Truth and Deception at the Rhetorical Structure Level

Victoria L. Rubin
Language and Information Technology Research Laboratory (LiT.RL), Faculty of Information and Media Studies, University of Western Ontario, North Campus Building, Room 260, London, Ontario, N6A 5B7 Canada. E-mail: vrubin@uwo.ca

Tatiana Lukoianova[1]
Fisher College of Business, Ohio State University, 250 Fisher Hall, Columbus, OH, 43210. E-mail: vashchilko.1@osu.edu

This paper furthers the development of methods to distinguish truth from deception in textual data. We use rhetorical structure theory (RST) as the analytic framework to identify systematic differences between deceptive and truthful stories in terms of their coherence and structure. A sample of 36 elicited personal stories, self-ranked as truthful or deceptive, is manually analyzed by assigning RST discourse relations among each story's constituent parts. A vector space model (VSM) assesses each story's position in multidimensional RST space with respect to its distance from truthful and deceptive centers as measures of the story's level of deception and truthfulness. Ten human judges evaluate independently whether each story is deceptive and assign their confidence levels (360 evaluations total), producing measures of the expected human ability to recognize deception. As a robustness check, a test sample of 18 truthful stories (with 180 additional evaluations) is used to determine the reliability of our RST-VSM method in determining deception. The contribution is in demonstration of the discourse structure analysis as a significant method for automated deception detection and an effective complement to lexicosemantic analysis. The potential is in developing novel discourse-based tools to alert information users to potential deception in computer-mediated texts.

Introduction

In recent years, the question of whether there is deception in textual information has been addressed algorithmically. Although deception has been traditionally studied outside of the library and information science field (in disciplines such as psychology and communication studies), automated deception detection methods are now of interest in information science because of their application (a) in tools to support and enhance human abilities in discerning information from misinformation, concepts originally linked by Fox (1983)[2]; (b) as an additional metric for information quality assessment discussed by Rubin and Vashchilko (2012)[3] and Lukoianova and Rubin (2014),[4] and (c) as a

[1]Tatiana Lukoianova is also known as Tatiana Vashchilko.

Received June 11, 2013; revised December 16, 2013; accepted December 17, 2013

[2]Fox (1983) analyzes the nature of information and misinformation as "contained in, transferred by, and conveyed by sets of sentences" (p. 212). "The difference between informing and misinforming seems to hinge on, or ultimately depend on, the fact that sometimes what one person tells another is true and sometimes it is false. Indeed one is inclined to divide all cases in which *X* tells *Y* that *P* into two mutually exclusive and exhaustive classes on the basis of whether *P* is true or false" (p. 118). Fox concludes that "informing does not require truth" but "misinforming does require falsity" (p. 160).

[3]In their summary, Rubin and Vashchilko (2012, p. 2) say that information quality (IQ) is traditionally defined and assessed "based on the usefulness of information or its 'fitness for use' by delineating various dimensions along which IQ can be measured quantitatively. One of the four major dimensions of IQ is intrinsic IQ, in which various authors assigned such components as accuracy, believability, reputation, and objectivity (Wang & Strong, 1996); accuracy and factuality (Zmud, 1978); believability, accuracy, credibility, consistency, and completeness (Jarke & Vassiliou, 1997); accuracy, precision, reliability, and freedom from bias (Delone & McLean, 1992); accuracy and reliability (Goodhue, 1995); accuracy and consistency (Ballou & Pazer, 1995); correctness and unambiguous (Wand & Wang, 1996)." Rubin and Vashchilko (2012) propose to extend IQ assessment methodology with a veracity/deception dimension because it differs contextually from accuracy and other well-studied components of intrinsic IQ, and almost no literature considers the deception as one of the components of the intrinsic IQ despite extensive research on deception detection and its importance for written communication.

measure for the assessment of source credibility, the notion traditionally seen in library and information science juxtaposed to cognitive authority such as that of Rieh (2010). Deceptive pieces of information, even authoritatively stated, can mislead or misinform information users, negatively affecting the outcome of their decision making. Users' mere awareness of the possibility of deception (and other types of information manipulation[5]) in computer-mediated communication contributes to users' information literacy (Rubin & Chen, 2012) and may lead to more discriminating ways of engaging with information. The potential of furthering deception detection research is in providing new methods as the basis for developing novel tools to assist information seekers, online readers, or decision makers by alerting them to potential deception in computer-mediated texts.

Automated deception detection is a challenging task (DePaulo, Charlton, Cooper, Lindsay, & Muhlenbruck, 1997), only recently proved feasible with natural language processing and machine learning techniques (Bachenko, Fitzpatrick, & Schonwetter, 2008; Fuller, Biros, & Wilson, 2009; Hancock, Curry, Goorha, & Woodworth, 2008; Rubin, 2010; Zhou, Burgoon, Nunamaker, & Twitchell, 2004). The idea is to distinguish truthful from deceptive information, where deception usually implies an intentional and knowing attempt on the part of the sender to create a false belief or false conclusion in the mind of the receiver of the information (e.g., Buller & Burgoon, 1996; Zhou et al., 2004). This article focuses solely on textual information, in particular, on computer-mediated personal communications such as e-mail and online posts.

Previously suggested techniques for detecting deception in text reach modest accuracy rates at the level of lexicosemantic analysis. Certain lexical items are considered to be predictive linguistic cues and could be derived, for example, from the statement validity analysis techniques used in law enforcement for credibility assessments (as described by Porter & Yuille, 1996). Although there is no clear consensus on reliable predictors of deception,[6] deceptive cues are identified in texts, extracted, and clustered conceptually, for instance, to represent diversity, complexity, specificity, and nonimmediacy of the analyzed texts (see, e.g., Zhou et al., 2004). When implemented with standard classification algorithms (such as neural nets, decision trees, and logistic regression), such methods achieve 74% accuracy (Fuller et al., 2009). Existing psycholinguistic lexicons (e.g., linguistic inquiry and word count [LIWC], described by Pennebaker, Chung, Ireland, Gonzales, & Booth, 2007) have been adapted to perform binary text classification for truthful versus deceptive opinions, with an average classifier demonstrating a 70% accuracy rate (Mihalcea & Strapparava, 2009).

These modest results, though usually achieved on restricted topics, are promising in that they supersede notoriously unreliable human abilities in lie–truth discrimination tasks. On average, people are not very good at spotting lies (Vrij, 2000), succeeding only about half of the time (Frank, Paolantinio, Feeley, & Servoss, 2004). For instance, a meta-analytical review of more than 100 experiments with more than 1,000 participants showed a 54% mean accuracy rate at identifying deception (DePaulo et al., 1997). Human judges achieve 50% to 63% success rates, depending on what is considered deceptive on a 7-point truth-to-deception continuum scale (Rubin & Conroy, 2011, 2012), but, the higher the actual self-reported deception level of the story, the more likely a story will be confidently assigned as deceptive. In other words, extreme degrees of deception are more apparent to judges.

The task for current automated deception detection techniques has been formulated as binary text categorization (Is a message deceptive or truthful?), and the decision applies to the entire analyzed text. This an overall discourse-level decision, so it may be reasonable to consider discourse or pragmatic features of each message. Thus far, discourse is surprisingly rarely considered, if at all, and most of the effort has been restricted to lexicosemantic verbal predictors. A rare exception is the Bachenko et al. (2008) study, which focuses on the truth or falsity of individual propositions, achieving a finer-grained level of analysis,[7] but the propositional interrelations within the discourse structure are not considered. To the best of our knowledge, there have been no attempts in the task of automated deception detection to incorporate discourse structural features and/or text coherence analysis at the pragmatic levels of story interpretation.

Study Objective

With the recent advances in the identification of verbal cues of deception in mind and the realization that they focus on linguistic levels below discourse and pragmatic analysis,

[4]Lukoianova and Rubin (2014) recognize a broader concept of *veracity* as a necessary property for effective utilization of big data (complementing the three previously established big data quality dimensions: volume, variety, and velocity). Big data is seen as having biases, ambiguities, and inaccuracies that affect data quality and have to be identified and accounted for to reduce inference errors and improve the accuracy of generated insights. The composite big data veracity index combines three main dimensions (truthfulness/deception dimension, discussed in this article, as well as objectivity/subjectivity and credibility/implausibility dimensions) and proposes a useful way to measure systematic variations in quality across big data sets with textual information.

[5]Rubin and Chen's (2012) information manipulation classification theory conceptualizes and differentiates broad types of information manipulation (including falsification, exaggeration, concealment, and hoax) by 12 salient factors (such as intentionality, accuracy, and social acceptability).

[6]In retrospect, some researchers question whether there is "a suitably computable basis" for a mechanism of identifying deception and building a deception detection system based on lexicosemantic deception cues because of disagreements in measuring and interpreting such cues (Vartapetiance & Gillam, 2012).

