Identification of Speakers in Novels

Hua He[†] **Denilson Barbosa** [‡] **Grzegorz Kondrak**[‡]

[†]Department of Computer Science
University of Maryland
huah@cs.umd.edu

[‡]Department of Computing Science
University of Alberta
{denilson,gkondrak}@ualberta.ca

Abstract

Speaker identification is the task of attributing utterances to characters in a literary narrative. It is challenging to automate because the speakers of the majority of utterances are not explicitly identified in novels. In this paper, we present a supervised machine learning approach for the task that incorporates several novel features. The experimental results show that our method is more accurate and general than previous approaches to the problem.

1 Introduction

Novels are important as social communication documents, in which novelists develop the plot by means of discourse between various characters. In spite of a frequently expressed opinion that all novels are simply variations of a certain number of basic plots (Tobias, 2012), every novel has a unique plot (or several plots) and a different set of characters. The interactions among characters, especially in the form of conversations, help the readers construct a mental model of the plot and the changing relationships between characters. Many of the complexities of interpersonal relationships, such as romantic interests, family ties, and rivalries, are conveyed by utterances.

A precondition for understanding the relationship between characters and plot development in a novel is the identification of speakers behind all utterances. However, the majority of utterances are not explicitly tagged with speaker names, as is the case in stage plays and film scripts. In most cases, authors rely instead on the readers' comprehension of the story and of the differences between characters.

Since manual annotation of novels is costly, a system for automatically determining speakers of utterances would facilitate other tasks related to the processing of literary texts. Speaker identification could also be applied on its own, for instance in generating high quality audio books without human lectors, where each character would be identifiable by a distinct way of speaking. In addition, research on spoken language processing for broadcast and multi-party meetings (Salamin et al., 2010; Favre et al., 2009) has demonstrated that the analysis of dialogues is useful for the study of social interactions.

In this paper, we investigate the task of speaker identification in novels. Departing from previous approaches, we develop a general system that can be trained on relatively small annotated data sets, and subsequently applied to other novels for which no annotation is available. Since every novel has its own set of characters, speaker identification cannot be formulated as a straightforward tagging problem with a universal set of fixed tags. Instead, we adopt a ranking approach, which enables our model to be applied to literary texts that are different from the ones it has been trained on.

Our approach is grounded in a variety of features that are easily generalizable across different novels. Rather than attempt to construct complete semantic models of the interactions, we exploit lexical and syntactic clues in the text itself. We propose several novel features, including the speaker alternation pattern, the presence of vocatives in utterances, and unsupervised actor-topic features that associate speakers with utterances on the basis of their content. Experimental evaluation shows that our approach not only outperforms the baseline, but also compares favorably to previous approaches in terms of accuracy and generality, even when tested on novels and authors that are different from those used for training.

The paper is organized as follows. After discussing previous work, and defining the terminology, we present our approach and the features that it is based on. Next, we describe the data, the an-

Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, pages 1312–1320,
Sofia, Bulgaria, August 4-9 2013. ©2013 Association for Computational Linguistics

notation details, and the results of our experimental evaluation. At the end, we discuss an application to extracting a set of family relationships from a novel.

2 Related Work

Previous work on speaker identification includes both rule-based and machine-learning approaches. Glass and Bangay (2007) propose a rule generalization method with a scoring scheme that focuses on the speech verbs. The verbs, such as *said* and *cried*, are extracted from the communication category of WordNet (Miller, 1995). The speech-verb-actor pattern is applied to the utterance, and the speaker is chosen from the available candidates on the basis of a scoring scheme. Sarmento and Nunes (2009) present a similar approach for extracting speech quotes from online news texts. They manually define 19 variations of frequent speaker patterns, and identify a total of 35 candidate speech verbs. The rule-based methods are typically characterized by low coverage, and are too brittle to be reliably applied to different domains and changing styles.

Elson and McKeown (2010) (henceforth referred to as EM2010) apply the supervised machine learning paradigm to a corpus of utterances extracted from novels. They construct a single feature vector for each pair of an utterance and a speaker candidate, and experiment with various WEKA classifiers and score-combination methods. To identify the speaker of a given utterance, they assume that all previous utterances are already correctly assigned to their speakers. Our approach differs in considering the utterances in a sequence, rather than independently from each other, and in removing the unrealistic assumption that the previous utterances are correctly identified.

