

# Towards Automatic Detection of Narrative Structure

Jessica Ouyang, Kathleen McKeown

Columbia University  
New York, NY 10027, USA  
ouyangj@cs.columbia.edu, kathy@cs.columbia.edu

## Abstract

We present novel computational experiments using William Labov’s theory of narrative analysis. We describe his six elements of narrative structure and construct a new corpus based on his most recent work on narrative. Using this corpus, we explore the correspondence between Labov’s elements of narrative structure and the implicit discourse relations of the Penn Discourse Treebank, and we construct a mapping between the elements of narrative structure and the discourse relation classes of the PDTB. We present first experiments on detecting Complicating Actions, the most common of the elements of narrative structure, achieving an f-score of 71.55. We compare the contributions of features derived from narrative analysis, such as the length of clauses and the tenses of main verbs, with those of features drawn from work on detecting implicit discourse relations. Finally, we suggest directions for future research on narrative structure, such as applications in assessing text quality and in narrative generation.

**Keywords:** narrative, annotation, discourse

## 1 Introduction

Labov and Waletzky (1967) pioneered the study of narrative in their analysis of oral narratives of personal experience – true stories from the narrators’ lives, collected through sociolinguistic interviews. From their seminal paper, the discipline of narrative analysis emerged, but it is only recently that computational linguists have begun to tackle questions of narrative. The theory put forth by Labov and Waletzky (1967) and refined by Labov (2013) has yet to be approached computationally, and its relationship with other tasks in computational linguistics has remained unexplored.

In this work, we describe Labov’s theory of narrative analysis (Section 2). We construct a new corpus of narrative structure directly from narratives collected and annotated by Labov (2013) (Section 3). We use this corpus to explore the similarities between narrative structure and implicit discourse relations (Section 4). Detecting implicit discourse relations is difficult, and state-of-the-art systems achieve results ranging from 26.57 f-score on some classes of discourse relations to 79.22 f-score on others (Pitler et al., 2009; Zhou et al., 2010; Park and Cardie, 2012; Biran and McKeown, 2013). We show that the elements of narrative structure defined by Labov (2013) can be mapped to the four classes of discourse relations annotated in the Penn Discourse Treebank (Prasad et al., 2008a).

We propose a computational linguistics task that embodies Labov’s theory of narrative structure (Section 5) and conduct preliminary experiments using supervised learning on our corpus to automatically detect Complicating Actions, a narrative structure element, achieving 71.55 f-score (Sections 6). Our results show that the theory proposed by Labov and Waletzky (1967) describes a narrative structure that can be detected automatically. In addition, we compare the informativeness of features adapted from Labov’s theory of narrative analysis to that of features drawn from work on implicit discourse relations and find that, while the narrative features outperform the discourse features, the

best results are achieved using both feature sets (Section 7). Our findings suggest that integrating the task of detecting narrative structure with that of detecting discourse relations may help both.

Finally, we suggest directions for future work on detecting narrative structure and argue that this work can be leveraged to improve the detection of implicit discourse relations (Section 8).

## 2 Narrative Analysis

Labov and Waletzky (1967) defined a structure of narrative consisting of three elements: the orientation, the complicating action, and the evaluation. Labov (2013) refined this structure to include additional elements: the abstract, the resolution, and the coda. Each clause of a narrative is assigned to one of these elements of narrative structure, but not all elements are necessary in every narrative – the original three defined by Labov and Waletzky (1967) are sufficient for a narrative.

**The Abstract** Narratives are organized around a single “most reportable event”. Of all the events in the story, this is the most fantastic (the least credible) and has the greatest effect on the lives of the characters. The abstract is an introduction to the story and often contains a description of the most reportable event. For example,

Shall I tell you about the first man got kilt – killed by a car here... Well, I can tell you that.

is the abstract of a narrative collected by Labov (2013).