[7]Using a corpus of criminal statements, police interrogations, and legal testimonies, their regression and tree-based classification automatic tagger performs at average 69% recall and 85% precision rates compared with the performance of human taggers on the same subset (Bachenko et al., 2008).

DOI: 10.1002/asi

this study focuses on one main question: What is the impact of the relations among discourse constituent parts on the discourse composition of deceptive and truthful messages?

We hypothesize that if the relations among discourse constituent parts in deceptive messages differ from those in truthful messages, then systematic analysis of such relations will help to detect deception. To investigate this question, we propose to use a novel method for deception detection research: rhetorical structure theory (RST) analysis with subsequent application of the vector space model (VSM). We name this approach the RST-VSM method. RST analysis is promising in deception detection because RST analysis captures coherence of a story in terms of functional relations among different meaningful text units and describes a hierarchical structure of each story (Mann & Thompson, 1988). The result is that each story is represented by a set of RST relations connected in a hierarchical manner, with more salient text units heading this hierarchical tree. We also propose to utilize the VSM model to convert the derived RST relations' frequencies into meaningful clusters of diverse deception levels. To evaluate the proposed RST-VSM approach to deception detection in texts, we compare human assessment with the RST analysis of deception levels for the set of 36 deceptive and truthful stories. To determine whether the results are generalizable, we use a second sample of 18 completely truthful stories to validate the method.

The article proceeds with three main parts. The next part discusses the methodological foundations of the RST-VSM approach. The following part describes the data and the method of collecting the sample. Finally, the Results section demonstrates the identified levels of deception and truthfulness, their distribution across truthful and deceptive stories, and the out-of-sample analysis.

RST-VSM Method: Combining the Vector Space Model and Rhetorical Structure Theory

The VSM seems to be very useful in the identification of truth and deception types of written stories, especially if the meaning of the stories is represented as RST relations. RST differentiates between rhetorically stand-alone parts of a text, some of which are more salient (nucleus) than the others (satellite). During the past couple of decades, empirical observations and previous RST research have confirmed that writers tend to emphasize certain parts of a text to make sure that their most essential ideas reach the reader. These parts can be systematically identified through analysis of the rhetorical connections among more and less essential parts of a text. RST helps to describe and quantify text coherence through a set of constraints on nucleus and satellites. The main function of these constraints is to describe in a meaningful way why and how one part of a text connects to the others within a hierarchical tree structure, which is an RST representation of a coded text. The names of the RST relations also resemble the purpose of using the connected text parts together.

For example, one of the RST relations that appears in truthful stories but never appears in the deceptive stories in our sample is "evidence." The main purpose of using "evidence" to connect two parts of text is to present additional information in a satellite so that the reader's belief in the information in the nucleus increases. However, this can happen only if the information in the satellite is credible from the reader's point of view. For some reason, the RST coding has never used evidence in 18 deceptive stories, but used it rather often in 18 truthful stories. This might indicate that (a) writers of deceptive stories did not see any purpose in supplying additional information to the readers to increase their beliefs in communicating the writers' essential ideas; (b) the information presented in the satellite was not credible from the reader's point of view, which did not qualify the relationship between nucleus and satellite to be classified as the evidence relation; or (c) both (see example of an RST diagram in Figure 1).

Our premise is that if there are systematic differences between deceptive and truthful written stories in terms of their coherence and structure, then the RST analysis of these stories can identify two sets of RST relations and their structure. One set is specific for the deceptive stories, and the other one is specific for the truthful stories. We propose to use a VSM for the identification of these sets of RST relations. Mathematically speaking, written stories have to be modeled in a way suitable for the application of various computational algorithms based on linear algebra. Using a VSM, the written stories can be represented as vectors in a high-dimensional space (Manning & Schütze, 1999; Salton & McGill, 1983). According to the VSM, stories are represented as vectors, and the dimension of the vector space is equal to the number of RST relations in a set of all written stories under consideration. Such a representation of written stories makes the VSM attractive in terms of its simplicity and applicability (Baeza-Yates & Ribeiro-Neto, 1999).

The VSM[8] is the basis for almost all clustering techniques when dealing with the analysis of texts. Once the texts are represented according to VSM, as vectors in an n-dimensional space, we can apply myriad cluster methods that have been developed in computational science, data mining, and bioinformatics. Cluster analysis methods can be divided into two large groups (Zhong & Ghosh, 2003): discriminative (or similarity-based) approaches (Indyk, 1999; Scholkopf & Smola, 2001; Vapnik, 1998) and generative (or model-based) approaches (Bilmes, 1998; Cadez, Gaffney, & Smyth, 2000; Rose, 1998).

The main benefit of applying the VSM to RST analysis is that the VSM allows a formal identification of coherence and structural similarities among stories of the same type (truthful or deceptive). For this purpose, RST relations are dimensions of the RST space, in which each story is an element or a vector in the RST space (Figure 2).

[8]Tombros (2002) maintains that most of the research related to the retrieval of information is based on vector space modeling.

DOI: 10.1002/asi

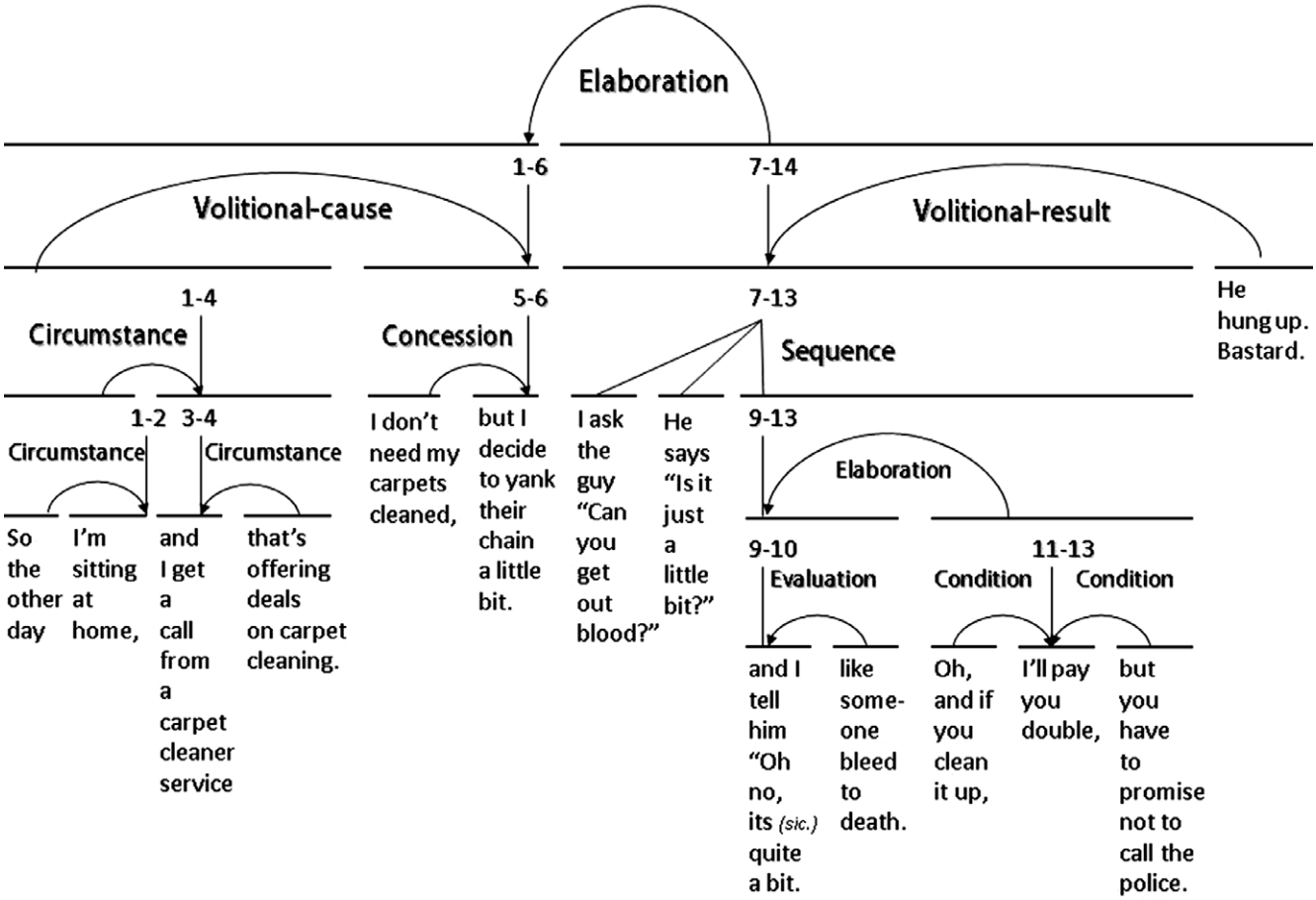

FIG. 1. Sample diagram for RST analysis.