The speaker identification task has also been investigated in other domains. Bethard et al. (2004) identify opinion holders by using semantic parsing techniques with additional linguistic features. Pouliquen et al. (2007) aim at detecting direct speech quotations in multilingual news. Krestel et al. (2008) automatically tag speech sentences in newspaper articles. Finally, Ruppenhofer et al. (2010) implement a rule-based system to enrich German cabinet protocols with automatic speaker attribution.

3 Definitions and Conventions

In this section, we introduce the terminology used in the remainder of the paper. Our definitions are different from those of EM2010 partly because we developed our method independently, and partly because we disagree with some of their choices. The examples are from Jane Austen's *Pride and Prejudice*, which was the source of our development set.

An *utterance* is a connected text that can be attributed to a single speaker. Our task is to associate each utterance with a single speaker. Utterances that are attributable to more than one speaker are rare; in such cases, we accept correctly identifying one of the speakers as sufficient. In some cases, an utterance may include more than one quotation-delimited sequence of words, as in the following example.

> *"Miss Bingley told me," said Jane, "that he never speaks much."*

In this case, the words *said Jane* are simply a speaker tag inserted into the middle of the quoted sentence. Unlike EM2010, we consider this a single utterance, rather than two separate ones.

We assume that all utterances within a paragraph can be attributed to a single speaker. This "one speaker per paragraph" property is rarely violated in novels — we identified only five such cases in *Pride & Prejudice*, usually involving one character citing another, or characters reading letters containing quotations. We consider this an acceptable simplification, much like assigning a single part of speech to each word in a corpus. We further assume that each utterance is contained within a single paragraph. Exceptions to this rule can be easily identified and resolved by detecting quotation marks and other typographical conventions.

The paragraphs without any quotations are referred to as *narratives*. The term *dialogue* denotes a series of utterances together with related narratives, which provide the context of conversations. We define a dialogue as a series of utterances and intervening narratives, with no more than three continuous narratives. The rationale here is that more than three narratives without any utterances are likely to signal the end of a particular dialogue.

We distinguish three types of utterances, which are listed with examples in Table 1: *explicit speaker* (identified by name within the paragraph),

Category	Example
Implicit speaker	*“Don’t keep coughing so, Kitty, for heaven’s sake!”*
Explicit speaker	*“I do not cough for my own amusement,” replied* **Kitty**.
Anaphoric speaker	*“Kitty has no discretion in her coughs,” said* **her father**.

Table 1: Three types of utterances.

anaphoric speaker (identified by an anaphoric expression), and *implicit speaker* (no speaker information within the paragraph). Typically, the majority of utterances belong to the implicit-speaker category. In *Pride & Prejudice* only roughly 25% of the utterances have explicit speakers, and an even smaller 15% belong to the anaphoric-speaker category. In modern fiction, the percentage of explicit attributions is even lower.

4 Speaker Identification

In this section, we describe our method of extracting explicit speakers, and our ranking approach, which is designed to capture the speaker alternation pattern.

4.1 Extracting Speakers

We extract explicit speakers by focusing on the speech verbs that appear before, after, or between quotations. The following verbs cover most cases in our development data: *say, speak, talk, ask, reply, answer, add, continue, go on, cry, sigh,* and *think*. If a verb from the above short list cannot be found, any verb that is preceded by a name or a personal pronoun in the vicinity of the utterance is selected as the speech verb.

In order to locate the speaker's name or anaphoric expression, we apply a deterministic method based on syntactic rules. First, all paragraphs that include narrations are parsed with a dependency parser. For example, consider the following paragraph:

> *As they went downstairs together, Charlotte said, “I shall depend on hearing from you very often, Eliza.”*

The parser identifies a number of dependency relations in the text, such as *dobj(went-3, downstairs-4)* and *advmod(went-3, together-5)*. Our method extracts the speaker's name from the dependency relation *nsubj(said-8, Charlotte-7)*, which links a speech verb with a noun phrase that is the syntactic subject of a clause.

