You can be Shakespeare!
A Case Study in Paraphrase Targeting Writing Styles

*Author*1[1,2] *Author*2[1,3]
(1) INSTITUTE_1, address 1
(2) INSTITUTE_2, address 2
(3) INSTITUTE_3, address 3
author1@institute1, author@institute2

ABSTRACT
We present initial investigation into the task of paraphrasing language while targeting a particular writing style. The plays of William Shakespeare and their modern translations are used as a testbed for evaluating paraphrase systems targeting a specific style of writing. We empirically show that even with a relatively small amount of parallel training data available, it is possible to learn paraphrase models which capture stylistic phenomenon, and these models outperform baselines based on dictionaries and out-of-domain parallel text. In addition we present an initial investigation into automatic evaluation metrics for paraphrasing writing style. To the best of our knowledge this is the first work to investigate the task of paraphrasing text with the goal of targeting a specific style of writing.

KEYWORDS: Paraphrase, Writing Style.

1 Introduction

Identical meaning can be expressed or *paraphrased* in many different ways; automatically detecting or generating different expressions with the same meaning is fundamental to many natural language understanding tasks(Giampiccolo et al., 2007), so much previous work has investigated methods for automatic paraphrasing(Barzilay and Lee, 2003; Dolan et al., 2004; Shinyama and Sekine, 2003; Das and Smith, 2009; Bannard and Callison-Burch, 2005). Although two utternaces may be semantically equivelant, they can still be stylistically quite different. For example, the same information information about a new product is likely to be conveyed using very different lexical and grammatical patterns in advertising materials v.s. technical manuals, or in Shakespearean plays v.s. Hollywood movies.

In this paper, we investigate the task of automatic paraphrasing when targeting a specific writing style, focusing specifically on the style of Early Modern English employed by William Shakespeare. We exploit modern translations of 17 plays written to help literature students to better understand Shakespeare. These modern translations are used to generate a parallel corpus of Shakespere's style and modern English, which is then used to train phrase-based translation models which are capable of automatically paraphrasing ordinary sentences into Shakesperean English. In addition we develop several baseline systems which don't make use of this source of parallel text and instead rely on dictionaries of expressions commonly found in shakesperean english, or parallel monolingual text gathered through Amazon's Mechanical Turk (Chen and Dolan, 2011).

We evaluate these models both through human judgements and standard evaluation metrics

corpus	initial size	aligned size	No-Change BLEU
http://nfs.sparknotes.com	31,718	21,079	23.67
http://enotes.com	13,640	10,365	49.60

Table 1: Parallel corpora generated form modern translations of Shakespeare's plays

from the Machine Translation and paraphrase literature, however no previous work has investigated the ability of these automatic metrics to capture the notion of writing style. We propose several new metrics for evaluating the task of paraphrasing while targeting a specific style, and measure correlation with human judgement showing promising results for this particular style.

Systems which are capable of automatically paraphrasing literary writing styles could be directly benefical for educational applications, for example helping students to experiment with writing literature in the style of authors they are studying. Additionally note that out of the 37 surviving plays written by William Shakespeare, only 17 currently have modern translations available; although we have not yet formally evaluated paraphrasing in the other direction, this work also has the potential to make the other 20 plays more accessable to students of Shakespeare.

2 Data

We propose to use Shakespeare's plays and their modern English translations as a testbed the task of paraphrasing targeting a specific writing style. Having access to parallel text in the target style allows us to train statistical models for generating paraphrases, and also perform automatic evaluation using BLEU which requires access to a set of reference translations. For this purpose we scraped modern translations of 17 Shakespeare plays from `http://nfs.sparknotes.com`, and an additional 8 translations of overlapping plays from `http://enotes.com`, giving us two reference translations for 8 out of the 17 plays.

After tokenizing and lowercasing, the plays and their modern translations were aligned using Bob Moore's bilingual sentence (Moore, 2002) aligner producing about 21,079 alignments out of 31,718 sentences in the Sparknotes data, and 10,365 aligned sentence pairs out of 13,640 sentences in the enotes data. In addition, we note that the modern translations from each source are qualitatively quite different. The Sparknotes paraphrases tend to differ more significantly from the original texts, whereas the enotes data tends to stick closer to the original text, being much more conservative in it's use of paraphrase, although it does include many useful paraphrses which do tend to be qualitatively different from those used in the Sparknotes data. To illustrate these differences empircally, we evaluated a very simple paraphrase system using the standard BLEU evaluation metric which has been shown to be a useful measure of semantic equivelance in Parapharse (Chen and Dolan, 2011). These corpus statistics are summarized in table 1.

- Motivate the need for a benchmark dataset for evaluating the writing style paraphrase task
- Present Shakespeare / Modern translation data as a situation where we have parallel data available (useful for building models & evaluating automatic evaluation metrics)

3 Evaluation Metrics

- Describe the need for automatic evaluation metrics.

- Describe previously used evaluation metrics for paraphrase.
- Highlight problems with previous metrics when targeting a specific writing style.
- Propose new metrics.

4 Experiments

- Experimental setup.
- Present results from human evaluation comparing various systems.
- Analyze correlation between evaluation metrics and human judgments.

5 Related Work

- Kevin Knight's work on poetry generation
- Any work on writing style (e.g. classification)? Possibly cite work on author attribution...
- work on paraphrase evaluation metrics (David Chen, CCB, etc...)

6 Conclusions

References

Bannard, C. and Callison-Burch, C. (2005). Paraphrasing with bilingual parallel corpora. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*.

Barzilay, R. and Lee, L. (2003). Learning to paraphrase: an unsupervised approach using multiple-sequence alignment. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1*, NAACL '03.

Chen, D. L. and Dolan, W. B. (2011). Collecting highly parallel data for paraphrase evaluation. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL-2011)*, Portland, OR.

Das, D. and Smith, N. A. (2009). Paraphrase identification as probabilistic quasi-synchronous recognition. 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: Volume 1 - Volume 1*, ACL '09.

Dolan, B., Quirk, C., and Brockett, C. (2004). Unsupervised construction of large paraphrase corpora: Exploiting massively parallel news sources. In *Proceedings of Coling 2004*.

Giampiccolo, D., Magnini, B., Dagan, I., and Dolan, B. (2007). The third pascal recognizing textual entailment challenge. In *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*, RTE '07.

Moore, R. C. (2002). Fast and accurate sentence alignment of bilingual corpora. In *Proceedings of the 5th Conference of the Association for Machine Translation in the Americas on Machine Translation: From Research to Real Users*, AMTA '02.

Shinyama, Y. and Sekine, S. (2003). Paraphrase acquisition for information extraction. In *Proceedings of the second international workshop on Paraphrasing - Volume 16*, PARAPHRASE '03.

Snow, R., O'Connor, B., Jurafsky, D., and Ng, A. Y. (2008). Cheap and fast—but is it good?: evaluating non-expert annotations for natural language tasks. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '08.