The Words Themselves:
A Content-Based Approach to Quote Attribution

Quote attribution is the identification of the speaker of a quotation in a given text. It requires reasoning about conversational patterns and contextual clues, and is especially complex in literary texts. In the analysis of novels, which is our focus here, accurate quote attribution is a prerequisite for studies that seek insight into a given author's style, including their ability to create "dialogic" novels with meaningfully differentiated character voices. Failure to consider novels as composed of the differentiated voices of characters and narrators — and instead to regard them as the uniform expression of their author's style — had led to disputed results, most famously in the work of Jockers (2010; c.f. Hammond 2017).

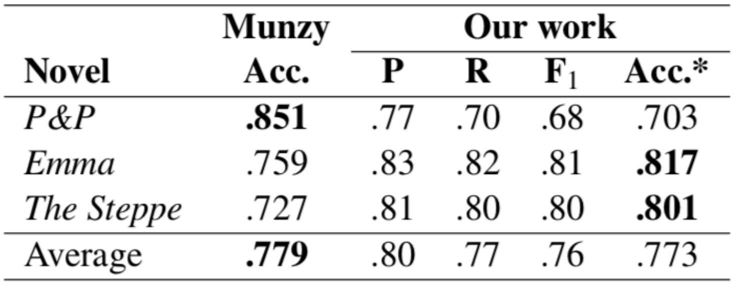
Our work improves on previous approaches by considering both contextual information (the words around a given quotation) and the content of the quotation itself (the words spoken by the character). Consider the following example from the opening of Jane Austen's Pride and Prejudice:

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

The contextual information in this example is famously difficult. Identifying the speaker first requires identifying that the addressee is Mr. Bennet, and that the speaker is "his lady." We then must infer that "his lady" refers to "his wife", where "his" refers to the addressee, Mr. Bennet. The speaker is therefore Mrs. Bennet, a character whose name is not mentioned until several paragraphs later. Most previous approaches to quote attribution fall into one of two camps, both of which rely on contextual information: rule-based systems (Glass, 2007; Sarmento, 2009) and machine learning systems (Elson, 2010; 2012, O'Keefe et al., 2012). Very few systems, however, consider the actual content of the quotation. In instances in which the contextual cues are very clear — if the text above had included a phrase such as "said Mrs. Bennet" — our system relies upon them. However, in cases like that above, where the cues are ambiguous or vague, we look at the style and content of the quotation itself (distinctive lexical choices like "let", syntax like "have you heard", or topics such as real estate) to assist in guessing who is most likely to be speaking.

For our experiment, we worked with the QuoteLi corpus provided by Muzny et al. (2017); despite consisting of only three annotated novels, it is the largest and most accurate extant corpus of quote-attributed novels. We first build our seed training set by extracting high-confidence quotations, referring to contextual information and using simple trigram matching as employed by Muzny et al. (2017), Elson and McKeown (2010), and O'Keefe et al. (2011). In order to avoid the long tail of minor characters affecting the performance of our classifier, we restrict our experiments to those characters with at least 15 attributed quotes in the seed set. Once our seed training set is extracted, we follow the semi-supervised self-training procedure of Yarowsky (1995) to classify the rest of the quotations. In each iteration, quotations that are classified with a confidence score above a certain threshold are added to the training set, and the remaining form the test set for the next iteration of classification.

For the classification itself, we use features based on weights from the Sparse Additive Generative model of text (SAGE) introduced by Eisensetein et al. (2011), along with GloVe word embeddings (Pennington et al., 2014). SAGE is a generative model of text that models a datapoint by estimating the log deviations of its word frequencies from a background lexical distribution. Thus, for each quotation, we obtain the weighting coefficients by finding its SAGE coefficients with respect to the entire data distribution. Our word vectors are pre-trained 300-dimensional GloVe embeddings. A SAGE-weighted average of the embeddings of the words in each quotation results in a 300-dimensional vector representation for each datapoint, which is then passed through a Maximum Entropy classifier.



	Munzy	**Our work**			
Novel	**Acc.**	**P**	**R**	**F_1**	**Acc.***
P&P	**.851**	.77	.70	.68	.703
Emma	.759	.83	.82	.81	**.817**
The Steppe	.727	.81	.80	.80	**.801**
Average	**.779**	.80	.77	.76	.773

Table 1: Precision (P), Recall (R), F1 score and accuracy scores of our system. The accuracy scores reported by Muzny et al. (2017) on the complete corpus are shown for comparison. Best accuracy on each novel is shown in boldface. *Evaluated only on a subset of the complete dataset.

Table 1 presents our classification results for the three novels considered in our experiment. Our method achieves an average accuracy almost exactly equal to that of the current state-of-the-art (Muzny et al., 2017) which relies heavily on contextual information. Our system performs particularly well on implicit and anaphoric quotations, on which context-based systems perform poorly. This shows that the distinctiveness of character dialogues in these texts is by itself a strong indicator of speaker identity, even with our relatively simple, lexical unigram-based definition of style. Conversely, our system performs poorly on some major characters. The blame for this may lie in part with the authors of individual texts: stylistic distinctiveness of characters is the consequence of conscious effort on the author's part, and our system is reliant on this.

Given the importance of quote attribution to computational literary research, we believe that our content- and style-based approach is highly promising and merits further investigation. The major barrier to further research is the lack of reliable corpora. We are currently engaged in an effort to build such corpora ourselves; we have secured funding to develop annotation software and hire a team of graduate research assistants to annotate a set of representative texts. We hope that our presentation at DH will inspire others to join us in this important annotation project.

References

1. Jacob Eisenstein, Amr Ahmed, and Eric P Xing. 2011. Sparse additive generative models of text. In Proceedings of the 28th International Conference on International Conference on Machine Learning, pages 1041–1048. Omnipress.

2. David K Elson and Kathleen R McKeown. 2010. Automatic attribution of quoted speech in literary narrative. In Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, pages 1013–1019. AAAI Press.

3. Adam Hammond. 2017. The Double Bind of Validation: Distant Reading and the Digital Humanities' "Trough of Disillusionment." Literature Compass, volume 14, pages 1–13.

4. Matthew Jockers. 2013. Macroanalysis: Digital Methods and Literary History. Urbana-Champaign: University of Illinois Press.

5. Tim O'Keefe, Silvia Pareti, James R Curran, Irena Koprinska, and Matthew Honnibal. 2012. A sequence labelling approach to quote attribution. In Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pages 790–799. Association for Computational Linguistics.

6. Grace Muzny, Michael Fang, Angel Chang, and Dan Jurafsky. 2017. A two-stage sieve approach for quote attribution. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, volume 1, pages 460–470.

7. Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* pages 1532–1543.

8. David Yarowsky. 1995. Unsupervised word sense disambiguation rivaling supervised methods. In Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics, pages 189–196, Cambridge, MA.