Once an explicit speaker's name or an anaphoric expression is located, we determine the corresponding gender information by referring to the character list or by following straightforward rules to handle the anaphora. For example, if the utterance is followed by the phrase *she said*, we infer that the gender of the speaker is female.

4.2 Ranking Model

In spite of the highly sequential nature of the chains of utterances, the speaker identification task is difficult to model as sequential prediction. The principal problem is that, unlike in many NLP problems, a general fixed tag set cannot be defined beyond the level of an individual novel. Since we aim at a system that could be applied to any novel with minimal pre-processing, sequential prediction algorithms such as Conditional Random Fields are not directly applicable.

We propose a more flexible approach that assigns scores to candidate speakers for each utterance. Although the sequential information is not directly modeled with tags, our system is able to indirectly utilize the speaker alternation pattern using the method described in the following section. We implement our approach with *SVM-rank* (Joachims, 2006).

4.3 Speaker Alternation Pattern

The speaker alternation pattern is often employed by authors in dialogues between two characters. After the speakers are identified explicitly at the beginning of a dialogue, the remaining odd-numbered and even-numbered utterances are attributable to the first and second speaker, respectively. If one of the speakers “misses their turn”, a clue is provided in the text to reset the pattern.

Based on the speaker alternation pattern, we make the following two observations:

1. The speakers of consecutive utterances are usually different.
2. The speaker of the n-th utterance in a dialogue is likely to be the same as the speaker of the $(n - 2)$-th utterance.

Our ranking model incorporates the speaker alternation pattern by utilizing a feature expansion scheme. For each utterance n, we first generate its own features (described in Section 5), and

Features	Novelty
Distance to Utterance	No
Speaker Appearance Count	No
Speaker Name in Utterance	No
Unsupervised Actor-Topic Model	Yes
Vocative Speaker Name	Yes
Neighboring Utterances	Yes
Gender Matching	Yes
Presence Matching	Yes

Table 2: Principal feature sets.

subsequently we add three more feature sets that represent the following neighboring utterances: $n-2, n-1$ and $n+1$. Informally, the features of the utterances $n-1$ and $n+1$ encode the first observation, while the features representing the utterance $n-2$ encode the second observation. In addition, we include a set of four binary features that are set for the utterances in the range $[n-2, n+1]$ if the corresponding explicit speaker matches the candidate speaker of the current utterance.

5 Features

In this section, we describe the set of features used in our ranking approach. The principal feature sets are listed in Table 2, together with an indication whether they are novel or have been used in previous work.

5.1 Basic Features

A subset of our features correspond to the features that were proposed by EM2010. These are mostly features related to speaker names. For example, since names of speakers are often mentioned in the vicinity of their utterances, we count the number of words separating the utterance and a name mention. However, unlike EM2010, we consider only the two nearest characters in each direction, to reflect the observation that speakers tend to be mentioned by name immediately before or after their corresponding utterances. Another feature is used to represent the number of appearances for speaker candidates. This feature reflects the relative importance of a given character in the novel. Finally, we use a feature to indicate the presence or absence of a candidate speaker's name within the utterance. The intuition is that speakers are unlikely to mention their own name.

Feature	Example
start of utterance	*"Kitty . . .*
before period	*. . . Jane.*
between commas	*. . . , Elizabeth, . . .*
between comma & period	*. . . , Mrs. Hurst.*
before exclamation mark	*. . . Mrs. Bennet!*
before question mark	*. . . Lizzy?. . .*
vocative phrase	*Dear . . .*
after vocative phrase	*Oh! Lydia . . .*
2nd person pronoun	*. . . you . . .*

Table 3: Features for the vocative identification.

5.2 Vocatives

We propose a novel vocative feature, which encodes the character that is explicitly addressed in an utterance. For example, consider the following utterance:

> "*I hope Mr. Bingley will like it, Lizzy.*"

Intuitively, the speaker of the utterance is neither *Mr. Bingley* nor *Lizzy*; however, the speaker of the next utterance is likely to be *Lizzy*. We aim at capturing this intuition by identifying the addressee of the utterance.

