Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 5906–5915
Marseille, 20-25 June 2022
© European Language Resources Association (ELRA), licensed under CC-BY-NC-4.0

Rhetorical Structure Approach for Online Deception Detection: A Survey

Francielle Vargas*†, Jonas D'Alessandro∝, Zohar Rabinovich‡
Fabrício Benevenuto†, Thiago A. S. Pardo*
*Institute of Mathematical and Computer Sciences, University of São Paulo, Brazil
‡Viterbi School of Engineering, University of Southern California, United States
†Computer Science Department, Federal University of Minas Gerais, Brazil
∝Linguistics Department, Federal University of Minas Gerais, Brazil
francielleavargas@usp.br, jonasd@let.grad.ufmg.br, zoharrab@usc.edu
fabricio@dcc.ufmg.br, taspardo@icmc.usp.br

Abstract

Most information is passed on in the form of language. Therefore, research on how people use language to inform and misinform, and how this knowledge may be automatically extracted from large amounts of text is surely relevant. This survey provides first-hand experiences and a comprehensive review of rhetorical-level structure analysis for online deception detection. We systematically analyze how discourse structure, aligned or not with other approaches, is applied to automatic fake news and fake reviews detection on the web and social media. Moreover, we categorize discourse-tagged corpora along with results, hence offering a summary and accessible introductions to new researchers.

Keywords: rhetorical structure theory, discourse-level structure analysis, online deception detection, natural language processing.

1. Introduction

Assessing whether reported statements are intentionally misstated (or manipulated) is of considerable interest to researchers, financial companies, security, and governmental regulators (Larcker and Zakolyukina, 2012; Larcker and Zakolyukina, 2012). According to Hancock and Guillory (2015), there are reliable cues for deception detection, and the belief that liars give cues that may indicate their deception is nearly universal. Moreover, the most relevant literature on deception suggest that liars may be identified by their words (Newman et al., 2003; DePaulo et al., 2003), and a fairly straightforward element for mitigating risks of deceptive activities is to identify deceptive intentions (Ho and Hancock, 2019).

For the last few years, there has been a growth in the number of web and social media users; consequently, the potential for deceptive activities has also increased, such as fake news, fake reviews (also known as opinion spam), deceptive discussion, and simple lies. Particularly, fake news detection is defined as the prediction of the chances of a news article being intentionally deceptive (Rubin et al., 2015). Fake reviews - also known as opinion spam - are inappropriate or fraudulent reviews (Li et al., 2014; Ott et al., 2011). Deceptive discussions consist of intentionally misstated (or manipulated) narratives or statements (Larcker and Zakolyukina, 2012).

Notwithstanding the lack of discourse processing for deceptive detection, more recently, discourse frameworks such as Rhetorical Structure Theory (RST) (Mann and Thompson, 1987; Thompson and Mann, 1988) have been adopted to automatically detect deceptive stories. Figure 1 shows an example of a deceptive story (fake news) annotated using the RST framework.

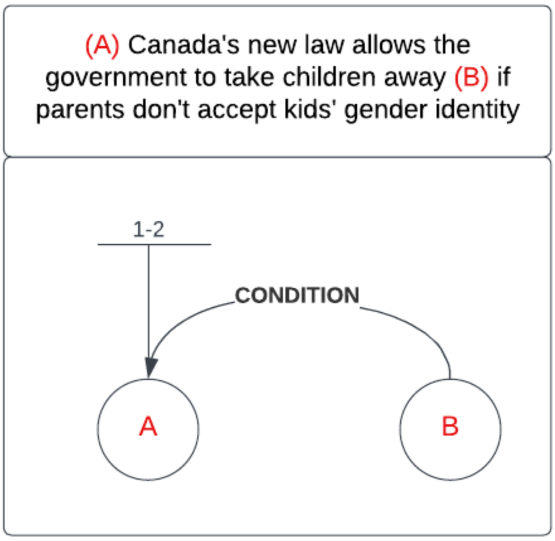

(A) Canada's new law allows the government to take children away (B) if parents don't accept kids' gender identity
1-2
CONDITION
A
B

Figure 1: Discourse-tagged fake news using RST framework. This fake news was extracted from Politifact stated on April 24, 2022 in a Instagram post.