**The Orientation** The orientation contains information on the time, the place, and the persons involved in the story – the background information. It usually occurs at the beginning of the narrative, but some orienting information may be postponed until later in the narrative, just before it becomes relevant. An example of this is found in “Jacob Schissel’s story”, a narrative collected by Labov (2013): the orienting information

When I let go his arm, there was a knife on the table,

is given towards the end of the narrative, just before the Schissel is stabbed with the knife.

**The Complicating Action** The complicating action is a chain of causal or instrumental events that culminates in the most reportable event. The complicating action chain tells what happened in the story. In “Jacob Schissel’s story”, the chain of complicating actions is as follows:

1. He saw a rat out in the yard
2. and he started talk about it [sic]
3. and I told him to cut it out.
4. ...I grabbed his arm
5. and twisted it up behind him.
6. he picked up [the knife]
7. and he let me have it.

Each event is causally related to the one before it, except for events 5 and 7, which are instrumentally related to events 4 and 6.

**The Evaluation** The evaluation is where the narrator gives his opinions on the events of the story, considers alternative outcomes, assigns praise or blame to the characters, or attempts to add credibility to the story. Evaluations usually come at the end of a narrative, but like orientations, they can be interjected among the events of the complicating action. For example, Jacob Schissel gave this evaluation on being stabbed:

And the doctor just says, “Just that much more,” he says, “and you’d a been dead.”

This evaluation serves two purposes: first, it presents an alternative outcome in which the narrator did not survive the stabbing; second, it adds credibility to the stabbing by quoting a third party witness, the doctor.

**The Resolution** Some narratives extend the chain of events to a final resolution of the situation created by the most reportable event. For example, in a narrative about a fight, the narrator gives the resolution,

An’ they took us – they took us to the hospital.

**The Coda** The coda signals the end of the story by bringing the listener back to the present. For example, in the story about the fight, the lines

An’ that was it. That’s the only fight I can say I ever lost.

relates the events of the narrative to the present.

### 3 Corpus

We created a corpus from 20 oral narratives of personal experience collected by Labov (2013). Of these, ten were longer, episodic narratives, where each episode had its own internal structure. We treated each episode as an independent narrative, for a final count of 49 such narratives. Labov (2013) also discusses biblical and historical narratives, but these are very different in style from the narratives of personal experience that make up the bulk of his work, so we

did not include them. The corpus is small, but we refrained from including additional narratives. We used only narratives collected and annotated by Labov (2013) because we tried to represent his theory as accurately as possible in our work.

We did measure interannotator agreement in order to establish a ceiling on performance for our computational experiments and to determine how difficult it would be to train new annotators. We compared two annotators to Labov’s gold standard annotations: one of the authors, who was trained by Labov in a semester-long course on narrative analysis and had high agreement with Labov (86.04%;  $\kappa = 0.85$ ), and an undergraduate computer science student who was trained by the authors over the course of two weeks and had only moderate agreement with Labov (61.08%;  $\kappa = 0.58$ ). Most disagreements between the annotators and Labov were on Evaluation labels. In the example previously given,

And the doctor just says, “Just that much more,” he says, “and you’d a been dead.”

Labov chose the Evaluation annotation because he believed that the narrator intended to use this quote to add credibility to his story. However, there is no way to know for sure what the narrator intended, so whether an annotator marks this line as Evaluation or Complicating Action depends on what he or she believes the narrator was trying to say.

Labov was careful to transcribe each narrative exactly as it was presented by the original narrator, including the narrator’s gestures and non-verbal reactions from the interviewer and other listeners. We removed all non-verbal communications and retained only what was said by the narrator. We also manually normalized speakers’ accents. These normalizations are systematic throughout the narratives where they appear and therefore are unrelated to narrative structure. For instance, in a narrative taken from an interview with a Scottish speaker, Labov (2013) transcribes the speaker’s accent in words such as “doon” and “ootside”, “couldnae” and “wasnae”; we normalized these words to “down” and “outside”, “could not” and “was not”. These normalizations allowed us to use off-the-shelf tools for part-of-speech tagging and parsing.