We manually annotated vocatives in about 900 utterances from the training set. About 25% of the names within utterance were tagged as vocatives. A Logistic Regression classifier (Agresti, 2006) was trained to identify the vocatives. The classifier features are shown in Table 3. The features are designed to capture punctuation context, as well as the presence of typical phrases that accompany vocatives. We also incorporate interjections like "oh!" and fixed phrases like "my dear", which are strong indicators of vocatives. Under 10-fold cross validation, the model achieved an F-measure of 93.5% on the training set.

We incorporate vocatives in our speaker identification system by means of three binary features that correspond to the utterances $n-1$, $n-2$, and $n-3$. The features are set if the detected vocative matches the candidate speaker of the current utterance n.

5.3 Matching Features

We incorporate two binary features for indicating the gender and the presence of a candidate speaker. The gender matching feature encodes the gender agreement between a speaker candidate and the speaker of the current utterance. The gender information extraction is applied to two utterance groups: the anaphoric-speaker utterances, and the explicit-speaker utterances. We use the technique described in Section 4.1 to determine the gender of a speaker of the current utterance. In contrast with EM2010, this is not a hard constraint.

The presence matching feature indicates whether a speaker candidate is a likely participant in a dialogue. Each dialogue consists of continuous utterance paragraphs together with neighboring narration paragraphs as defined in Section 3. The feature is set for a given character if its name or alias appears within the dialogue.

5.4 Unsupervised Actor-Topic Features

The final set of features is generated by the unsupervised actor-topic model (ACTM) (Celikyilmaz et al., 2010), which requires no annotated training data. The ACTM, as shown in Figure 1, extends the work of author-topic model in (Rosen-Zvi et al., 2010). It can model dialogues in a literary text, which take place between two or more speakers conversing on different topics, as distributions over topics, which are also mixtures of the term distributions associated with multiple speakers. This follows the linguistic intuition that rich contextual information can be useful in understanding dialogues.

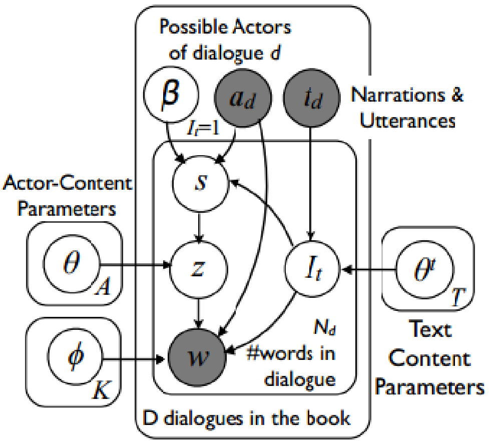
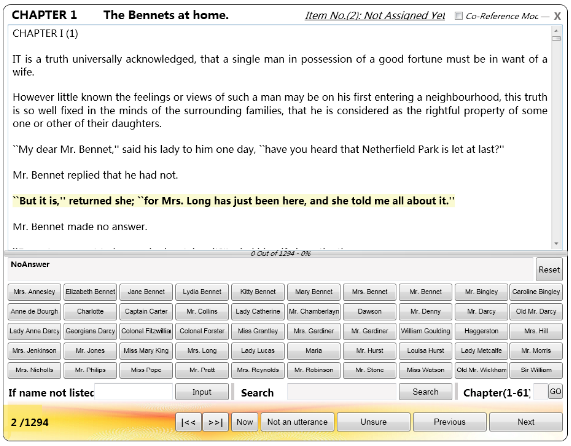


Figure 1: Graphical Representation of ACTM.

The ACTM predicts the most likely speakers of a given utterance by considering the content of an utterance and its surrounding contexts. The Actor-Topic-Term probabilities are calculated by using both the relationship of utterances and the surrounding textual clues. In our system, we utilize four binary features that correspond to the four top ranking positions from the ACTM model.



CHAPTER I (1)

IT is a truth universally acknowledged, that a single man in possession of a good fortune must be in want of a wife.

However little known the feelings or views of such a man may be on his first entering a neighbourhood, this truth is so well fixed in the minds of the surrounding families, that he is considered as the rightful property of some one or other of their daughters.

``My dear Mr. Bennet,'' said his lady to him one day, ``have you heard that Netherfield Park is let at last?''