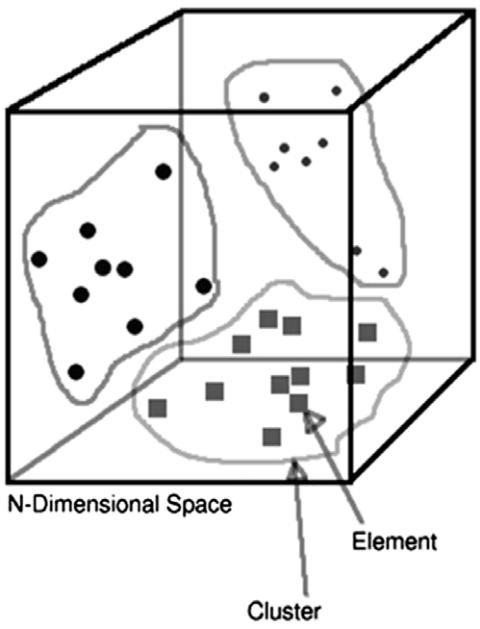

FIG. 2. Cluster representation of story sets or RST relations (Karypis, 2002; Rasmussen & Karypis, 2004).

Visually, one can think about the RST relations as a cube (Figure 2) in which each dimension is an RST relation.

The main subsets of this set of stories are two clusters, deceptive stories and truthful stories. The element of a cluster is a story, and a cluster is a set of elements that share enough similarity to be grouped together as either deceptive stories or truthful stories (Berkhin, 2002). That is, there are several distinctive features (RST relations, their co-occurrences, and their positions in a hierarchical structure) that make each story unique and cause it to be classified in a particular cluster. These distinctive features of the stories are compared, and, when some similarity threshold is met, they are placed in one of two groups, deceptive or truthful stories.

Similarity[9] is one of the key concepts in cluster analysis, and most of the classical techniques (k-means, unsupervised Bayes, hierarchical agglomerative clustering) and rather recent ones (CLARANS, DBSCAN, BIRCH, CLIQUE, CURE, etc.) "are based on distances between the samples in the original vector space" (Strehl, Ghosh, & Mooney, 2000, p. 58). Such algorithms form a similarity-based clustering

[9]"Interobject similarity is a measure of correspondence or resemblance between objects to be clustered" (Hair, Anderson, Tatham, & Black, 1995, p. 429).

DOI: 10.1002/asi

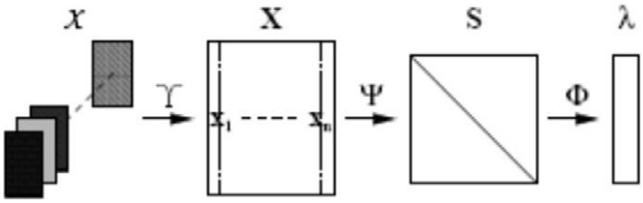

FIG. 3. Similarity-based clustering framework (Strehl et al., 2000).

framework (Figure 1) as described by Strehl et al. (2000) or as Zhong and Ghosh (2003) define it, as discriminative (or similarity-based) clustering approaches. That is why this article modifies Strehl et al.'s (2000) similarity-based clustering framework (Figure 3) to develop a unique RST-VSM methodology for deception detection in text.

The RST-VSM method includes three main steps.

1. The set of written stories, *X*, is transformed into the vector space description, X, using some rule, *Y*, that in our case corresponds to an RST analysis and identification of RST relations as well as their hierarchy in each story.
2. This vector space description X is transformed into a similarity space description, S, using some rule, Ψ, which in our case is the Euclidean distance of every story from deception and truth centers correspondingly based on normalized frequency of RST relations in a written story.[10]
3. The similarity space description, S, is mapped into clusters based on the rule Φ, which we define as the relative closeness of a story to a deception or a truth center; if a story is closer to the truth center, then the story is placed in a truth cluster, whereas, if a story is closer to a deception center, then the story is placed in a deception cluster.

Data Collection and Sample

The data set contains 36 unique personal stories in the main training sample and 18 truthful stories in the second sample for out-of-sample testing, elicited using Amazon's online survey service, Mechanical Turk (http://www.mturk.com). Respondents in one group were asked to write a rich, unique story, one that is completely true or with some degree of deception. Respondents in another group were asked to evaluate the stories written by the respondents in the first group.[11]

Two groups of 18 stories comprise the first main training data sample. The first group consists of 18 stories that were self-ranked by their writers as completely deceptive on a 7-point Likert scale from complete truth to complete deception (deceptive self-reported group). The second group includes stories that their writers rated as completely truthful stories (truthful self-reported group). The second group is matched in numbers for direct comparisons with the first group by selecting 18 stories from a larger group of 39 completely truthful stories at random (Rubin & Conroy, 2011, 2012). The second test sample for out-of-sample analysis consists of another 18 self-ranked truthful stories. Each story (in truthful self-reported and deceptive self-reported groups) has 10 unique human judgments associated with it. Each judgment is binary ("judged truthful" or "judged deceptive") and has an associated confidence level assigned by the judge (either "totally uncertain," "somewhat uncertain," "I'm guessing," "somewhat certain," or "totally certain"). Each writer and judge was encouraged to provide reasons for defining a story as truthful or deceptive and assigning a particular confidence level. In total, 396 participants contributed to the first main sample, 36 of whom were story writers and 360 of whom were judges performing the lie–truth discrimination task by confidence level. Another 198 participants contributed to the second test sample, 18 of whom were story writers and 180 of whom were judges performing the lie–truth discrimination task by confidence level.

We combined the 10 judges' evaluations of a story into one measure, the expected level of a story's deception or truthfulness. Because judges' confidence levels reflect the likelihood of a story being truthful or deceptive, the probability of a story being completely truthful or deceptive equals 1 and corresponds to a "totally certain" confidence level that the story is truthful or deceptive. In the same way, the other levels of confidence have the following probability correspondences: "totally uncertain" has probability 0.2 of a story being deceptive or truthful, "somewhat uncertain" 0.4, "I'm guessing" 0.6, and "somewhat certain" 0.8.

Two dummy variables are created for each story. One dummy, a deception dummy, equals 1 if a judge rated the story as "judged deceptive" and 0 otherwise. The second dummy, the truthfulness dummy, equals 1 if a judge rated the story as "judged truthful" and 0 otherwise. Then the expected level of deception of a story equals the product of the probability (confidence level) of deception and the deception dummy across 10 judges. Similarly, the expected level of truthfulness is equal to the product of the probability of truthfulness (confidence level) and the truthfulness dummy across 10 judges.

Thirty-six stories, evenly divided between truthful and deceptive self-report groups, were manually analyzed using the classical set of Mann and Thompson's (1988) RST relations, which has been extensively tested empirically (Taboada & Mann, 2006). As a first stage of the RST-VSM methodology development, the manual RST coding was required to deepen the understanding of the rhetorical relations and structures specific for deceptive and truthful stories. Moreover, the manual analysis aided by O'Donnell's

[10]RST stories when represented as vectors differ in length, so the normalization step is necessary to make them comparable. The coordinates of every story (the frequency of an RST relation in a story) are divided by the vector's length.

[11]For further details on the data collection process and the discussion of associated challenges, see Rubin and Conroy (2012).

DOI: 10.1002/asi

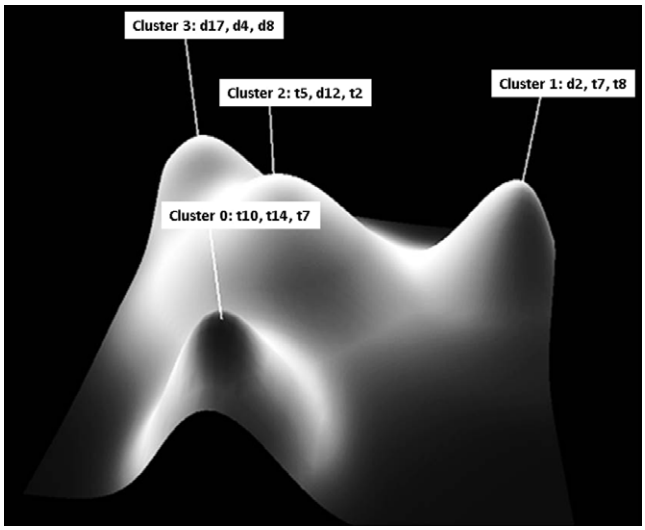

FIG. 4. Four clusters in RST space by level of deception.