Labov (2013) manually split each narrative into clauses and annotated each clause with one of the elements of narrative structure. However, because the narratives are transcribed speech, many of Labov’s clauses consist of run-on sentences or multiple fragments of sentences. As will be discussed in Section 6, some of our features are based on the verb phrase and main verb of the clause, so we need to consider each independent clause separately. As a result, in building the corpus, we used an automatic, rule-based chunker to split narratives into clauses. Each resulting clause was annotated with the label of the corresponding Labov clause. In many cases, our chunker split Labov clauses into multiple clauses. For example, the following is split into two clauses at the comma after “again”:

He said, “Well babe,” he says uh, he said, “you wanna get – you can get married again, you can do anything you want!”

Arguments		Predicted Relation Class				Total
		Temporal	Contingency	Comparison	Expansion	
Abstract,	Abstract	0	6	4	1	11
Abstract,	Orientation	0	6	4	7	17
Orientation,	Orientation	3	77	78	86	244
Orientation,	Complicating Action	0	25	13	29	67
Complicating Action,	Orientation	0	9	9	15	33
Complicating Action,	Complicating Action	2	188	127	134	451
Complicating Action,	Evaluation	0	42	12	14	68
Evaluation,	Complicating Action	0	22	10	8	40
Evaluation,	Evaluation	0	71	33	33	137
Evaluation,	Resolution	0	8	4	5	17
Resolution,	Resolution	0	11	8	7	26
Total		5	467	307	340	1,119

Table 2: Mapping from narrative structure to discourse relations.

Structural Element	Number of Clauses
Abstract	25
Orientation	340
Complicating Action	573
Evaluation	262
Resolution	47
Coda	30
Total	1,277

Table 1: Distribution of clauses in the corpus.

In other cases, our automatically chunked clauses consisted of multiple, consecutive Labov clauses joined together; these were annotated with the Labov clause with which they had the greatest overlap. For example, Labov (2013) treats the following as four clauses:

CA So he walked into Martin Cassidy’s house,  
his own house,  
CA came out with a double-bitted axe,  
RS hit him across the head once,  
RS turned over,

We treat this as a single clause with the label CA (complicating action). For convenience, we will refer to a clause that is labeled as part of the abstract as an Abstract, a clause that is labeled as part of the orientation as an Orientation, and so on.

Our finished corpus consists of 1,277 clauses distributed among the structural elements as shown in Table 1. In Labov’s gold standard annotations, there are only 1,233 clauses: 20 fewer Orientations, 3 fewer Complicating Actions, 29 fewer Evaluations, 5 fewer Resolutions, and 3 fewer Codas than our corpus. Our automatic clause chunker performed 82 joins, where two consecutive Labov clauses were joined into one clause, and 127 splits, where one Labov clause was split into two clauses.

## 4 Comparison to Implicit Discourse Relations

We hypothesize that the tasks of detecting narrative structure and detecting implicit discourse relations are related. While discourse relations hold between two arguments and narrative structure elements are assigned to individual clauses, the four sense classes of discourse relations defined by Prasad et al. (2008b) can be mapped to the elements of narrative structure that serve as their arguments. For clarity, PDTB relations are in *italics*.

To test this hypothesis, we use a modified version of the system described by Biran and McKeown (2013) to identify implicit discourse relations between adjacent clauses in our corpus. In private communication, Biran noted two differences between this system and that which was published. First, this system uses culled word-pair lists that are more practical for applications than the very large word-pair lists used in the publication. Here, only the top 10% of pairs are used, with only a slight decrease in performance. Second, this system implements five-way classification, which the publication did not. The four binary classifiers described in the publication are cascaded in order of precision, and *NoRel*, or no relation, is predicted if none of the binary classifiers predicts a relation. The accuracy of this modified classifier on a balanced test set with *NoRel* and the four classes of discourse relations equally represented is 51.75%. The output of the classifier on our corpus is shown in Table 2. We only show narrative structure pairs that occurred more than ten times in the corpus.

