). These texts come from web blogs, where people often tell personal stories from their perspective, or use the blog as a public journal of their daily life. In contrast with newswire text, there is no expectation that a blog will report the truth in an unbiased manner. The distribution of the POV on this corpus reflects this tendency, with 66% of the texts being first person.

The diegesis distribution for the web blog stories was not unexpected: slightly more than half of the blog stories with first person narrators are homodiegetic. These are the most *personal* stories of the web blog story corpus, in which the narrator is involved in the story's action.

To estimate the true accuracy of the POV classifier on the Spinn3r corpus, we randomly sampled 20 texts, 13 from those classified as first person and 7 classified as third person. Of the 13 first person texts 9 were confirmed correct, while of the 7 third person texts only 3 were confirmed correct. Overall, our classifier has trouble classifying the web blog texts. This might be due to syntactic irregularities of blog posts, which vary in their degree of adherence to proper English grammar. With respect to third person narrators we estimate that the POV classifier has an accuracy of 42% over the web blog text (34% margin of error at 99% confidence).

To estimate the true accuracy of the diegesis classifier over this corpus we randomly sampled and checked the diegesis of 20 texts, including six homodiegetic and 14 heterodiegetic texts. Of the six homodiegetic texts, all were correct, and of the 14 heterodiegetic texts, three were correct. With re-

spect to the heterodiegetic narrators we estimate that the diegesis classifier has an accuracy of 21% over the press release text (27% margin of error at 99% confidence).

5.5 Islamic Extremist Texts

The CSC Islamist Extremist corpus contained 3,300 story texts, as identified by Corman *et al.* (2012). These texts were originally posted on Islamist Extremist websites or forums. Our POV classifier found that 99.7% of the extremist stories were written in the third person. For the most part, the extremist stories were second hand accounts of events, often to share news about the outcome of battles or recount the deeds of Jihadists.

To estimate the true accuracy of the POV classifier on this corpus, we randomly sampled 150 texts, 2 from those classified as first person, and 148 classified as third person. Both of the texts classified as first person were verified to be first person narrators. Of the 148 texts classified as third person, 139 were verified correct. With respect to third person narrators, we can estimate the classifier has an accuracy of 93.9% over the extremist texts (4.92% margin of error at 99% confidence).

To estimate the true accuracy of the diegesis classifier over this corpus we randomly sampled and checked the diegesis of 150 texts, including 2 homodiegetic and 148 heterodiegetic texts. Of the 2 homodiegetic texts, 1 was correct, and of the 148 heterodiegetic texts, 137. With respect to heterodiegetic narrators, we can estimate the classifier has an accuracy of 92% over the press release text (5.6% margin of error at 99% confidence).

6 Related Work

As far as we know this is the first study on the automatic classification of point of view and diegesis at the level of the text. In his book "Computational Modeling of Narrative", Mani framed the problem of computational classification of narrative characteristics, including point of view and diegesis, defining with reference to narratology (Mani, 2012). He gives a framework for representing features and characteristics of narrative in his markup language *NarrativeML*. However, he does not actually implement a classifier for these characteristics.

Corpus	# Texts	1st Person	3rd Person	Homo.	Heterodiegetic	Accuracy Estm.
Reuters-21578	19,043	1 (<1%)	19,042 (~100%)	1 (<1%)	19042 (~100%)	99% / 99%
EurekAlert	12,135	31 (<1%)	12,104 (~100%)	5 (<1%)	12,129 (~100%)	97% / 94%
CSC Extremist	3,300	42 (1%)	3,258 (99%)	15 (<1%)	3,285 (~100%)	94% / 92%
Lexis Nexis	4,974	1,290 (26%)	3,684 (74%)	818 (16%)	4,156 (84%)	70% / 40%
Spinn3r	201	133 (66%)	68 (34%)	67 (33%)	134 (67%)	42% / 21%

Table 5: POV and Diegesis classifications of texts across corpora. Total number of texts was 39,653. The columns labeled “1st Person”, “3rd Person”, “Homo.”, and “Heterodiegetic” indicate the number of texts placed in each class by the classifiers trained on the CEN corpus. Percentages in parentheses indicate the fraction of that corpus falling into the specified category. The last column reports the measured accuracy of the classifiers as determined by randomly sampling and checking the results: the first percentage refers to the POV classifier and the second percentage to the diegesis classifier.