As shown in Figure 1, a deceptive story [1] was segmented into two elementary discourse units (EDUs), and the relationship among them uses the CONDITION coherence relation. Furthermore, there are spans of texts that are more central (nucleus) to the text's purpose than others (satellite). The nucleus is signaled by the direction of the arrow. And, although there is a kind of relation between nucleus and satellite (mononuclear), we also find a relation with two nuclei (multinuclear).

[1] https://tinyurl.com/ytdff5ep

Since deceiving action requires advanced cognitive development and mechanisms that honesty simply does not require, research on people's cognitive mechanisms of deception detection has promising guidance for the detection and refutation of fake content on the web and social media (Kumar and Geethakumari, 2014). Besides, regarding that deceptive stories also lack "evidence", a very plausible assumption would be that coherence relations may be an efficient strategy. Nevertheless, despite the acknowledged potential of discourse analysis as a cognitive approach to detect deceptive activities, there is a considerable lack of research on deceptive stories in this field. We do not have significant knowledge of embedded lies in texts or discourses (Meibauer, 2018), with the notable exception of the studies proposed by Galasiński (2000) and Meibauer and Dynel (2016); nonetheless, these studies deal with fictional discourse in American television shows. Thus, there is a lack of research on deception in non-fictional discourse and empirical research at the discourse-level analysis of deceptive texts. A plausible assumption to explain this research gap would be that the investigation concerning the discursive structure of lies is challenging.

In this study, we embrace the challenges and opportunities of the application of the RST and discourse-level analysis for deception detection, whereas it provides the first survey that encompasses works addressing the RST discourse framework to tackle the problem of deceptive activities on the web and social media. Although the discourse-aware approach may be applied in a wide variety of deceptive activities, until the present moment, it was only applied for fake news and fake review detection tasks.

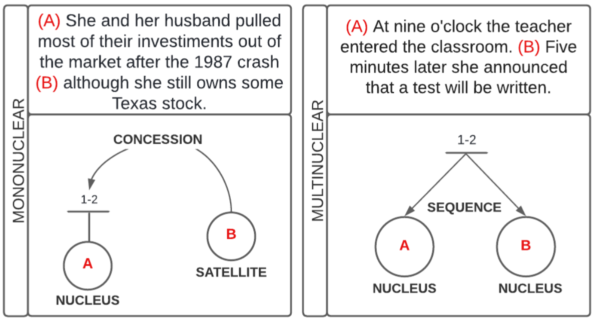
In what follows, we present in Sections 2 and 3 the main definitions related to RST and Deception Detection. Section 4 introduces a summarized section containing corpora, models and methods of the literature that applied RST and discourse-level structure for fake news or fake reviews detection. A detailed description of them is also presented. In Section 5, we discuss the main challenges of rhetorical structure approach for online deception detection. Finally, in Section 6, conclusions are presented.

2. Rhetorical Structure Theory

RST is a relevant framework in Artificial Intelligence dealing with Computational Linguistics at discourse-level structure analysis. According to Mann and Thompson (1987), RST consists of a theory to help us to understand texts as instruments of communication. Therefore, RST provides a consistent framework for investigating relational propositions, which are unstated but inferred propositions that arise from the text structure in the process of interpreting texts. RST relies on three mechanisms that are central: *nuclearity*, *schemas*, and *coherence relations* (Thompson and Mann, 1988).

2.1. Nuclearity

Nuclearity consists of the identification of prominent and complementary text spans. Organizing them hierarchically within the schema encompasses four main concepts: *nucleus* and *satellite*, as well as *mononuclear* and *multinuclear* coherence relations. The nucleus consists of text elements that are more central and relevant in the relation. In contrast to nucleus, the supporting units are called satellites. These drift towards mononuclear and multinuclear coherence relations. An example is shown in Figure 2.


(A) She and her husband pulled most of their investments out of the market after the 1987 crash (B) although she still owns some Texas stock.

(A) At nine o'clock the teacher entered the classroom. (B) Five minutes later she announced that a test will be written.

Figure 2: Nuclearity structure in mononuclear and multinuclear coherence relations.

As exemplified by Figure 2, in the first example, there is a span that is more relevant for understanding the text. Consequently, this type of coherence relation is mononuclear, and the nucleus is signaled by an arrow. In contrast, in the second example, both spans present the same relevance for understanding the text. This type of coherence is multinuclear, in which both spans are considered nuclei and signaled by arrows.