*NoRel* is never predicted, and the distribution of discourse relation sense tags in our corpus is different from that in the Penn Discourse Treebank (PDTB). In our corpus, *Contingency* relations are the most common, while *Expansion* relations are the most common in the PDTB (Prasad et al., 2008a). In both corpora, implicit *Temporal* relations are relatively rare.

Table 2 also shows that we can frame detecting narrative structure as a coarse-grained discourse task. Assigning a clause to one of the elements of narrative structure allows us to narrow down the classes of discourse relations in

which it can be an argument because certain relation classes are more likely to occur between particular narrative structures. For example, about half of all relations in which one argument is an Evaluation are *Contingency* relations, while relations in which one argument is an Orientation are more likely to be *Expansions*.

To better understand the correspondences between discourse relations and elements of narrative structure, we analyzed a subset of ten narratives from our corpus (289 pairs of consecutive clauses) that Biran annotated by hand with the relation classes, types, and subtypes used in the PDTB. For each class of PDTB relations below, we identify the frequency of PDTB relations within this class in our corpus and how they corresponds to Labov’s narrative structure.

**Temporal** The two types, *Temporal:Synchrony* and *Temporal:Asynchronous*, occur with equal frequency in this subset of our corpus.

**Contingency** *Contingency:Cause* relations occur between pairs of Orientations and Complicating Actions and pairs of two Complicating Actions. This corresponds to Labov’s assertion that the events of a complicating action chain must be causally or instrumentally related to each other or to background information or characters’ motivations. *Contingency:Pragmatic Cause* relations, where the first argument makes a claim that the second argument justifies, are rare and occur between pairs of Complicating Actions and Evaluations. These relations occur when the narrator is explaining a character’s motivations:

Arg1 I talked him out of it and says, “Well, we’ll go look for her, and if we can’t find her well you can – go ahead, pull the trigger if you want to.”  
Arg2 I was maneuvering.

*Pragmatic Cause* relations in which the justification is stated first, and the claim second, do not occur in the PDTB, but they do appear in our corpus:

I hurt my one knee... And I thought that was what my wife was coming to the shipyard to see me for, right? And here it is, I mean.  
Arg1 “Oh, hey babe,” I said, “they got limousines that take you home if you’re hurt. You know.  
Arg2 You didn’t have to come see me.”

This may be due to a difference in genre; the PDTB annotates the *Wall Street Journal* corpus, while our narratives are transcribed speech.

**Comparison** *Comparison:Concession* relations, where one argument denies an expectation presented by the other, are the only type of *Comparison* to occur in this subset of our corpus. This may be because the narratives in our corpus are about life-and-death situations, which Labov asserts are intrinsically surprising and unusual. For example, the above example about the narrator’s wife coming to see him at work continues,

“Oh, hey babe,” I said, “they got limousines that take you home if you’re hurt. You know.  
Arg1 You didn’t have to come see me.”  
Arg2 And here it is she’s coming to tell me that my son died...

Here, the first argument sets an expectation for why the narrator’s wife has come, and the second argument violates that expectation. Another example is,

And I thought – I thought I’d bought it. I said, “Whoa, whoa, wait a minute, I’m not dead yet!”  
Arg1 And the guys said, “No, we’re just shielding you from the sun.”  
Arg2 Boy, I thought I was dead!

Here the denial is given by the first argument, and the expectation is given by the first.

**Expansion** *Expansion:Conjunction* relations are extremely common in this subset of our corpus. *Expansion:Restatement* relations are also relatively common and likely correspond to the narrator emphasizing a particular piece of information. *Expansion:Alternative* relations are rare and are used when the narrator is unsure of or does not clearly remember a piece of information.

## 5 Task Description

In the rest of the paper, we describe our experiments on automatically detecting narrative structure using supervised learning. Our ultimate goal is, for every clause in a narrative, to label it with one of the elements of narrative structure, as annotated by Labov (2013); in this paper, we focus on labeling the complicating action. For example, the clause

and Dad said, “George, I’m so sorry to hear about the death of –”

should be labeled as Complicating Action, while the clause

This was a good many years ago.

should be labeled as Other (it is part of the orientation), and the clause

“If he had died, I would have been one of the first people to know.”

should also be labeled as Other (it is part of the evaluation).