Wiebe proposed an algorithm for classifying psychological point of view in third person fictional narratives (Wiebe, 1994). The algorithm is a complex rule-based classifier which tracks broadening and narrowing of POV, and reasons whether each sentence is objective or subjective. She discusses a study where people used the algorithm to classify sentences, but the accuracy of people in that task was not given. Thus, while intriguing, it is not clear how well this algorithm performs since its correctness was not verified with a human annotated corpus.

In more recent work, Sagae *et al.* employed a data-driven approach for classifying spans of objective and subjective narrations (Sagae *et al.*, 2013). Their experiments were performed on a corpus of 40 web blog posts from the Spinn3r 2009 web blog corpus (Burton *et al.*, 2009). Their features included lexical, part of speech, and word/part of speech tag n-grams. The granularity of their classifier is fine grained, in that the system tags spans of text within a document, as opposed to our classifiers which classify the whole document.

7 Discussion

Our best classifier for POV uses the occurrence of all pronouns as features, with an F_1 of 0.857 for first person POV, and 0.954 for third person POV. The weighted average over the two classes is a 0.928 F_1 . Table 3 contains the results for the POV classification experiments. This is a great start for the automatic classification of POV, and comes close to human performance. It is reasonable and expected from narratological discussion that the best set of features is the number of first, second, and third per-

son pronouns in non-quoted text.

The best diegesis classifier in our study, the one that counts the first person pronouns as verb arguments as well as the occurrence of each pronoun, has an F_1 of 0.721 for homodiegetic, and 0.947 for heterodiegetic. The weighted average over the two classes is a 0.898 F_1 . Table 4 contains the results for the diegesis classification experiments. This is a good first start for diegesis classification, but the performance for homodiegetic narrators falls short. The features for this classifier are also reasonable: first person pronouns in verb arguments shows that the narrator is either causing action to happen or being affected by actions, and so should naturally correlate with homodiegesis. The inclusion of all pronouns as a feature for diegesis also makes sense, as point of view and diegesis are closely correlated. As noted previously, third person narrators cannot refer to themselves, so they cannot be related to the story.

The best performing POV and diegesis classifiers performed significantly than their respective baseline classifiers. In Table 3, the majority class baseline classifier has 0.607 F_1 , while the best POV classifier has 0.928 F_1 . Table 4 shows that the majority baseline classifier for diegesis has 0.706 F_1 , while the best diegesis classifier has 0.898 F_1 .

Diegesis classification might be improved by restricting pronoun argument detection only to those verbs that actually indicate events in the story. This focuses the classifier on places where the narrator is involved in driving the story forward, which is more closely aligned with the definition of diegesis. To do this, we would need to incorporate an automatic event detector (Verhagen *et al.*, 2007, *e.g.*). On the other hand, event detection currently is not espe-

cially accurate, and incorporating such a feature may very well depress our classification performance.

Another approach of interest would be to adapt our classifiers to detect if a narrative characteristic changes over the course of a text. Our study focused on short spans of traditional, formal, edited novels where the point of view and diegesis remained constant. In longer texts it is possible that these characteristics could change, for example, in a stream of text comprised of multiple narratives, or in a text which explicitly is trying to defy convention (e.g., in highly literary texts such as James Joyce's *Ulysses*).

Finally, our classifier assumed that the classified texts were all approximately the same length (i.e., the first page, or approximately 60 lines). A modification that would be important to explore is using densities or ratios for the occurrences of the pronouns, instead of raw counts, for classifying texts that are less than 60 lines long.

8 Contributions

In this paper, we described and made significant progress against the problem of automatic classification of narrative point of view and diegesis. We demonstrated a high performing classifier for point of view with 0.928 F_1 , and a good classifier for diegesis with 0.898 F_1 . To evaluate our classifiers we created a doubly annotated corpus with gold-standard annotations for point of view and diegesis—based on the first 60 lines—of 270 English novels. We applied these classifiers to almost 40,000 news story texts drawn from five different corpora, and show that the classifiers remain highly accurate and that the proportions of POV and diegesis they identify correlates in an expected way with the genre of the news texts. We provide the annotation guide, annotated corpus, and the software as resources for the community.