2.2. Schemas

A schema is defined as predefined patterns specifying how regions of text combine to form larger regions, until to whole texts. Figure 3 shows five types of schemas originally proposed by this theory.

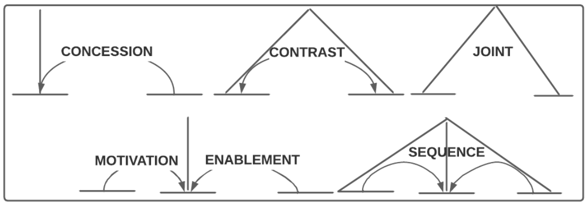
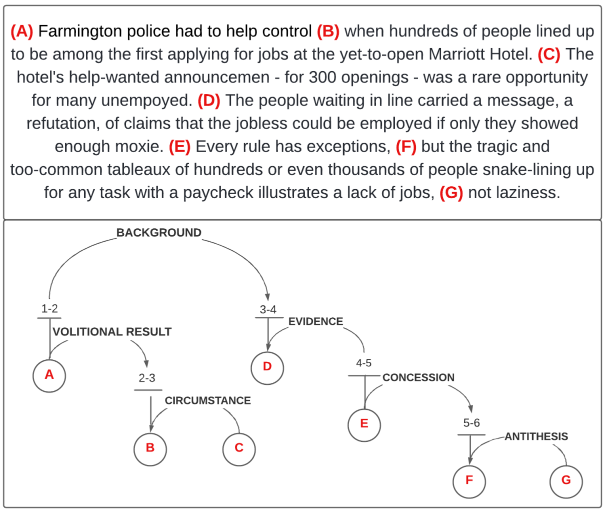


Figure 3: Five different schemas according to RST.

As it is shown in Figure 3, a schema is characterized by a vertical line pointing to one of the text spans that the schema covers, titled "nucleus". The other spans are linked to the nucleus by relations, represented by labeled curved lines, and these spans are titled "satellites".

2.3. Coherence Relations

RST predicts the construction of a tree of coherence relations (also known as rhetorical or discourse relations), which is mainly based on the premise that the content of text units may be hierarchically organized. Accordingly, RST assumes that some units are more central (nucleus) to the text than others (satellite). Coherence relations also are described in terms of schemas (i.e., how one or more satellites or nuclei associated with each other). For instance, observe the RST tree shown in Figure 4.



(A) Farmington police had to help control (B) when hundreds of people lined up to be among the first applying for jobs at the yet-to-open Marriott Hotel. (C) The hotel's help-wanted announcemen - for 300 openings - was a rare opportunity for many unempoyed. (D) The people waiting in line carried a message, a refutation, of claims that the jobless could be employed if only they showed enough moxie. (E) Every rule has exceptions, (F) but the tragic and too-common tableaux of hundreds or even thousands of people snake-lining up for any task with a paycheck illustrates a lack of jobs, (G) not laziness.

Figure 4: Rhetorical structure tree.

As it is shown in Figure 4, the span G (satellite) is connected with the span F (nucleus) using the ANTITHESIS coherence relation. In the same settings, the span C is connected with the span B (nucleus) using the CIRCUMSTANCE relation. Furthermore, the span E (satellite) is connected with the block 5-6, which is composed by the spans [F-G] (nucleus) using the CONCESSION relation. The span D (nucleus) is connected with the block 4-5, which is composed by the spans [E-[F-G]] (satellite) using the EVIDENCE relation. Then, the span A (satellite) is connected with the block 2-3, which is composed by the spans [B-C] (nucleus) using the CIRCUMSTANCE relation. At last, the block 1-2, which is composed by the spans [A1-[A2-A3]] (nucleus) is connected with the block 3-4, which is composed by the spans [A4-[A5[A6-A7]]] (satellite) using the BACKGROUND relation.

3. Online Deception Detection

Over the past year, on the sharp growth of the web and social media, cyber-crimes such as identity blows, thief, fraud, and misinformation have become increasingly common. Theses deceptive activities often are characterized by the ease of deception and concealment of one's real identity (Pérez-Rosas et al., 2017). The research area responsible for investigating and providing methods to detect deceptive activities is known as deception detection. According to Rubin et al. (2015), automated deception detection, as a field within NLP and Information Science (IS), is responsible for the development of methods to distinguish truth from deception in textual data, identifying linguistic predictors of deception with text processing and machine learning techniques. Deception detection in textual information has became a relevant study area within NLP, mainly due to the sharing of fake news on the web and social media around the world.

