A refined methodology for paraphrase acquisition.

ABSTRACT

Paraphrase acquisition is a process consists of extraction and validation but so far only the validation part was emphasized and yet the measurement of overall performance is validation-based. Also, many paraphrase acquisition approaches have been proposed and contributed to the research society. However, they were tested under different versions of methodologies individually which results in increasing the difficulty of benchmarking. In this paper, both the drawbacks and the importance of previous methodologies settings are discussed. Then we present a refined methodology with some modifications of the settings of previous methodologies. The way to carry out the experiment is explained step-by-step in detail. We also propose a new measurement criterion for extraction part which will be considered in overall performance measurement. Finally, this methodology is expected to yield more accurate, more comparable and better coverage result over previous methodologies. In the future, we plan to apply this methodology to our approach.