RSTTool (http://www.wagsoft.com/RSTTool/) ensures higher reliability of the analysis and avoids compilation of errors, because the RST output further served as the VSM input. Taboada (2004) reports on Marcu's RST Annotation Tool (http://www.isi.edu/licensed-sw/RSTTool/) and Ghorbel's RhetAnnotate and provides a good overview of other recent RST resources and applications. The acquired knowledge during manual coding of deceptive stories along with recent advances in automated RST analysis will later help in evaluating the RST-VSM method, reduce subjective variations in data coding, and design a completely automated deception detection tool relying on the automated procedures to recognize rhetorical relations which utilize the full rhetorical parsing (Marcu, 1997; Marcu & Echihabi, 2002).

Recent efforts in developing a fully automated RST-based discourse parser confirm that it is feasible to extend traditional sentence-level discourse parsing to text-level parsing. The first fully implemented text-level discourse parser, the HILDA discourse parser, was proposed by Hernault, Prendinger, duVerle, and Ishizuka (2010) and later extended by Feng and Hirst (2012) by adding original rich linguistic features as well as those suggested by Lin, Kan, and Ng (2009). With time and continued improvements, the subjectivity in establishing rhetorical relations by manual coders will become less of a challenge; the task is being systematized algorithmically.

Results

The preliminary clustering of 36 stories in RST space using various clustering algorithms shows that RST dimensions can systematically differentiate between truthful and deceptive stories as well as diverse levels of deception (Figure 4).

The visualization uses CLUTO software,[12] which finds the clustering solution as a result of the optimization of a "particular function that reflects the underlying definition of the 'goodness' of the cluster" (Rasmussen & Karypis, 2004, p. 3). Among the four clusters in RST space, two clusters are composed of completely deceptive stories (far back left peak, cluster 3) or entirely truthful stories (front peak, cluster 0), the other two clusters have a mixture with the prevalence of truthful stories (clusters 1 and 2). This preliminary investigation of using RST space for deception detection indicates that the RST analysis seems to offer a systematical way of distinguishing between truthful and deceptive features of texts.

This article develops an RST-VSM method by using RST analysis of each story in n-dimensional RST space, with subsequent application of the vector space model to identify the level of a story's deception. Every normalized frequency of an RST relation in a story is a distinct coordinate in the RST space. Each story has 26 coordinates in RST space, the number of identified RST relations in the entire main sample. The story writers' ratings are used to calculate the centers for the truth and deception clusters based on corresponding writers' self-rated truthful and deceptive sets of stories in the sample. The normalized Euclidean distances between a story and each of the centers are defined as the degree of deception of that story depending on its closeness to the deception center. The closer a story is to the deception center, the higher is its level of deception. The closer a story is to the truthful center, the higher is its level of truthfulness.[13]

Our RST-VSM-based approach seems to differentiate between truthful and deceptive stories. The difference in means (*t*-test) demonstrates that the truthful stories have a statistically significantly lower average number of text units per statement than the deceptive stories ($t = -1.3104$), although these differences are not large statistically, at only the 10% significance level. The normalized frequencies of the RST relations appearing in the truthful and deceptive stories differ for about one third of all RST relations based on the statistical significance testing (Table 1).

The comparison of the distribution of RST relations across deception and truth centers constructed based on the 36 stories from the first main training sample demonstrates that, on average, the frequencies and the usage of such RST relations as conjunction, elaboration, evaluation, list, means, nonvolitional cause, nonvolitional result, sequence, and solutionhood in deceptive stories exceed those in the truthful ones (Figure 5).

On the other hand, the average usage and frequencies of RST relations such as volitional result, volitional cause, purpose, interpretation, concession, circumstance, and antithesis in truthful stories exceed those in the deceptive ones. Some of the RST relations are specific for only one type of story: enablement, restatement, and evidence appear

[12]http://glaros.dtc.umn.edu/gkhome/cluto/gcluto/overview

[13]All calculations were performed in STATA.

DOI: 10.1002/asi

TABLE 1. Comparison of the normalized frequencies of the RST relationships in truthful and deceptive stories: Difference in means test.

RST relationships appearing in truthful and deceptive stories with *no* statistically significant differences	RST relationships appearing in the truthful stories with statistically significantly *greater* normalized frequencies than the deceptive ones	RST relationships appearing in the truthful stories with statistically significantly *lower* normalized frequencies than the deceptive ones
Background, circumstance, concession, condition, conjunction, elaboration, enablement, interpretation, list, means, nonvolitional cause, nonvolitional result, purpose, restatement, sequence, solutionhood, summary, unconditional	Antithesis ($t = 2.3299$) Evidence ($t = 3.7996$) Joint ($t = 1.5961$) Volitional cause ($t = 1.8597$) Volitional result ($t = 1.8960$)	Disjunction ($t = -1.7850$) Evaluation ($t = -2.0762$) Preparation ($t = -1.7533$)

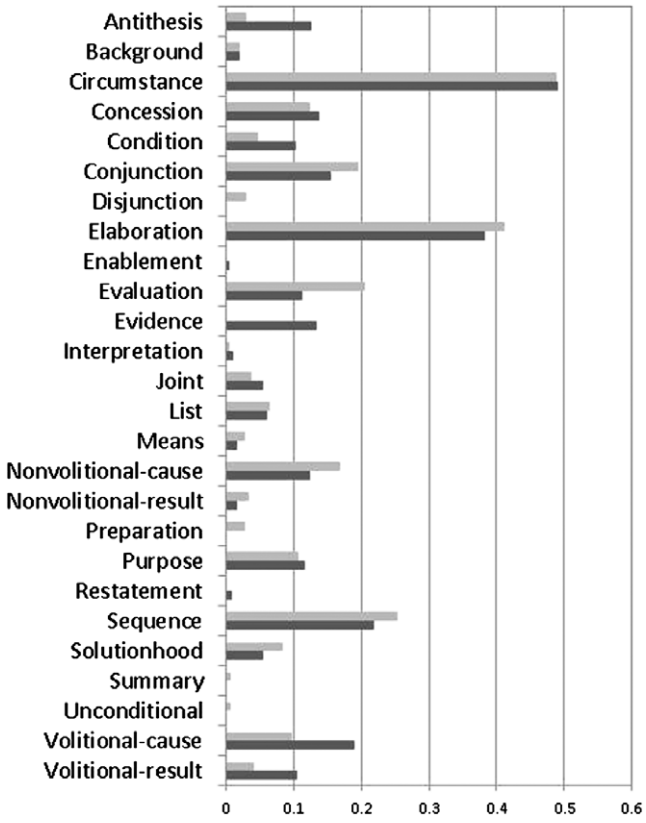

FIG. 5. Comparison of the RST relations composing the deceptive cluster center (top light gray bar) and the truthful cluster center (bottom dark gray bar).

only in truthful stories, whereas summary, preparation, unconditional, and disjunction appear only in deceptive stories.

The histograms of distributions of deception (truthfulness) levels assigned by judges and derived from RST-VSM analysis demonstrate some similarities between the two for truthful and for deceptive stories (Figures 6 and 7). More rigorous statistical testing reveals that only truthfulness levels in deceptive stories assigned by judges do not have a statistically significant difference from the RST-VSM ones. We used the Wilcoxon signed rank sum test, the nonparametric version of a paired-samples *t*-test (STATA 12 command signrank). For other groups, the judges' and RST assessments differ significantly. The difference is especially apparent in the distributions of deception and truthfulness in truthful stories. Among these differences, the RST-VSM method counted 44.44% of stories having 50% deception level, whereas judges counted 61.11% of the same stories having a low deception level of no more than 20%. The level of truthfulness was also much higher in the judges' assessment than the level based on RST-VSM calculations.

The distribution of the levels of deception and truthfulness across all deceptive stories (Figure 6) and across all truthful stories (Figure 7) shows variations in patterns of deception levels based on RST-VSM. In deceptive stories, the RST-VSM levels of deception are consistently higher than the RST-VSM levels of truthfulness. Assuming that the writers of the stories did make them up, the RST-VSM approach seems to offer a systematic way of detecting a high level of deception with rather good precision.

The RST-VSM deception levels are not as high as those of human judges, with human judges assigning much higher levels of deception to deceptive stories than to truthful stories (Figures 8 and 9). Assuming that the stories are indeed made up, the human judges have greater precision than the RST-VSM method. Nevertheless, RST-VSM analysis assigns higher deception levels to those stories that also receive higher human judges' deception levels. This pattern is consistent across all deceptive stories.

Validation of RST-VSM Method

To validate the method, we use the second test sample of 18 self-ranked truthful stories. We identify the coordinates of the 18 stories in RST space. To evaluate whether a story is deceptive or truthful, we measure the distance of each of the 18 stories in the second sample from the corresponding centers. The coordinates of truthful and deceptive centers in RST space are derived from the 36 stories of the training sample and are exactly the same as in the Results section.