Online deceptive activities are addressed by literature on different tasks, which handle a wide range of aspects, such as credibility of users and sources, information veracity, information verification, and linguistic aspects of deceptive language (Atanasova et al., 2019). Unless otherwise stated, these tasks include the discovery of fake news (Lazer et al., 2018); rumor detection in social media (Vosoughi et al., 2018); information verification in question answering systems (Mihaylova et al., 2018); detection of information manipulation agents (Chen et al., 2013; Mubarak et al., 2020); assertive technologies for investigative journalism (Hassan et al., 2015); detection of fake reviews (Ott et al., 2011); detection of deceptive discussions (Larcker and Zakolyukina, 2012).

A definition with relevance for the area rotates around the concept of "deceptive language". Deceptive language is defined by Communication, Linguistics and Psychology literature as a type of language deliberately used with aim of attempting to mislead others. For instance, falsehoods communicated by people who are mistaken or self-deceived are not lies, nevertheless, literal truths designed to mislead are lies as a deliberate attempt to mislead others. Besides that, most relevant literature on deception refers mainly to levels of deceit and typology of media (e.g., face-to-face, voice, text) (Zhou et al., 2003). DePaulo et al. (2003) claim that deceptive linguistic style may present weak employment of singular and third-person pronouns, negative polarity, and high employment of movement verbs. Nahari et al. (2019) suggest that a basic assumption related to deceptive language is that liars differ from truth-tellers in their verbal behavior, making it possible to classify them by inspecting their verbal accounts. Additionally, a set of linguistic behaviors may predict deception, as tones of words and kinds of preposition, conjunctions, and pronouns (Newman et al., 2003).

Taking advantage of the discourse-level analysis, Galasińki (2000) presents a pioneer study on fictional deceptive stories. According to the author, discourse analysis of deceptive texts deception is intrinsically tied with "information manipulation", which consists of presenting a reality that is misrepresented. The author argues that deception should be classified in three different levels: (i) *falsification* (i.e., attributing false statements to a debater), (ii) *distortion* (i.e., manipulating by understating or overstating what a debater states), and (iii) *de-contextualization* (i.e., taking the words a debater uses out of their context).

4. Discourse-Aware Deception Detection

In this section, we summarize and categorize discourse-aware deception detection corpora, models and methods (see Section 4.1). Moreover, we also describe in detail the proposals of literature that address RST and discourse-level structure for deception detection, more specifically, for fake news and fake reviews detection (see Sections 4.2 and 4.3).

4.1. Corpora, Models and Methods

A very plausible assumption, when one opts for the discourse-aware approach applied to deception detection, would be that there are significant differences between structures of truthful and deceptive stories. Indeed, it has been proposed by various authors. While the research community currently lacks discourse annotated corpora for deception detection tasks, recent works have proposed discourse-tagged corpora for the English, Portuguese and Russian languages. Table 1 provides a summary of the discourse-tagged corpora proposed in literature.

Table 1: Discourse-tagged corpora overview for the fake news and fake reviews detection tasks.

Authors	Total	Classes	Lang.	Type	Task
Vargas et al. (2021)	600	300-Fake 300-Truth	Portuguese, English	Mult	Fake News
Pisarevskaya (2017)	174	87-Fake 87-Truth	Russian	Mono	Fake News
Popoola (2017)	50	25-Fake 25-Truth	English	Mono	Fake Reviews
Rubin et al. (2015)	132	66-Fake 66-Truth	English	Mono	Fake News

As it is shown in Table 1, the discourse-tagged corpora for the fake news and fake reviews detection tasks were proposed for the English, Russian, and Portuguese languages. As being particularly a human time-onerous task and a kind of challenging annotation process, the corpora present a small set of documents. Furthermore, both monolingual and multilingual corpora were proposed.

Moving forward, as it is known from research proposals on fake news and fake reviews, a wide variety of models have been proposed to tackle online deception detection. Most of them rely on linguistic features such as n-grams, language complexity, part-of-speech tags, and syntactic and semantic features. On the other hand, discourse-level structure approach is usually framed as a supervised learning problem, which embodies in a model coherence relations followed by hierarchical nuclearity information to build automatic classifiers. In Table 2, we also summarize discourse-aware models and methods proposed in literature. Notice that models use bag-of-rst, dependency parsing, embeddings and BERT tokenizer as features, and both classical and neural machine learning have been applied. Finally, f1-score performance is reported in column "%", except for Karimi and Tang (2019), whose authors reported values related to accuracy.