### 5.1 Related Work on Narrative

Computational approaches to narrative are largely split between work on characters and work on plot. Our experiments fall into the category of plot. Previous work in this area has focused on causality and temporal relations in narrative.

Chambers and Jurafsky (2008) described methods for learning narrative event chains, which are sequences of temporally ordered events – pairs of verbs and typed dependencies – involving a single actor called the protagonist. Narrative chains are similar to the complicating action chains defined by Labov (2013); a complicating action chain contains multiple narrative chains, one for each character in the story. Chambers and Jurafsky (2008) discussed two tasks: narrative cloze, in which they predicted an event that was missing from a narrative chain, and ordering events, in which they classified whether or not there was a *before* relation between two events. Both of these tasks

are related to Chambers and Jurafsky’s later work on narrative schemas, sequences of events that are characteristic of a particular domain (Chambers and Jurafsky, 2009). In contrast, our work focuses not on whether a Complicating Action chain is typical of stories in a particular domain, but rather on whether the Complicating Actions can be distinguished from the other structural elements in the story.

Elson and McKeown (2009) created a tool for generating semantic encodings of narratives. They later extended this work to produce a system for generating narrative text from semantic encodings, using tense and aspect to express the temporal relationships among events and states (Elson and McKeown, 2010). This differs from our work in that Labov’s theory of narrative structure focuses on the functions of clauses in a narrative and does not consider the semantics behind them.

Reidl and Young (2010) created a narrative planner that considered both causality among events and character intentionality to generate coherent narratives with believable characters. Rather than detecting Complicating Actions, as we do in this work, Reidl and Young (2010) generate them by choosing a sequence of actions to transition from an initial state to a pre-defined outcome state. However, the structure of the narratives produced by this system consists of a block of Orientations describing the initial state, followed by a block of Complicating Actions, and ending with a block of Resolutions describing the outcome state. This structure is relatively simple compared to the structures produced by human narrators, which are the focus of our work.

## 6 Experiments

We focused on the task of distinguishing the Complicating Action from the rest of the narrative – just under half of the clauses in our corpus are Complicating Actions, while Orientations and Evaluations are less common, and Abstracts, Resolutions, and Codas are rare.

### 6.1 Features

We used a bag of words over each clause, with proper nouns replaced with their named entity types as tagged by the Stanford Named Entity Recognizer (Finkel et al., 2005). We also used several features adapted from Labov’s theory of narrative analysis, as well as features drawn from work on detecting implicit discourse relations (Pitler et al., 2009). The organization of these features is shown in Figure 1.

For each clause, we also included the features of  $n$  preceding clauses to account for some narrative elements being more likely to follow others. For example, Complicating Actions are very likely to be preceded by Orientations and other Complicating Actions, somewhat likely to be preceded by Evaluations, and highly unlikely to be preceded by Abstracts, Resolutions, and Codas. We experimented with  $n$  ranging from 0 to 9. An experiment with  $n = 3$ , for example, would consider  $n + 1$ , or four, times as many features as an experiment with  $n = 0$ .

#### 6.1.1. Narrative Features

**Length of Clause** The number of words in the clause. This serves as a simple approximation of the complexity of

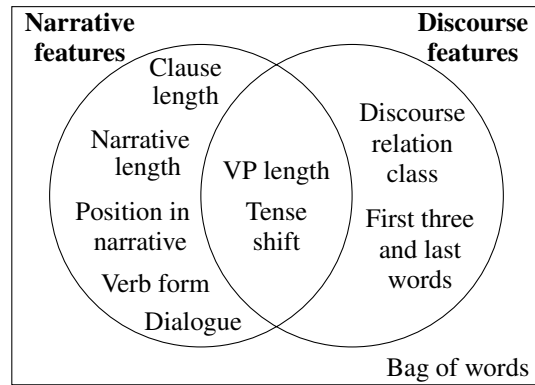


Figure 1: The organization of features.

a clause; Complicating Action clauses tend to be short and simple, while the other structural elements, and especially Evaluations, tend to be longer and more complex.