The distributions of the expected levels of deception and truthfulness (Figure 10) demonstrate that the human judges consider most of the stories more truthful than deceptive (the lowest level of truthfulness is 0.4, with only 22.22% of the stories in this category, and the maximum level of

DOI: 10.1002/asi

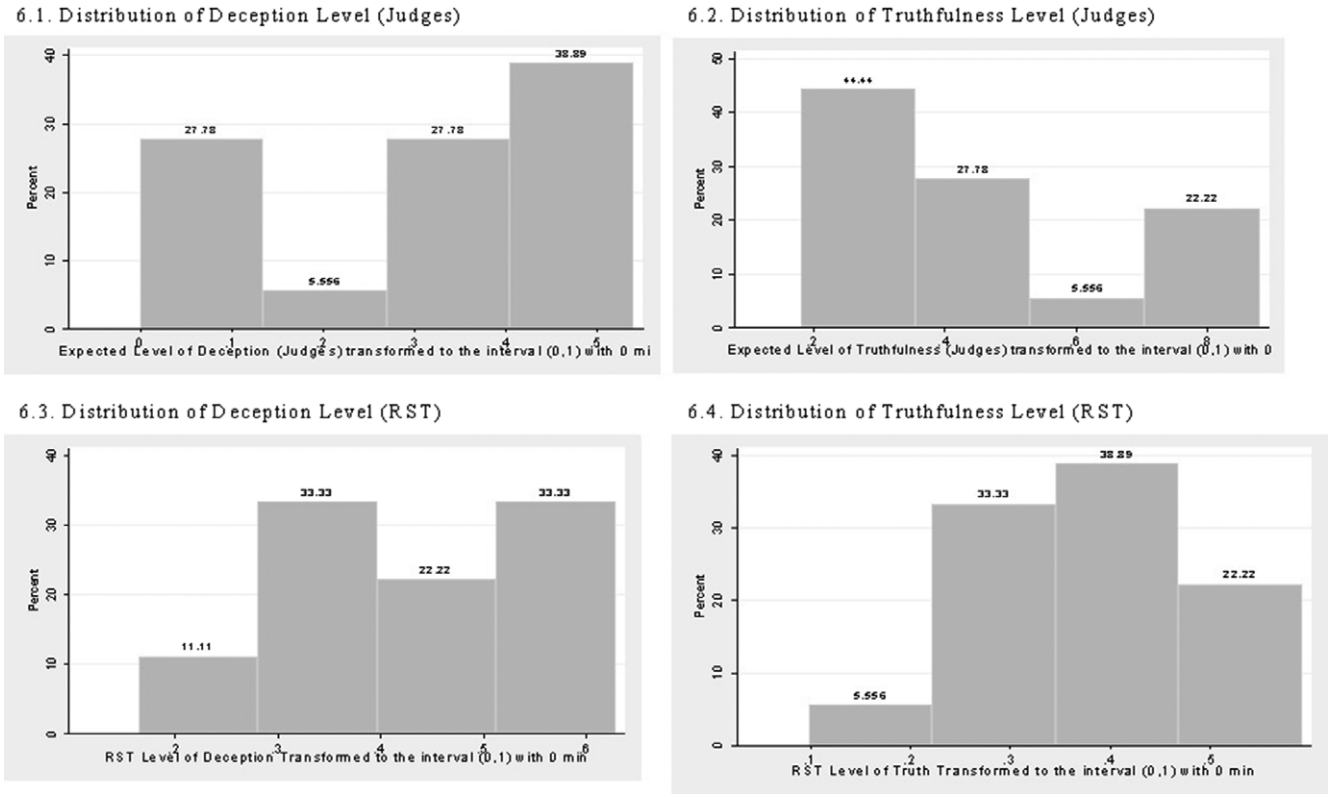

FIG. 6. Distribution of deception and truthfulness levels for deceptive stories (main sample).

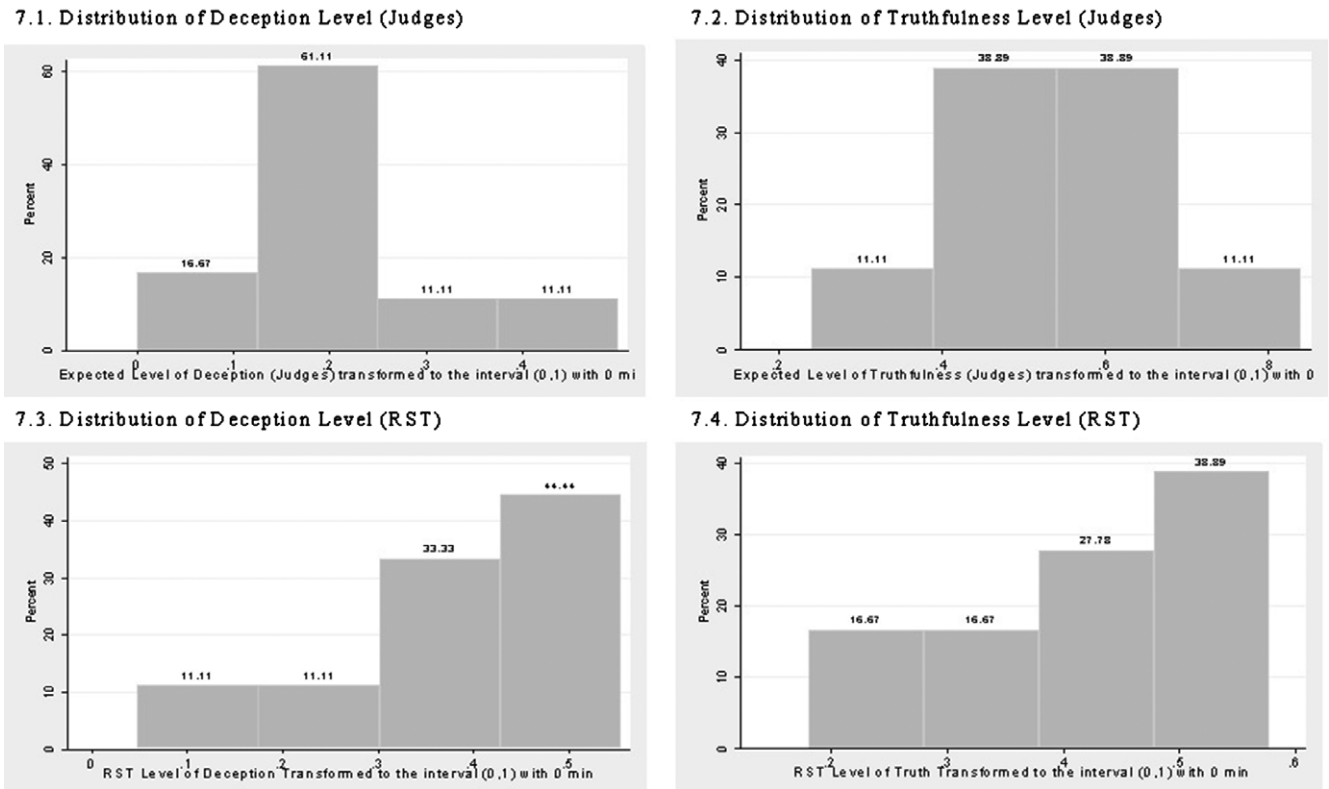

FIG. 7. Distribution of deception and truthfulness levels for truthful stories (main sample).