Table 2: Discourse-aware models and methods for fake news detection.

Authors	Set of Features	ML Method	Lang	Fscore	Task
Kuzmin et al. (2020)	RuBERT, Bag-of-rst	BERT, SVM, LR	Russian	90%	Fake news
Karimi and Tang (2019)	Dependency parsing	LSTM	English	82%	Fake news
Atanasova et al. (2019)	Embeddings	LSTM	English	69%	Fake news
Pisarevskaya (2017)	Bag-of-rst	SVM, Random Forest	Russian	65%	Fake news
Rubin et al. (2015)	Bag-of-rst	SVM	English	63%	Fake News

4.2. Fake News Detection

Kuzmin et al. (2020)

Fake news prediction is a global problem, and most of approaches have been developed for the English language (Kuzmin et al., 2020). Nevertheless, fake news is spread around the world, and it may be written originally in several languages. In this proposal, the authors trained and compared different models for fake news detection in Russian. They assess whether different language-based features including the vectorization of rhetorical structure obtained from both - a RST parsing and a rst manually annotated corpus - could be helpful for the fake news detection task. This proposal was implemented and evaluated using classical machine learning methods, as Support Vector Machine (SVM) and Logistic Regression (LR) over bag-of-n-grams and bag-of-rst representations. Besides that, sophisticated machine learning techniques, as BERT (Devlin et al., 2019) were also implemented. The authors used three different corpora of fake news in Russian. The first one was proposed by Pisarevskaya (2017) (see Table 1 - manually annotated). The second one was proposed by Zaynutdinova et al. (2019); it is composed of 1,366 fake news and 7,501 true news. Finally, Taiga Corpus [2] was also applied. Furthermore, three distinct representations were used (i) bag-of-ngrams with tf-idf preprocessing, (ii) bag-of-rst, which consists of the vectorization of coherence relations and nuclearity, and (iii) pre-trained BERT-based model, more specifically, the RuBERT2 obtained using DeepPavlov [3] (Burtsev et al., 2018) with Transformers (Wolf et al., 2020). The authors reported that classical approaches using bag-of-n-grams and bag-of-rst presented high results (90% of F1-score) overcoming the neural network approach, which uses the RuBERT ((88% of F1-score). Moreover, the authors suggest that satire is similar to fake news, and satire differs from real news. The authors also concluded that humans rarely perform better than chance at detecting deceptive activities. Therefore, humans performed worse than the best automated model.

[2] https://tatianashavrina.github.io/taiga_site/

[3] https://deeppavlov.ai/

Karimi and Tang (2019)
Discourse-level structure analysis of deceptive and truthful news is a tremendous challenge, mainly due to existing methods for capturing discourse-level structure rely on annotated corpora, which are not available for fake news datasets (Karimi and Tang, 2019). In this proposal, the authors provide a new *dependency parsing approach*, titled "**H**ierarchical **D**iscourse-level **S**tructure for **F**ake news detection". The HDSF consists of an automated manner to learn a discourse-level structure for a given document through an approach based on the dependency parsing at the sentence level. It should be noted that in this approach, sentences are classified as elementary discourse units (EDU's). An example of discourse-level structure of a document (fake news) using the proposed dependency tree is shown in Figure 5. Note that a document is segmented into sentences (S1, S2, S3, S4 and S5), and hierarchically organized.

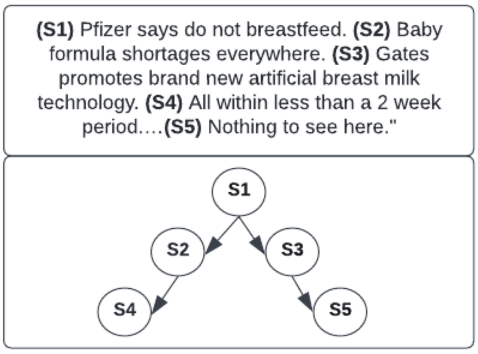

(S1) Pfizer says do not breastfeed. **(S2)** Baby formula shortages everywhere. **(S3)** Gates promotes brand new artificial breast milk technology. **(S4)** All within less than a 2 week period….**(S5)** Nothing to see here."