Mr. Bennet replied that he had not.

``But it is,'' returned she; ``for Mrs. Long has just been here, and she told me all about it.''

Mr. Bennet made no answer.

Figure 2: Annotation Tool GUI.

6 Data

Our principal data set is derived from the text of *Pride and Prejudice*, with chapters 19–26 as the test set, chapters 27–33 as the development set, and the remaining 46 chapters as the training set. In order to ensure high-quality speaker annotations, we developed a graphical interface (Figure 2), which displays the current utterance in context, and a list of characters in the novel. After the speaker is selected by clicking a button, the text is scrolled automatically, with the next utterance highlighted in yellow. The complete novel was annotated by a student of English literature. The annotations are publicly available[1].

For the purpose of a generalization experiment, we also utilize a corpus of utterances from the 19th and 20th century English novels compiled by EM2010. The corpus differs from our data set in three aspects. First, as discussed in Section 3, we treat all quoted text within a single paragraph as a single utterance, which reduces the total number of utterances, and results in a more realistic reporting of accuracy. Second, our data set includes annotations for all utterances in the novel, as opposed to only a subset of utterances from several novels, which are not necessarily contiguous. Lastly, our annotations come from a single expert, while the annotations in the EM2010 corpus were collected through Amazon's Mechanical Turk, and filtered by voting. For example, out of 308 utterances from *The Steppe*, 244 are in fact annotated, which raises the question whether the discarded utterances tend to be more difficult to annotate.

Table 4 shows the number of utterances in all

[1] `www.cs.ualberta.ca/~kondrak/austen`

	IS	AS	ES	Total
Pride & P. (all)	663	292	305	1260
Pride & P. (test)	65	29	32	126
Emma	236	55	106	397
The Steppe	93	39	112	244

Table 4: The number of utterances in various data sets by the type (IS - Implicit Speaker; AS - Anaphoric Speaker; ES - Explicit Speaker).

	Pride & P.	*Emma*	*Steppe*
BASELINE	42.0	44.1	66.8
INDIVIDUAL	77.8	67.3	74.2
NEIGHBORS	82.5	74.8	80.3
ORACLE	86.5	80.1	83.6

Table 5: Speaker identification accuracy (in %) on *Pride & Prejudice*, *Emma*, and *The Steppe*.

data sets. We selected Jane Austen's *Emma* as a different novel by the same author, and Anton Chekhov's *The Steppe* as a novel by a different author for our generalization experiments.

Since our goal is to match utterances to characters rather than to name mentions, a preprocessing step is performed to produce a list of characters in the novel and their aliases. For example, *Elizabeth Bennet* may be referred to as *Liz, Lizzy, Miss Lizzy, Miss Bennet, Miss Eliza,* and *Miss Elizabeth Bennet.* We apply a name entity tagger, and then group the names into sets of character aliases, together with their gender information. The sets of aliases are typically small, except for major characters, and can be compiled with the help of web resources, such as Wikipedia, or study guides, such as CliffsNotes^{TM}. This preprocessing step could also be performed automatically using a canonicalization method (Andrews et al., 2012); however, since our focus is on speaker identification, we decided to avoid introducing annotation errors at this stage.

Other preprocessing steps that are required for processing a new novel include standarizing the typographical conventions, and performing POS tagging, NER tagging, and dependency parsing. We utilize the Stanford tools (Toutanova et al., 2003; Finkel et al., 2005; Marneffe et al., 2006).

7 Evaluation

In this section, we describe experiments conducted to evaluate our speaker identification approach. We refer to our main model as NEIGHBORS, because it incorporates features from the neighboring utterances, as described in Section 4.3. In contrast, the INDIVIDUAL model relies only on features from the current utterance. In an attempt to reproduce the evaluation methodology of EM2010, we also test the ORACLE model, which has access to the gold-standard information about the speakers of eight neighboring utterances in the range $[n-4, n+4]$. Lastly, the BASELINE approach selects the name that is the closest in the narration, which is more accurate than the "most recent name" baseline.

7.1 Results

Table 5 shows the results of the models trained on annotated utterances from *Pride & Prejudice* on three test sets. As expected, the accuracy of all learning models on the test set that comes from the same novel is higher than on unseen novels. However, in both cases, the drop in accuracy for the NEIGHBORS model is less than 10%.