DOI: 10.1002/asi

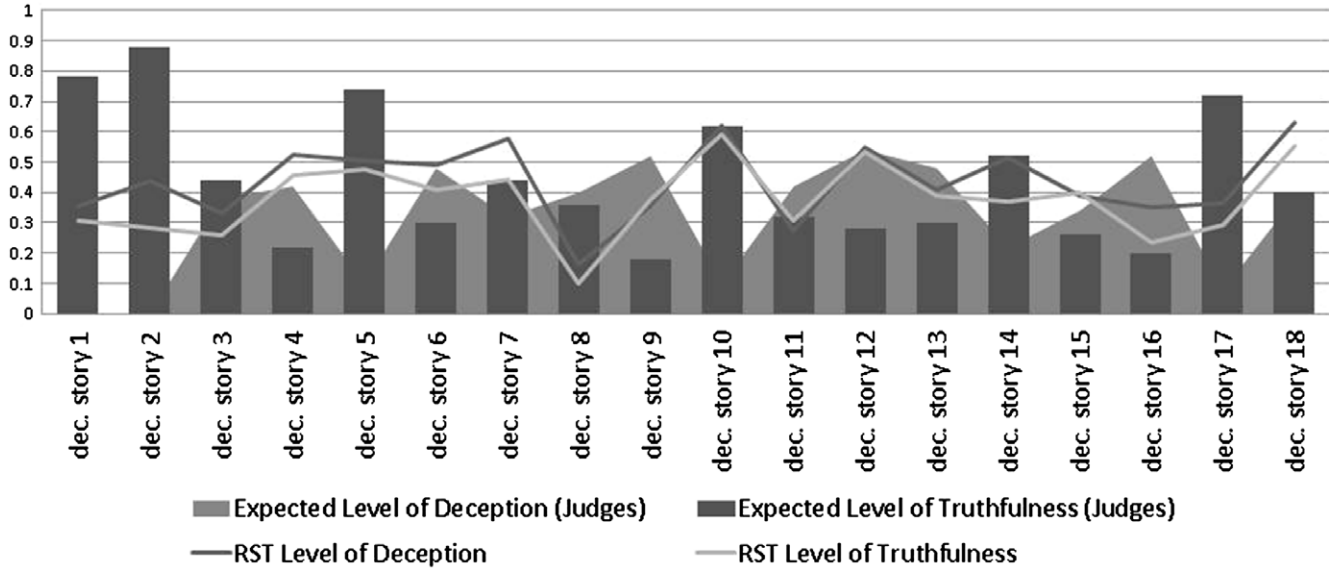

FIG. 8. Distributions of expected levels of deception and truthfulness in deceptive stories (main sample).

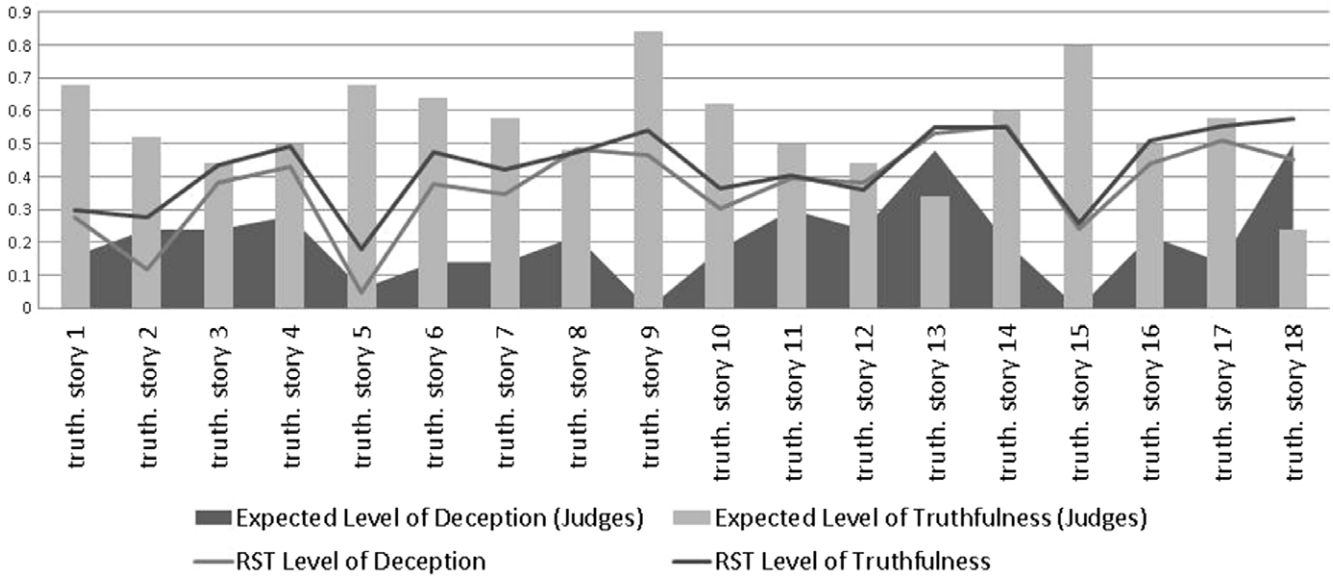

FIG. 9. Distributions of expected levels of deception and truthfulness in truthful stories (main sample).

truthfulness is 0.7, with 33.33% of the stories in this category). The expected levels of deceptions assigned by judges are very low for these 18 stories, the highest level of deception being no more than 0.4.

RST analysis assigns slightly higher levels of deception than human judges to all these originally truthful stories (Figure 11). Figure 11 provides a comparison of expected levels of deception assigned using the RST-VSM model and levels assigned by human judges. The graph demonstrates that the RST-VSM methodology and human judges have agreed in classifying 12 of 18 stories (67%) as truthful. In other words, the metric was performing as expected. In six cases the test demonstrates false positives by both human judges and RST analysis. The RST misidentified four stories as deceptive, whereas the human judges were inclined to characterize the stories as more truthful than deceptive. In two remaining stories (of the six false positives that were erroneously identified as deceptive), the human judges were also incorrect and considered them more deceptive than truthful.

This analysis raises the question of whether there are any particular features in the stories that can mislead both metrics and humans in correctly identifying a story as deceptive or truthful. In particular, the analysis reveals that

DOI: 10.1002/asi

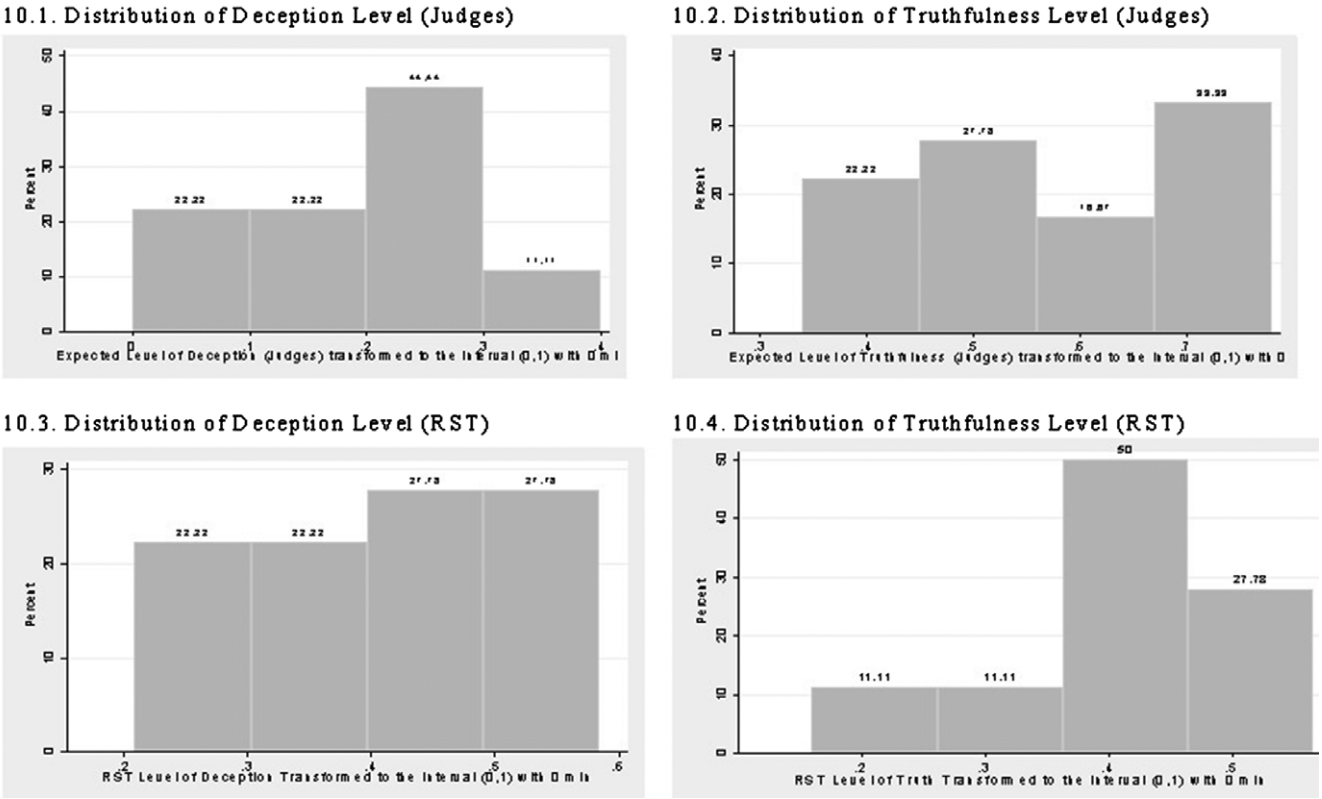

FIG. 10. Distribution of deception and truthfulness levels for truthful stories (second sample).