Acknowledgments

This work was partially supported by National Institutes of Health (NIH) grant number 5R01GM105033-02. Thanks to Fernando Serrano and Victor Alvarez for their work annotating the Corpus of English Novels. We also thank Professor Steve Corman for providing access to his corpus of Islamic Extremist Texts.

References

- [Aufderheide1997] Patricia Aufderheide. 1997. Public intimacy: The development of first-person documentary. *Afterimage: The Journal of Media Arts and Cultural Criticism*, 25(1):16–18.
- [Bal2009] Mieke Bal. 2009. *Narratology: Introduction to the Theory of Narrative*. University of Toronto Press, Toronto.
- [Burton et al.2009] Kevin Burton, Akshay Java, and Ian Soboroff. 2009. The ICWSM 2009 Spinn3r dataset. In *Proceedings of the 3rd Annual Conference on Weblogs and Social Media (ICWSM 2009)*, San Jose, CA.
- [Ceran et al.2012] Betul Ceran, Ravi Karad, Steven Corman, and Hasan Davulcu. 2012. A hybrid model and memory based story classifier. In *Proceedings of the 3rd International Workshop on Computational Models of Narrative (CMN'12)*, pages 60–64, Istanbul, Turkey.
- [Chang and Lin2011] Chih-Chung Chang and Chih-Jen Lin. 2011. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):1–27.
- [Davison1983] W. Phillips Davison. 1983. The third-person effect in communication. *Public Opinion Quarterly*, 47(1):1–15.
- [De Smet2008] Hendrik De Smet. 2008. Corpus of english novels. <https://perswww.kuleuven.be/u0044428/cen.htm>.
- [Devore2011] Jay L. Devore. 2011. *Probability and Statistics for Engineering and the Sciences*. Cengage Learning, Boston, MA, 8th edition.
- [Finlayson2008] Mark A. Finlayson. 2008. Collecting semantics in the wild: The story workbench. In *Proceedings of the AAAI Fall Symposium on Naturally Inspired Artificial Intelligence (NIAI)*, pages 46–53, Arlington, VA.
- [Finlayson2011] Mark A. Finlayson. 2011. The Story Workbench: An extensible semi-automatic text annotation tool. In *Proceedings of the 4th Workshop on Intelligent Narrative Technologies (INT4)*, pages 21–24, Stanford, CA.
- [Gordon and Swanson2009] Andrew Gordon and Reid Swanson. 2009. Identifying personal stories in millions of weblog entries. In *Proceedings of the 3rd International Conference on Weblogs and Social Media (ICWSM 2009), Data Challenge Workshop*, San Jose, CA.
- [Hart1971] Michael Hart. 1971. Project Gutenberg. <https://www.gutenberg.org/>.
- [Landis and Koch1977] Richard Landis and Gary Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.

- [Mani2012] Inderjeet Mani. 2012. *Computational Modeling of Narrative*. Morgan & Claypool Publishers, Williston, VT.
- [Manning et al.2014] Christopher D Manning, Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, System Demonstrations*, pages 55–60, Baltimore, MD.
- [Sagae et al.2013] Kenji Sagae, Andrew Gordon, Morteza Dehghani, Mike Metke, Jackie Kim, Sarah Gimbel, Christine Tipper, Jonas Kaplan, and Mary Helen Immordino-Yang. 2013. A data-driven approach for classification of subjectivity in personal narratives. In *Proceedings of the 5th International Workshop on Computational Models of Narrative (CMN'13)*, pages 198–213, Hamburg, Germany.
- [Verhagen et al.2007] Marc Verhagen, Robert Gaizauskas, Frank Schilder, Mark Hepple, Graham Katz, and James Pustejovsky. 2007. SemEval-2007 task 15: TempEval temporal relation identification. In *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007)*, pages 75–80, Prague, Czech Republic.
- [Wiebe1994] Janyce M Wiebe. 1994. Tracking point of view in narrative. *Computational Linguistics*, 20(2):233–287.