**Length of Narrative** The total number of clauses in the narrative containing the clause. Shorter narratives tend to consist mostly of Complicating Actions, while the other structural elements are more likely to occur in longer narratives.

**Position in Narrative** Which quarter of the narrative the clause is in. Complicating Actions occupy the middles of narratives, while the other structural elements tend to occur near the beginnings (Abstracts, Orientations) or the ends (Evaluations, Resolutions, Codas).

**Verb Form** The tense of the main verb and the presence of modals in the clause. Complicating Action clauses tend to be in either the past tense or what Labov terms the “historical present”, which is most commonly used with dialogue (“And then he says...”). The tenses of the other structural elements are more varied. The verbs and modals were tagged using the Stanford Parser (Klein and Manning, 2003) and were represented by binary features enumerating all possible tenses.

**Dialogue** Whether or not the clause contains a line of dialogue. Dialogue is more common in Complicating Actions than in the other elements of narrative structure.

#### 6.1.2. Discourse Features

**Explicit Relation Class** The sense classes of any explicit PDTB connectives in the clause. Pitler et al. (2009) reported that some implicit discourse relations were more likely to appear immediately before or after certain explicit discourse relations. The sense classes were automatically extracted using the tool described by Pitler and Nenkova (2009) and represented by four binary features, one for each class.

**Implicit Relation Class** The sense classes of the implicit relation between the clause and the preceding clause and that between the clause and the following clause. The relations were automatically identified using the classifier described in Section 4.

**First Three and Last Words** The first three words and the last word of the clause. Pitler et al. (2009) noted that these words tend to be connective-like and often correspond to the alternatively lexicalized relations in the PDTB. These

were represented as binary features enumerating the words that occur in the first three and last positions of all clauses in the corpus.

### 6.13. Shared Features

These features were motivated both by Labov’s theory of narrative analysis and work on implicit discourse relations.

**Length of Verb Phrase** The number of words in the largest VP subtree in the clause, as produced by the Stanford Parser (Klein and Manning, 2003). Like the length of the clause, this is also an approximation for clause complexity. In addition, Pitler et al. (2009) suggested that longer verb phrases may correspond to *Contingency* relations, while shorter verb phrases correspond to *Expansion* and *Temporal* relations.

**Tense Shift** Whether or not the tense of the main verb in the clause is different from that of the main verb in the preceding clause. A tense shift may occur when transitioning between the different elements of narrative structure, such as from a past perfect Orientation to a historical present Complicating Action. From a discourse perspective, Pitler et al. (2009) suggested that tense shifts occur in *Contingency* and *Temporal* relations but not in *Expansion* relations. In addition, Labov (2013) proposes that tense shifts occur when the narrator tries to emphasize a part of the narrative. In a narrative told mostly in the past tense, the most reportable event may be given in the historical present, and in a narrative told in the historical present, the most reportable event may be given in the past.

## 6.2 Methodology

We partitioned our corpus into a balanced training set of 918 clauses: 459 Complicating Actions and 459 Other clauses. Our testing set consisted of 114 Complicating Actions and 140 Other clauses. In both cases, the Other clauses were selected following the natural distribution of structural elements, shown in Table 1.

We trained a logistic regression classifier in Mallet (McCallum, 2002) and chose the number of preceding clauses  $n = 9$  using 10 runs of two-fold cross-validation. We experimented with other classifiers, including C4.5 and Naive Bayes, but these did not perform as well in cross-validation.

## 7 Results

We achieved the best results using a logistic regression classifier on all features with  $n = 9$ , which we present in Table 3.

Experiment	f-score
All features	71.55
Bag of words features	64.42
Majority (Other) baseline	61.96
Random baseline	49.60

Table 3: Best results.