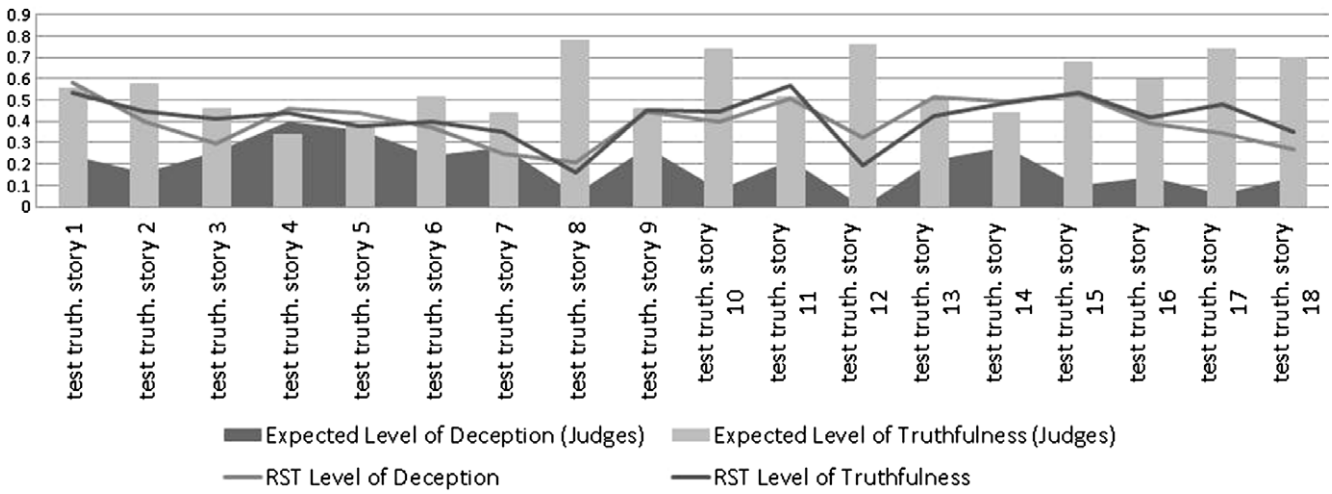

FIG. 11. Distributions of expected levels of deception and truthfulness in truthful stories (second sample).

there might be some systematic features of the truthful and deceptive stories that existing tools could misidentify. This demonstrates the need for more nuanced analysis of stories that are flagged as deceptive even though the stories are truthful and vice versa.

We suggest that a more nuanced deception detection analysis should be a two-stage process. The first stage of the deception detection process should continue using the existing automated tools of deception identification in written communication. These existing tools typically stop at this stage and report the results. We, however, propose to add a second, more nuanced stage of the analysis, which should focus only on those stories that were identified as truthful (if the goal is to determine deception) or as deceptive (if the goal is to determine truth). The second stage requires developing a method for identifying false positives by determining their specific features. For example, some of the false positives in our sample were stories describing extraordinary or

DOI: 10.1002/asi

strange events that would be erroneously interpreted as truthful or deceptive. The story below is a truthful story, which was identified by RST and humans as deceptive.

> I was doing an audio recording one day using my digital recorder and an external microphone. I was just talking out loud . . . expressing myself and talking about all of the things that were bothering me at the time. I downloaded the recording and started listening to it with headphones on. In between my speaking . . . during the lapses, I heard other voices . . . voices that were not heard at the time of the recording. The voices were varied and were distinctly commenting on the things I was saying out loud. Some of the things that I heard included, "she's sad" and "can we help her?" I belie [sic] these voices to be EVP "electronic voice phenomenon."

Perhaps written stories of extraordinary or unbelievable events have a distinguished set of deceptive/truthful features compared with stories about ordinary events. In a closer content analysis of these truthful stories, it becomes obvious that a reported event nested within an honestly narrated discourse seems rather strange and unbelievable. Perhaps a straightforward account of events, observations, or experiences (such as extraneous voices on recordings) should be distinguished from how the narrator interprets the events, explains causal relations, and states any conclusions.

This signals a warning of the need for more nuanced analysis of stories that are flagged as deceptive in the identification of truthful stories. This is in line with a general idea of deception detection tools, which at this point in their development are meant as assistive tools and can spot the cases that require human attention. On the other hand, such cases demonstrate that the RST method can pick out a lie nested within a truthful story.

The out-of-sample analysis indicates that the RST analysis provides rather good results in terms of determining whether the story is deceptive or truthful, even for a small sample. In addition, the RST analysis indicates the need to perform specialized analysis of false positives by developing a new method to identify their specific features. This means that RST-VSM approach is one of the potential tools for deception detection.

Discussion

The analysis of truthful stories shows some systematic and some slightly contradictory findings. On one hand, the levels of truthfulness assigned by judges are predominantly higher than the levels of deception. Again, assuming that the stories in the truthful set are completely true because their writers ranked them so, the human judges have greater likelihood of rating these stories as truthful than as deceptive. This can be an indicator of good precision of deception detection by human judges.

On the other hand, the RST-VSM analysis also demonstrates that a large subsample (but not as large as indicated by human judges) of truthful stories is closer to the truth center than to the deceptive one. However, it seems that the RST-VSM approach overestimates the levels of deception in the truthful stories compared with human judges.

Overall, however, the RST-VSM analysis demonstrates positive support for the proposed hypothesis. The apparent and consistent closeness of deceptive stories to the RST deception center (compared with the closeness of the deceptive stories to the truthful center) and truthful stories with the RST truthful center can indicate that the relations between discourse constituent parts differ between truthful and deceptive messages. Thus, because the truthful and deceptive relations exhibit systematic differences in RST space, the proposed RST-VSM method seems to be a valuable tool in deception detection. The results, however, have to be interpreted with caution, because the sample was very small, and only one expert carried out the RST coding.

The discussion, however, might be extended to the case in which the assumption of self-ranked levels of deception and truthfulness does not hold. In this case, we still suspect that even a deceptive story might contain elements of truth (though fewer), and a truthful story may have some elements of deception. RST-VSM analysis produced greater levels of deception in truthful and deceptive stories than did the human judges. This might indicate that RST-VSM potentially offers an alternative to human judges' way of detecting deception when it is least expected in text (as in the example of supposedly truthful stories) or detecting it in a more accurate way (if some level of deception is assumed, as in the completely deceptive stories). The advantage of the RST-VSM approach is its rigorous and systematic approach of coding discourse relations and their subsequent analysis in RST space by using vector space models. As a result, the relations among units exhibiting different degrees of salience in text because of writers' attempts to shape their subsequent readers' perceptions become indicators of diversity in deception levels.

Conclusions

Relations among discourse parts as well as the overall rhetorical structure seem to have different patterns in truthful and deceptive stories. If this is so, RST-VSM method can be an effective way of complementing the existing lexicosemantic analyses and thus a valuable tool in detecting deception.

The main findings in this research demonstrate that the proposed RST-based analysis resembles, to some degree, that of human judges of deceptive and truthful stories and that RST deception detection in self-rated deceptive stories has greater consistency than in truthful ones, which indicates the value of using the RST-VSM approach for deception detection.[14] However, RST conclusions regarding levels of deception in truthful stories require further investigation into the diversity of RST relations to express truths and deception

[14]The authors recognize that the results are preliminary and should be generalized with caution as a result of a small data set and certain methodological issues that require further development.

DOI: 10.1002/asi

as well as the types of clustering algorithms most suitable for clustering written communication in RST space without human judges' involvement and detecting deception with a certain degree of precision.

Our contribution to deception detection research and RST is threefold. (a) We demonstrate that discourse structure and pragmatics provide a promising basis for automated deception detection and an effective complement to lexicosemantic analysis. (b) We develop the novel RST-VSM method using RST analysis to identify previously unseen deceptive texts. (c) We suggest that a more accurate two-stage procedure with special methodology for detecting false positives could improve the deception detection process. The potential of this research lies in developing novel discourse-based tools to assist information seekers, online readers, or decision makers by alerting them to potential deception in computer-mediated texts. This, in turn, addresses several problems recognized in library and information science: discerning information from misinformation, screening information for its quality, and assessing the credibility of its resources.

Acknowledgments

This research was funded by the New Research and Scholarly Initiative Award (10–303) from the Academic Development Fund at the University of Western Ontario, London, Ontario, Canada. The authors thank the anonymous reviewers for their suggestions.

References

Bachenko, J., Fitzpatrick, E., & Schonwetter, M. (2008). Verification and implementation of language-based deception indicators in civil and criminal narratives. In D. Scott & H. Uszkoreit (Eds.), Proceedings of the 22nd International Conference on Computational Linguistics (COLING 2008) (pp. 41–48). Manchester, UK: COLING 2008 Organization Committee.

Baeza-Yates, R., & Ribeiro-Neto, B. (1999). Modern information retrieval. New York: ACM Press.

