Discourse-information-based automatic evaluation of public legal education texts

Juan Liu

School of Foreign Languages,
Guangzhou Maritime University,
No. 101, Hongshan No.3 Road,
Huangpu District,
Guangzhou 510330,
China

Awarding Institution: Guangdong University of Foreign Studies, School of Foreign Languages

Date of Award: 21 June 2018

KEYWORDS: DISCOURSE INFORMATION, PUBLIC LEGAL EDUCATION, AUTOMATIC EVALUATION, EVALUATIVE VALUES, EVALUATIVE CONCEPTS
Since Public Legal Education (PLE) websites have become one of the primary means of communication between the dominant organiser of PLE in China – the government – and the public, Public Legal Education Texts (PLET), which report legal news or interviews of legal experts, or comment on legal cases for the purpose of communicating legal knowledge, normally in the form of a title and a body, need to be improved in order to attract readers and enhance the effectiveness of PLE. Nonetheless, it can hardly be achieved without identifying readers’ evaluative values of PLET. In view of this, this study aims to construct a general automatic evaluation mechanism for PLET on the basis of Discourse Information Theory (DIT) (Du, 2007, 2011, 2012, 2014), a study of information from the discourse perspective with the help of text-mining technologies. It is hoped that the automation of the evaluation mechanism can contribute to the promotion of the linguistic findings of this research in the circle of PLE. Such a research objective is formulated into three specific research questions as follows:

1. What core values serve as the criteria for readers to evaluate PLET?
2. What are the discourse information representations (DIRs) and discourse information patterns (DIPs) that reflect the evaluation values?
3. How can the general automatic evaluation mechanism be realised?

As to methodologies, both qualitative and quantitative analyses are utilised. Qualitatively, core evaluative values and concepts are explored following Classic Grounded Theory (CGT) (Glaser, 1978), a qualitative research methodology that incorporates guidelines for simultaneous data collection and analysis to develop theories about social processes that are grounded in real-life experiences. This is differentiated from other Grounded Theories in that it emphasises introduction or emergence, and the individual researcher’s creativity within a clear frame of stages.

Quantitatively, statistics are used for the correlation analysis of evaluative concepts and the regression modelling of the global score, and for the identification and induction of DIRs and DIPs. The data analysed in this study include two batches: one is the transcription of the interviews and the texts the interviewees read for the qualitative study; the other is a small-sized corpus of online PLE texts annotated with DIT for quantitative study which is one of the sub-corpora of the Corpus for the Legal Information Processing System (CLIPS) and the two experts’ rating of the PLE texts.

Specifically, this study first explores the core evaluative values and concepts in the readers’ cognition through intensive interviews and analysis following the steps in CGT. Eight evaluative concepts are generalised through coding, namely:

Concept 1 Topics: novelty and relatedness;
Concept 2 Language Level: funny, story-like, in a friendly tone, plain;
From these, four core evaluative values are induced: informativeness, social distance, objectivity and manner. After that, according to the global scores and analytical scores as to each concept given by expert raters for the texts from the corpus, a model which represents the relationship between the global score and the analytical scores for valid evaluative concepts is built up with the statistical method of Multi-linear Regression analysis. The valid evaluative concepts are namely, Concept 1 (the novelty and relatedness of the topics), Concept 4 (the brevity of the body) and Concept 5 (the variety of the language).

Furthermore, Discourse Information Analysis (DIA) is conducted both from macro and micro perspectives in order to identify valid DIRs and DIPs for distinguishing evaluative score sections of each evaluative concept.

Finally, a general evaluation mechanism is constructed consisting of four modules, which are the processing module for Concept 1, the processing module for Concepts 4 and 5, the global scoring module and the evaluation result output module. During the construction, the merging of the processing modules for Concepts 4 and 5 contributed to the operating effectiveness of the general mechanism. In addition, the programming ideas and the solutions to difficulties in programming are illustrated with codes and explanations.

The major contribution of this study lies in the exploration of core evaluative values in readers’ cognition through field study and the construction of an automatic evaluation mechanism for PLET based on discourse information by simulating human beings’ thinking modes. In addition, two new subtypes of information elements ‘Quality tip (Qt)’ and ‘Turn tip (Tt)’ were put forward, and the automatic annotating and processing of certain subtypes of information elements and information knots were realised. This study promotes DIT from the perspective of automatic processing, sheds light for the development of text mining, especially in predictive modelling, and meanwhile may offer a practical instrument for PLE practitioners to evaluate and improve PLE composition.

**Acknowledgments**

I am very grateful to Professor Du Jinbang, Professor Chen Jinshi, Professor Yuan Chuanyou, Professor Xu Zhanghong, Dr Zhang Shaomin and Dr Ge Yunfeng for their inspiring suggestions and comments. This work was also supported by the
National Social Science Project of China (Project No.: 18BYY073) led by Professor Chen Jinshi.

References