Surprisingly, the accuracy is higher on *The Steppe* than on *Emma*, even though the different writing style of Chekhov should make the task more difficult for models trained on Austen's prose. The protagonists of *The Steppe* are mostly male, and the few female characters rarely speak in the novel. This renders our gender feature virtually useless, and results in lower accuracy on anaphoric speakers than on explicit speakers. On the other hand, Chekhov prefers to mention speaker names in the dialogues (46% of utterances are in the explicit-speaker category), which makes his prose slightly easier in terms of speaker identification.

The relative order of the models is the same on all three test sets, with the NEIGHBORS model consistently outperforming the INDIVIDUAL model, which indicates the importance of capturing the speaker alternation pattern. The performance of the NEIGHBORS model is actually closer to the ORACLE model than to the INDIVIDUAL model.

Table 6 shows the results on *Emma* broken down according to the type of the utterance. Unsurprisingly, the explicit speaker is the easiest category, with nearly perfect accuracy. Both the INDIVIDUAL and the NEIGHBORS models do better on anaphoric speakers than on implicit speakers, which is also expected. However, it is not the

	IS	AS	ES	Total
INDIVIDUAL	52.5	67.3	100.0	67.3
NEIGHBORS	63.1	76.4	100.0	74.8
ORACLE	74.2	69.1	99.1	80.1

Table 6: Speaker identification accuracy (in %) on Austen's *Emma* by the type of utterance.

case for the ORACLE model. We conjecture that the ORACLE model relies heavily on the neighborhood features (which are rarely wrong), and consequently tends to downplay the gender information, which is the only information extracted from the anaphora. In addition, anaphoric speaker is the least frequent of the three categories.

Table 7 shows the results of an ablation study performed to investigate the relative importance of features. The INDIVIDUAL model serves as the base model from which we remove specific features. All tested features appear to contribute to the overall performance, with the distance features and the unsupervised actor-topic features having the most pronounced impact. We conclude that the incorporation of the neighboring features, which is responsible for the difference between the INDIVIDUAL and NEIGHBORS models, is similar in terms of importance to our strongest textual features.

Feature	Impact
Closest Mention	-6.3
Unsupervised ACTM	-5.6
Name within Utterance	-4.8
Vocative	-2.4

Table 7: Results of feature ablation (in % accuracy) on *Pride & Prejudice*.

7.2 Comparison to EM2010

In this section we analyze in more detail our results on *Emma* and *The Steppe* against the published results of the state-of-the-art EM2010 system. Recall that both novels form a part of the corpus that was created by EM2010 for the development of their system.

Direct comparison to EM2010 is difficult because they compute the accuracy separately for seven different categories of utterances. For each category, they experiment with all combinations of three different classifiers and four score combination methods, and report only the accuracy achieved by the best performing combination on that category. In addition, they utilize the ground truth speaker information of the preceding utterances. Therefore, their results are best compared against our ORACLE approach.

Character

id	*name*	*gender*
	. . .	
9	Mr. Collins	m
10	Charlotte	f
11	Jane Bennet	f
12	Elizabeth Bennet	f
	. . .	

Relation

from	*to*	*type*	*mode*
	. . .		
10	9	husband	explicit
9	10	wife	derived
10	12	friend	explicit
12	10	friend	derived
11	12	sister	explicit
	. . .		

Figure 3: Relational database with extracted social network.

Unfortunately, EM2010 do not break down their results by novel. They report the overall accuracy of 63% on both "anaphora trigram" (our *anaphoric speaker*), and "quote alone" (similar to our *implicit speaker*). If we combine the two categories, the numbers corresponding to our NEIGHBORS model are 65.6% on *Emma* and 64.4% on *The Steppe*, while ORACLE achieves 73.2% and 70.5%, respectively. Even though a direct comparison is not feasible, the numbers are remarkable considering the context of the experiment, which strongly favors the EM2010 system.