Ballou, D., & Pazer, H. (1995). Designing information systems to optimize the accuracy-timeliness tradeoff. Information Systems Research, 6, 51–72.

Berkhin, P. (2002). Survey of clustering data mining techniques. Technical Report. San Jose, CA: Accrue Software. doi: 10.1.1.18.3739

Bilmes, J.A. (1998). A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. University of California, Berkeley. Retrieved from http://www.icsi.berkeley.edu/ftp/pub/techreports/1997/tr-97-021.pdf

Buller, D.B., & Burgoon, J.K. (1996). Interpersonal deception theory. Communication Theory, 6, 203–242.

Cadez, I., V, Gaffney, S., & Smyth, P. (2000). A general probabilistic framework for clustering individuals and objects. In R. Ramakrishnan, S. Stolfo, R. Bayardo, & I. Parsa (Eds.), Proceedings of the 6th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD) (pp. 140–149). New York: ACM Press.

Delone, W.H., & McLean, E.R. (1992). Information systems success: The quest for the dependent variable. Information Systems Research, 3, 60–95.

DePaulo, B.M., Charlton, K., Cooper, H., Lindsay, J.J., & Muhlenbruck, L. (1997). The accuracy-confidence correlation in the detection of deception. Personality and Social Psychology Review, 1, 346–357.

Feng, V.W., & Hirst, G. (2012). Text-level discourse parsing with rich linguistic features. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL2012) (pp. 60–68). Jeju, Korea: Association for Computational Linguistics.

Fox, C.J. (1983). Information and misinformation: An investigation of the notions of information, misinformation, informing, and misinforming. Westport, CT: Greenwood Press.

Frank, M.G., Paolantinio, N., Feeley, T.H., & Servoss, T.J. (2004). Individual and small group accuracy in judging truthful and deceptive communication. Group Decision and Negotiation, 13, 45–59.

Fuller, C.M., Biros, D.P., & Wilson, R.L. (2009). Decision support for determining veracity via linguistic-based cues. Decision Support Systems, 46, 695–703.

Goodhue, D.L. (1995). Understanding user evaluations of information systems. Management Science, 41, 1827–1844.

Hair, J.F., Anderson, R.E., Tatham, R.L., & Black, W.C. (1995). Multivariate data analysis with readings. Upper Saddle River, NJ: Prentice-Hall.

Hancock, J.T., Curry, L.E., Goorha, S., & Woodworth, M. (2008). On lying and being lied to: A linguistic analysis of deception in computer-mediated communication. Discourse Processes, 45, 1–23.

Hernault, H., Prendinger, H., duVerle, D.A., & Ishizuka, M. (2010). HILDA: A discourse parser using support vector machine classification. Dialogue and Discourse, 1, 1–33.

Indyk, P. (1999). A sublinear time approximation scheme for clustering in metric spaces. In R. Ostrovsky (Ed.), Proceedings of the 40th Annual Symposium on Foundations of Computer Science. (IEEE FOCS 40) (pp. 144–149). Washington, DC: IEEE Computer Society.

Jarke, M., & Vassiliou, Y. (1997). Data warehouse quality: A review of the dwq project. Paper presented at Conference on Information Quality, Cambridge, MA.

Karypis, G. (2002). CLUTO: A clustering toolkit. Minneapolis: University of Minnesota, Computer Science Department.

Lin, Z., Kan, M.-Y., & Ng, H.T. (2009). Recognizing implicit discourse relations in the Penn Discourse Treebank. In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP2009), 1 (pp. 343–351). Stroudsburg, PA: Association for Computational Linguistics.

Lukoianova, T., & Rubin, V.L. (2014). Veracity roadmap: Is big data objective, truthful and credible? Advances in Classification Research Online , 24(1). doi: 10.7152/acro.v24i1.14671

Mann, W.C., & Thompson, S.A. (1988). Rhetorical structure theory: Toward a functional theory of text organization. Text, 8, 243–281.

Manning, C.D., & Schütze, H. (1999). Foundations of statistical natural language processing. Cambridge, MA: MIT Press.

Marcu, D. (1997). The rhetorical parsing of natural language texts. In Proceedings of the ACL/EACL (pp. 96–103). Madrid.

Marcu, D., & Echihabi, A. (2002). An unsupervised approach to recognizing discourse relations. In Proceedings of the ACL (pp. 368–375). Philadelphia.

Mihalcea, R., & Strapparava, C. (2009). The lie detector: explorations in the automatic recognition of deceptive language. In Proceedings of the ACL (pp. 309–312). Singapore. Retrieved from http://www.aclweb.org.

Pennebaker, J.W., Chung, C.K., Ireland, M., Gonzales, A., & Booth, R.J. (2007). The development and psychometric properties of LIWC2007. Austin, TX. Retrieved from http://www.liwc.net

Porter, S., & Yuille, J.C. (1996). The language of deceit: An investigation of the verbal clues to deception in the interrogation context. Law and Human Behavior, 20, 443–458.

Rasmussen, M., & Karypis, G. (2004). gCLUTO: An interactive clustering, visualization and analysis system. UMN-CS TR-04–021.

Rieh, S.Y. (2010). Credibility and cognitive authority of information. In M. Bates & M.N. Maack (Eds.), Encyclopedia of library and information sciences (3rd ed., pp. 1337–1344). New York: Taylor and Francis.

Rose, K. (1998). Deterministic annealing for clustering, compression, classification, regression, and related optimization problems. Proceedings of the IEEE, 86, 2210–2239.

Rubin, V.L. (2010). On deception and deception detection: Content analysis of computer-mediated stated beliefs. In Proceedings of the American Society for Information Science and Technology Annual Meeting, 47, 1–10. doi: 10.1002/meet.14504701124

Rubin, V.L., & Chen, Y. (2012). Information manipulation classification theory for LIS and NLP. In Proceedings of the American Society for Information Science and Technology Annual Meeting, 49, 1–5. doi: 10.1002/meet.14504901353

Rubin, V.L., & Conroy, N. (2011). Challenges in automated deception detection in computer-mediated communication. In Proceedings of the American Society for Information Science and Technology Annual Meeting, 48, 1–4. doi: 10.1002/meet.2011.14504801098

Rubin, V.L., & Conroy, N. (2012). Discerning truth from deception: human judgments and automation efforts. First Monday, 17(3). Retrieved from http://firstmonday.org/ojs/index.php/fm/article/view/3933

Rubin, V.L., & Vashchilko, T. (2012). Extending information quality assessment methodology: Information veracity measure. In Proceedings of the American Society for Information Science and Technology Annual Meeting, 49(1), 1–6. doi: 10.1002/meet.14504901350

Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.

Scholkopf, B., & Smola, A. (2001). Learning with kernels. Cambridge, MA: MIT Press.

Strehl, A., Ghosh, J., & Mooney, R. (2000). Impact of similarity measures on web-page clustering. In K. Bollacker (Ed.), Proceedings of the AAAI 2000 Workshop on Artificial Intelligence for Web Search (pp. 58–64). Palo Alto, CA: AAAI.

Taboada, M. (2004). Building coherence and cohesion: Task-oriented dialogue in English and Spanish. Amsterdam: Benjamins.

Taboada, M., & Mann, W.C. (2006). Applications of rhetorical structure theory. Thousand Oaks, CA: Sage Publications.

Tombros, A. (2002). The effectiveness of query-based hierarchic clustering of documents for information retrieval (Doctoral thesis, Department of Computing Science, University of Glasgow).

Vapnik, V. (1998). Statistical learning theory. New York: Wiley.

Vartapetiance, A., & Gillam, L. (2012). Deception detection for the tangled web. Special Issue of ACM SIGCAS Computer and Society.

Vrij, A. (2000). Detecting lies and deceit. New York: Wiley.

Wand, Y., & Wang, R.Y. (1996). Anchoring data quality dimensions in ontological foundations. Communications of the ACM, 39(11), 86–95.

Wang, R.Y., & Strong, D.M. (1996). Beyond accuracy: What data quality means to data consumers. Journal of Management Information Systems, 12(4), 5–34.

Zhong, S., & Ghosh, J. (2003). Scalable, Balanced Model-Based Clustering. In D. Barbara & C. Kamath (Eds.), Proceedings of the Third SIAM International Conference on Data Mining (pp. 71–82). San Francisco, CA.

Zhou, L., Burgoon, J.K., Nunamaker, J.F., & Twitchell, D. (2004). Automating linguistics-based cues for detecting deception in text-based asynchronous computer-mediated communications. Group Decision and Negotiation, 13, 81–106.

Zmud, R.W. (1978). An empirical investigation of the dimensionality of the concept of information. Decision Sciences, 9, 187–195.