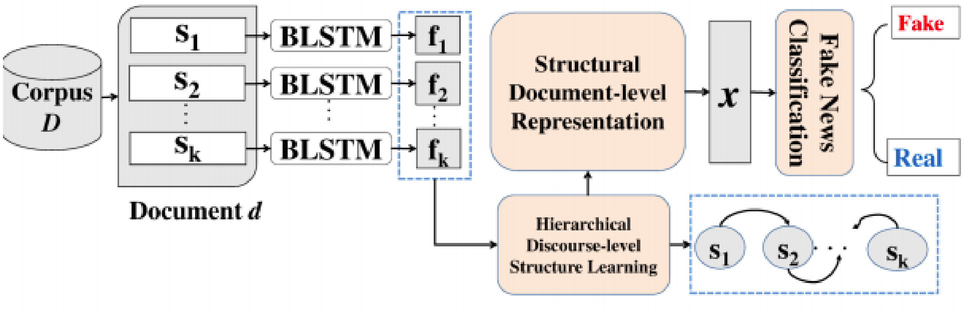
Figure 5: Hierarchical discourse-level structure of a document using a dependency tree. This fake news was extracted from Politifact.

The HDSF framework build a hierarchical structure between sentences without relying on an annotated corpus, as may be seen in Figure 6. Note that the HDSF receives as input a corpus of fake/real news documents (i.e., D). A model M may automatically learn hierarchical and structurally rich representations for documents in D. Meanwhile, given binary labels Y, model M uses the hierarchical representations to automatically predict the labels of unseen news documents.
In order to compare the HDSF approach with baseline and state-of-art models, the authors implemented seven different models including the proposed methods: N-grams, LIWC (Pennebaker et al., 2015), Bag-of-rst (Rubin and Lukoianova, 2015), BiGRNN-CNN (Ren and Zhang, 2016), LSTM and LSTM[w+s] (Karimi and Tang, 2019). Based on the obtained results, the HDSF overcame the other implemented approaches (82.19% of Accuracy). They concluded that discourse-level structure analysis is effectively rich for fake news prediction. In addition, the structures of fake news documents at the discourse level are substantially different from those of true ones, and real news documents indicate more degree of textual coherence.

Atanasova et al. (2019)

In this proposal, the authors focus on contextual and discourse-level structure information, which, according to them, provide important information that is typically not found over usual feature sets. The authors model the problem of fake news detection into two main tasks: (i) *claim classification*, which consists of automated identification of claims in political debates that a journalist should fact-check, and (ii) *answer fact-checking*, which consists of automatic verification of political answers in community-driven Web forums. They implemented an extensive block of experiments for both tasks using both classical and neural machine learning methods. The datasets used were: CW-USPD-2016 (Gencheva et al., 2017), which is annotated at the sentence level as check-worthy or not, and the context of the full debate was kept. it provides a binary annotation: whether a sentence was annotated for factuality by a given fact-checking, and composed of 4,5355 positive documents, and 880 negative documents; and CQA-QL-FACT (Nakov et al., 2016), which consists of a dataset composed by (i) a good vs. a bad answers, and (ii) a factually true vs. a factually false one. CQA-QL-FACT dataset provides 373 answers classified as factual, 689 answers classified as opinion, and 295 answers classified as socializing.

As previously stated, the authors propose methods for two different tasks: claim identification and answer fact-checking. For claim identification, a robust neural model that embodies a set of rich contextual and discourse features was proposed. Figure 7 shows the proposed models. A RST-based discourse parser (Joty et al., 2015) was used to obtain rhetorical structure features. As displayed in Figure 7, each segment is defined as a "maximal set of consecutive sentences by the same speaker, without intervention of another speaker or the moderator". In addition, the authors use a feed-forward neural network (FNN) with two hidden layers (with 200 and 50 neurons, respectively) and a softmax output unit for the binary classification. ReLU (Glorot et al., 2011) was used as the activation function and training happened for 300 epochs with a batch size of 550. They set the L2 regularization to 0.0001 and kept a constant learning rate of 0.04. On the other hand, for the answer fact-checking task, the authors built an interesting and robust model. The model combines an LSTM-based neural network with support vector machines to classify three question categories (factual, opinion, and socializing), as shown in Figure 8. Notice that a layer of pre-trained embeddings, and the blind layers are used in order to supplement the other features proposed by the authors (e.g., web support, ql support, similarities, etc.). Finally, the best performance for both tasks was obtained with the model that embodies discourse and contextual features as a supplement to other features (69% of F1-Score). Therefore, the authors concluded that contextual and discourse information may improve the performance of fake news detection systems.