8 Extracting Family Relationships

In this section, we describe an application of the speaker identification system to the extraction of family relationships. Elson et al. (2010) extract unlabeled networks where the nodes represent characters and edges indicate their *proximity*, as indicated by their interactions. Our goal is to construct networks in which edges are labeled by the mutual relationships between characters in a novel. We focus on family relationships, but also include social relationships, such as *friend*

```
INSERT INTO Relation (id1, id2, t, m)
     SELECT r.to AS id1, r.from AS id2 , 'wife' AS t, 'derived' AS m
     FROM Relation r
     WHERE r.type='husband' AND r.mode='explicit' AND
           NOT EXISTS(SELECT * FROM Relation r2
                      WHERE r2.from=r.to AND r2.to=r.from AND r2.type=t)
```

Figure 4: An example inference rule.

and *attracted-to*.

Our approach to building a social network from the novel is to build an active database of relationships explicitly mentioned in the text, which is expanded by triggering the execution of queries that deduce implicit relations. This inference process is repeated for every discovered relationship until no new knowledge can be inferred.

The following example illustrates how speaker identification helps in the extraction of social relations among characters. Consider, the following conversation:

> *"How so? how can it affect them?"*
> *"My dear Mr. Bennet," replied* ***his wife****,*
> *"how can you be so tiresome!"*

If the speakers are correctly identified, the utterances are attributed to *Mr. Bennet* and *Mrs. Bennet*, respectively. Furthermore, the second utterance implies that its speaker is the wife of the preceding speaker. This is an example of an explicit relationship which is included in our database. Several similar extraction rules are used to extract explicit mentions indicating family and affective relations, including *mother*, *nephew*, and *fiancee*. We can also derive relationships that are not explicitly mentioned in the text; for example, that *Mr. Bennet* is the husband of *Mrs. Bennet*.

Figure 3 shows a snippet of the relational database of the network extracted from *Pride & Prejudice*. Table *Character* contains all characters in the book, each with a unique identifier and gender information, while Table *Relation* contains all relationships that are explicitly mentioned in the text or derived through reasoning.

Figure 4 shows an example of an inference rule used in our system. The rule derives a new relationship indicating that character c_1 is the *wife* of character c_2 if it is known (through an explicit mention in the text) that c_2 is the *husband* of c_1. One condition for the rule to be applied is that the database must not already contain a record indicating the wife relationship. This inference rule would derive the tuple in Figure 3 indicating that the wife or Mr. Collins is Charlotte.

In our experiment with *Pride & Prejudice*, a total of 55 explicitly indicated relationships were automatically identified once the utterances were attributed to the characters. From those, another 57 implicit relationships were derived through inference. A preliminary manual inspection of the set of relations extracted by this method (Makazhanov et al., 2012) indicates that all of them are correct, and include about 40% all personal relations that can be inferred by a human reader from the text of the novel.

9 Conclusion and Future Work

We have presented a novel approach to identifying speakers of utterances in novels. Our system incorporates a variety of novel features which utilize vocatives, unsupervised actor-topic models, and the speaker alternation pattern. The results of our evaluation experiments indicate a substantial improvement over the current state of the art.

There are several interesting directions for the future work. Although the approach introduced in this paper appears to be sufficiently general to handle novels written in a different style and period, more sophisticated statistical graphical models may achieve higher accuracy on this task. A reliable automatic generation of characters and their aliases would remove the need for the preprocessing step outlined in Section 6. The extraction of social networks in novels that we discussed in Section 8 would benefit from the introduction of additional inference rules, and could be extended to capture more subtle notions of sentiment or relationship among characters, as well as their development over time.

We have demonstrated that speaker identification can help extract family relationships, but the converse is also true. Consider the following utterance:

> *"Lizzy," said her father, "I have given*
> *him my consent."*

In order to deduce the speaker of the utterance, we need to combine the three pieces of information: (a) the utterance is addressed to Lizzy (vocative prediction), (b) the utterance is produced by Lizzy's father (pronoun resolution), and (c) Mr. Bennet is the father of Lizzy (relationship extraction). Similarly, in the task of compiling a list of characters, which involves resolving aliases such as *Caroline*, *Caroline Bingley*, and *Miss Bingley*, simultaneous extraction of family relationships would help detect the ambiguity of *Miss Benett*, which can refer to any of several sisters. A joint approach to resolving speaker attribution, relationship extraction, co-reference resolution, and alias-to-character mapping would not only improve the accuracy on all these tasks, but also represent a step towards deeper understanding of complex plots and stories.

