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

$Author1^{1,2}$ $Author2^{1,3}$
(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 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.

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 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 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 students better understand Shakespeare's work. A parallel corpus is extracted from these modern translations, 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 from the Machine Translation and paraphrase literature, however no previous work has investi-

gated the ability of these automatic metrics to capture the notion of writing style. We propose several new metrics for evaluating paraphrases targeting a specific writing style, and measure correlation with human judgements showing promising, yet preliminary results.

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, we believe this work also has the potential to make the other 20 plays more accessible to students of Shakespeare by automatically generating relatively high-quality modern translations.

2 Data

We propose to use Shakespeare's plays as a testbed for the task of paraphrasing while targeting a specific writing style. Because these plays are some of the highest regarded examples of English literature and are also very unique in style, many linguistic resources are available such as parallel corpora of modern translations and dictionaries of stylistically representative words and their modern equivalents.

We compare 3 different stylisitic paraphrase systems targeting Shakesperean English. One which leverages parallel corpora of modern translations, another which makes use of dictionaries of styalistically representative expressions, and another which leverages out-of-domain monolingual parallel data.

2.1 Modern Translations

Having access to parallel text in the target style allows us to train statistical models for generating paraphrases, and also perform automatic evaluation of semantic adequacy 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 were aligned using Bob Moore's bilingual sentence (Moore, 2002) aligner, which produced about 21,079 alignments out of 31,718 sentences in the Sparknotes data, and 10,365 sentence pairs out of 13,640 sentences in the enotes data. The modern translations from each source are qualitatively quite different. The Sparknotes paraphrases tend to differ significantly from the original text, whereas the enotes translations are much more conservative, making fewer changes. To illustrate these differences empirically and provide an initial paraphrase baseline, we computed BLEU scores of the unchanged modern translations against Shakespeare's original text; the Sparknotes paraphrases result in a BLEU score of 23.67, whereas the Enotes paraphrases prouce a much higher BLEU of 49.60 indicating their similarity to the original text. These corpus statistics are summarized in table 1.

2.2 Baselines

Phrase-based translation has been demonstrated to be an effective approach to generating paraphrses (Chen and Dolan, 2011; **?**), however this approach does require the existence of parallel corpora which may not be available for many writing styles. For this reason we were motivated to investigate alternative approaches.

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

target	source	target	source
ABATE	shorten	ANIGHT	by night
CAUTEL	deceit	CHILDING	pregnant
FOIL	defeat	MORTAL	deadly

Table 2: Exaple dictionary entries

2.2.1 Dictionary Based Paraphrase

Several dictionaries of stylistically representative words of Shakesperean English and their modern equivelants are available on the web. These dictionaries can be used to define a translation model which is used in combination with a language model as in standard phrase-based MT (**?**).

We gathered a set of 2,386 dictionary entries which were scraped from `http:/www.william-shakespeare.info` and semi-automaticaly cleaned. Example dictionary entries are presented in table 2.

TODO: need to describe how parameters were learned and combined with LM

2.2.2 Out of Domain Monolingual Parallel Data

2.3 Initial Comparison

- List the various systems we evaluate (e.g. varying LM size, etc...)
- Plot BLEU and PINC scores for the various systems

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.