Corpus D
S1
S2
Sk
Document d
BLSTM
BLSTM
BLSTM
f1
f2
fk
Structural Document-level Representation
x
Fake News Classification
Fake
Real
Hierarchical Discourse-level Structure Learning
s1
s2
sk

Figure 6: HDSF framework for fake news detection.

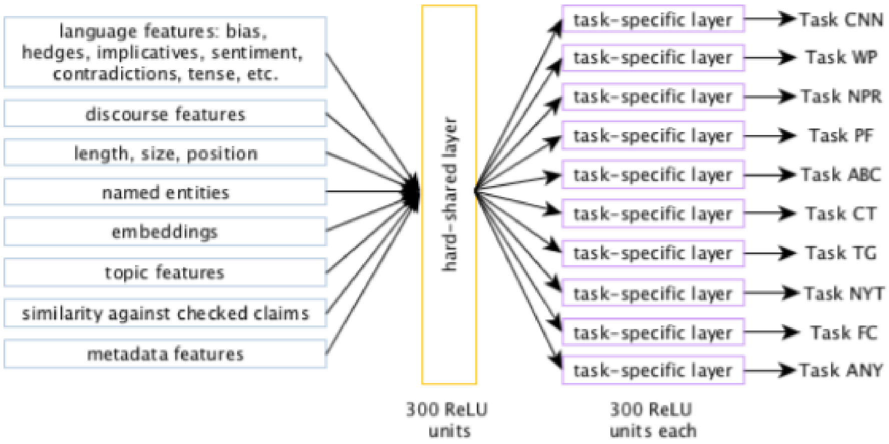
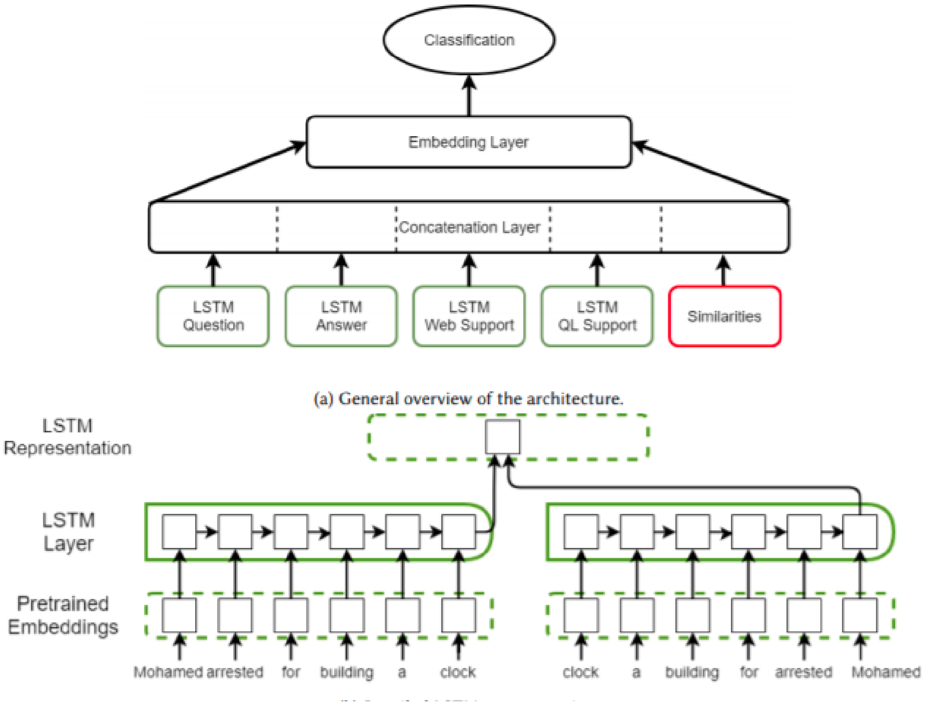


Figure 7: The neural architecture used for claim classification.



(a) General overview of the architecture.

Figure 8: Detailed LSTM architecture used for answer fact-cheeking classification.