### 7.1 Analysis and Discussion

Table 4 shows the confusion matrix for our best results. Abstracts and Resolutions were relatively easy to distin-

		Predicted	
		CA	Other
Actual	AB	0	5
	OR	10	58
	CA	81	33
	EV	20	32
	RS	1	8
	CD	2	4

Table 4: Confusion matrix for best results.

guish from Complicating Actions, while Evaluations and Codas were especially difficult.

We compared the performance of the narrative-motivated features with that of the discourse-motivated features (Table 5). Both feature sets include the bag of words features and the shared features described in Section 6.13.. The narrative features outperform the bag of words baseline ( $p = 0.0050$  by two-tailed paired  $t$ -test) and the discourse features ( $p = 0.0058$ ), but the difference between the discourse features and the bag of words baseline is not statistically significant.

Features	f-score
All features	71.55
Narrative features	68.18
Discourse features	65.12
Bag of words features	64.42

Table 5: Comparison of narrative and discourse features.

We also compared the performance of individual features by ablation in Table 6. The best feature was the position of the clause in the narrative, which suggests that a sequence-tagging model may be suited to this task. The second best feature was the classes of the implicit relations in which the clause was an argument. This supports our hypothesis that the tasks of detecting narrative structure and detecting implicit discourse relations are related and that improvements in one task may be applicable to the other.

Features	f-score
+Length of VP & narrative, first 3 & last words	71.55
+Explicit relation class	71.05
+Length of clause, dialogue	70.80
+Tense shift, verb form	69.68
+Implicit relation class	68.52
+Position in narrative	67.61
Bag of words features	64.42

Table 6: Comparison of individual features.

In Table 2, we showed that Complicating Actions are more likely to be arguments of *Contingency* relations than of *Expansion* relations and about as likely to be arguments of *Comparison* relations as of *Expansion* relations. This

suggests that successfully detecting Complicating Actions could help in detecting certain classes of implicit discourse relations. For comparison, Biran and McKeown (2013) report the best scores over several state-of-the-art implicit discourse relation detection systems (Table 7): *Expansion* relations are relatively easy to distinguish, while *Contingency* and *Comparison* relations are more difficult.

Relation class	f-score
Comparison	31.79
Contingency	49.82
Expansion	79.22
Temporal	26.57

Table 7: Best scores on implicit discourse relations.

## 8 Future Work

We hope to explore the relationship between narrative structure and discourse by examining whether first identifying a structural element for each clause in a narrative can assist in detecting implicit discourse relations.

To do this, we must extend our system from distinguishing only between Complicating Actions and Others to detecting all of the elements of narrative structure. The main obstacle we face is the small amounts of Abstracts, Resolutions, and Codas available. These structural elements do not occur in every narrative – as Table 1 shows, our corpus contains only 25 Abstracts, 47 Resolutions, and 30 Codas. To address this problem, we need to collect significantly more data. We plan to explore the use of blogs. Blogs are a likely source of personal narrative and are easier for automated tools to handle than is our current corpus of transcribed speech.

After collecting more data, we would need to annotate the data or explore semi-supervised approaches. While it is not difficult to train an annotator, we saw in Section 3 that it can be difficult to assign a single element of narrative structure to a clause – in particular, annotators tend to disagree on which clauses should be considered Evaluations. A possible solution would be to allow clauses to have multiple labels.

We would also like to try reframing the task of detecting narrative structure as a sequence tagging problem. As our experiments showed, information about the preceding clauses helped in labeling the current clause.

## 9 Conclusion

We have described a new task of detecting narrative structure elements based on the work of Labov and Waletzky (1967) and Labov (2013) and presented experiments on detecting one element, the Complicating Action. We achieved 71.55 f-score, performing comparably to state-of-the-art work on detecting implicit discourse relations, and argued that the integration of these two related tasks can benefit both.