Acknowledgments

We would like to thank Asli Celikyilmaz for collaboration in the early stages of this project, Susan Brown and Michelle Di Cintio for help with data annotation, and David Elson for the attempt to compute the accuracy of the EM2010 system on *Pride & Prejudice*. This research was partially supported by the Natural Sciences and Engineering Research Council of Canada.

References

Alan Agresti. 2006. Building and applying logistic regression models. In *An Introduction to Categorical Data Analysis*. John Wiley & Sons, Inc.

Nicholas Andrews, Jason Eisner, and Mark Dredze. 2012. Name phylogeny: A generative model of string variation. In *EMNLP-CoNLL*.

Steven Bethard, Hong Yu, Ashley Thornton, Vasileios Hatzivassiloglou, and Dan Jurafsky. 2004. Automatic extraction of opinion propositions and their holders. In *AAAI Spring Symposium on Exploring Attitude and Affect in Text*.

Asli Celikyilmaz, Dilek Hakkani-Tur, Hua He, Grzegorz Kondrak, and Denilson Barbosa. 2010. The actor-topic model for extracting social networks in literary narrative. In *Proceedings of the NIPS 2010 Workshop - Machine Learning for Social Computing*.

David K. Elson and Kathleen McKeown. 2010. Automatic attribution of quoted speech in literary narrative. In *AAAI*.

David K. Elson, Nicholas Dames, and Kathleen McKeown. 2010. Extracting social networks from literary fiction. In *ACL*.

Sarah Favre, Alfred Dielmann, and Alessandro Vinciarelli. 2009. Automatic role recognition in multiparty recordings using social networks and probabilistic sequential models. In *ACM Multimedia*.

Jenny Rose Finkel, Trond Grenager, and Christopher Manning. 2005. Incorporating non-local information into information extraction systems by gibbs sampling. In *ACL*.

Kevin Glass and Shaun Bangay. 2007. A naive salience-based method for speaker identification in fiction books. In *Proceedings of the 18th Annual Symposium of the Pattern Recognition*.

Thorsten Joachims. 2006. Training linear SVMs in linear time. In *KDD*.

Ralf Krestel, Sabine Bergler, and René Witte. 2008. Minding the source: Automatic tagging of reported speech in newspaper articles. In *LREC*.

Aibek Makazhanov, Denilson Barbosa, and Grzegorz Kondrak. 2012. Extracting family relations from literary fiction. Unpublished manuscript.

Marie Catherine De Marneffe, Bill Maccartney, and Christopher D. Manning. 2006. Generating typed dependency parses from phrase structure parses. In *LREC*.

George A. Miller. 1995. Wordnet: A lexical database for english. *Communications of the ACM*, 38:39–41.

Bruno Pouliquen, Ralf Steinberger, and Clive Best. 2007. Automatic detection of quotations in multilingual news. In *RANLP*.

Michal Rosen-Zvi, Chaitanya Chemudugunta, Thomas L. Griffiths, Padhraic Smyth, and Mark Steyvers. 2010. Learning author-topic models from text corpora. *ACM Trans. Inf. Syst.*, 28(1).

Josef Ruppenhofer, Caroline Sporleder, and Fabian Shirokov. 2010. Speaker attribution in cabinet protocols. In *LREC*.

Hugues Salamin, Alessandro Vinciarelli, Khiet Truong, and Gelareh Mohammadi. 2010. Automatic role recognition based on conversational and prosodic behaviour. In *ACM Multimedia*.

Luis Sarmento and Sergio Nunes. 2009. Automatic extraction of quotes and topics from news feeds. In *4th Doctoral Symposium on Informatics Engineering*.

Ronald B. Tobias. 2012. *20 Master Plots: And How to Build Them*. Writer's Digest Books, 3rd edition.

Kristina Toutanova, Dan Klein, Christopher D. Manning, and Yoram Singer. 2003. Feature-rich part-of-speech tagging with a cyclic dependency network. In *NAACL-HLT*.