The advantage of Labov’s theory of narrative structure over discourse relations is that, while discourse relations

are general across all domains, narrative structure is specific to narrative. As a result, it is well-suited to tasks of narrative analysis and generation.

From the analysis side, narrative structure provides some measure of what it means for a narrative to be well-formed. While Labov’s theory is intended to be descriptive rather than prescriptive, it would certainly be strange for a narrative to have all orientating information at the end. This description of the expected skeleton of a “good” narrative could have applications in assessing the quality of narrative writing. Louis and Nenkova (2013) explored predicting the quality of science journalism articles using features that included the proportion of narrative text in the article and discourse-based measures for how well-written an article was – both are related to this work. Information about narrative structure could also be useful in automatic essay grading or detecting which blog posts or news articles will become popular. If there are two blog posts about the same topic and written by equally well-connected bloggers, and one goes viral while the other does not, is there something about the way the popular post was written that makes it a more effective narrative?

From the generation side, learned models of narrative structure could improve narrative generation systems. A narrative generation system might naively present all orienting information at the beginning of the narrative, whereas a human narrator might save a piece of information for later so as not to give away the ending. For example, in “Jacob Schissel’s story,” which was discussed in Section 2, it comes as a shock to the listener that Schissel is suddenly stabbed – if the Orientation about the knife had been presented at the beginning, the narrative would have significantly less impact. A learned model of narrative structure could help narrative generation systems produce more fluent narratives.

## 10 References

- Biran, O. and McKeown, K. (2013). Aggregated word pair features for implicit discourse relation disambiguation. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*.
- Chambers, N. and Jurafsky, D. (2008). Unsupervised learning of narrative event chains. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*.
- Chambers, N. and Jurafsky, D. (2009). Unsupervised learning of narrative schemas and their participants. 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*.
- Elson, D. and McKeown, K. (2009). A tool for deep semantic encoding of narrative texts. In *Proceedings of the ACL-IJCNLP 2009 Software Demonstrations*.
- Elson, D. and McKeown, K. (2010). Tense and aspect assignment in narrative discourse. In *Proceedings of the Sixth International Conference in Natural Language Generation*.
- Finkel, J. R., Grenager, T., and Manning, C. (2005). Incorporating non-local information into information extrac-

- tion systems by gibbs sampling. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*.
- Klein, D. and Manning, C. D. (2003). Accurate unlexicalized parsing. In *Proceedings of the 41st Meeting of the Association for Computational Linguistics*.
- Labov, W. and Waletzky, J., (1967). *Narrative Analysis*, pages 12–44. University of Washington Press, Seattle, W.A.
- Labov, W. (2013). *The Language of Life and Death*. Cambridge University Press, Cambridge, U.K.
- Louis, A. and Nenkova, A. (2013). What makes writing great? first experiments on article quality prediction in the science journalism domain. *Transactions of Association for Computational Linguistics*.
- McCallum, A., (2002). *MALLET: A Machine Learning for Language Toolkit*.
- Park, J. and Cardie, C. (2012). Improving implicit discourse relation recognition through feature set optimization. In *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*.
- Pitler, E. and Nenkova, A. (2009). Using syntax to disambiguate explicit discourse connectives in text. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*.
- Pitler, E., Louis, A., and Nenkova, A. (2009). Automatic sense prediction for implicit discourse relations in text. In *Proceedings of the Joint Conference of the 47th annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*.
- Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A., and Webber, B. (2008a). The penn discourse treebank 2.0. In *Proceedings of the 6th International Conference on Language Resources and Evaluation*.
- Prasad, R., Miltsakaki, E., Dinesh, N., Lee, A., and Joshi, A., (2008b). *The PDTB 2.0 Annotation Manual*.
- Reidl, M. and Young, R. M. (2010). Narrative planning: Balancing plot and character. *Journal of Artificial Intelligence Research*, 39.
- Zhou, Z., Xu, Y., Niu, Z., Lan, M., Su, J., and Tam, C. L. (2010). Predicting discourse connectives for implicit discourse relation recognition. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*.