Pisarevskaya (2017)
In this research, the author reiterates the importance of understanding the difference between true and fake news evaluating the reliability of sources, mainly due to fact that large-scale data that have been shared daily on the web and social media. According to author, new methods for deception detection and information verification must be created for different languages. Accordingly, this proposal consists of investigating the suitability of RST-based coherence relation features in order to build a deception detection model for fake news detection in Russian. While this proposal is inspired by research of Rubin et al. (2015) for the English language, the author has also considered the linguistic distinctions between the English and Russian languages. At first, the authors proposed a new discourse-level manually annotated corpus using RST framework. For data collection, news stories were manually analyzed in retrospect, when the factuality was already known, and fake stories were classified with a negative class (0) and truthful stories were classified with a positive class (1). According to the author, towards class balancing, the texts were collected from different sources: well-known news agencies' websites, local or topic-based news portals, online newspapers from different countries (Russia, Ukraine, Armenia, etc.). Blog texts, social media content, news reports based on opinions (not on facts) were excluded from the data. Over the annotation process, the author reports the average number of rhetorical relations for each document such as 17.43. In the same settings, the reported total of rhetorical relations in the corpus is equal to 2.340. Clauses were taken as elementary discourse units (EDU's). In order to support the annotation process, an accurate RST guideline was proposed towards minimize the problem of subjectivity of annotators' interpretation. Moreover, an evaluation measure was applied, obtaining a human agreement of 75%. An overview of this annotated corpus is shown in Table 1. Moving forward, a subjective lexicon-based analysis was also performed. More specifically, this analysis consists of assessing behavior of positive and negative lemmatized words using a lexicon composed of 5,000 sentiment words from reviews devoted to various topics. Consideration the building of the model, RST coherence relations occurrences and their respective nuclearity were represented as features into a machine learning-based model. The authors titled this representation as "bag-of-rst-relation-types". Support Vector Machine (Scholkopf and Smola, 2001) and Random Forest (Breiman, 2001) were used as learning methods. The results reported by the author are quite incipient. The best obtained performance reached F1-score of 65%. The author also proposed the evaluation of human performance to classify deceptive and truthful stories in Russian, whose results evidenced that human performance is highly unsatisfactory (50% of F1-score) and worse than the best performance automated classification.

Rubin et al. (2015), Rubin and Lukoianova (2015)
This proposal is the first one that uses RST applied to deception detection. Therefore, it may be considered a baseline method. The authors examined the rhetorical structure, discourse constituent parts, and their coherence relations for deceptive (fabricated) and truthful (authentic) news to uncover systematic language differences and inform deception verification systems. The proposed approach for fake news detection using RST-annotated corpus was performed using a dataset of 132 documents, with an equal amount of deceptive and non-deceptive news. An overview of this dataset is exhibited in Table 1. The data was collected from the US National Public Radio (NPR)[4], during the period of March 2010 to May 2014, and contains transcripts of the weekly radio show "Wait, Wait, Don't Tell Me" with its "Bluff the Listener". According to the authors, most news reports are typically humorous and a set of them are highly unlikely or unbelievable (e.g., a ship captain plotting his ship's course across land or a swim instructor not knowing how to swim). Moreover, the corpus was manually annotated by two analysts using RST. The authors report a human inter-annotator agreement of 69% using Cohen's kappa. In this proposal, an automated news verification method using RST and Vector Space Modeling (VSM) was proposed. The method titled "RST-VSM approach" applies the RST towards discourse analysis of true and fake news, and VSM in order to interpret the discourse features. The RST-VSM proposed approach was divided into three different experiments: (i) *centering*, (ii) *clustering*, and (iii) *predictive model*. The first experiment - centering - consists of the VSM representation used to assess each news report's position in a multi-dimensional RST space. Clustering of truthful and deceptive data points in this space was evaluated based on distances to hypothetical cluster centers. The authors obtained relevant differences between truthful and deceptive centers for each set of rhetorical relations. The second experiment - clustering - consists of a clusterization process of fake news and true news based on their similarity according to a chosen agglomerative clustering algorithm, with k-nearest neighbor clustering. They used the gCLUTO clustering package [5]. As a result of this experiment, four similarity clusters were formed. The clustering model was able to correctly assess 63% (20 out of 32 stories). The third experiment - predictive model - consists of a logistic regression model based on the training lumped dataset. Based on the performed experiments the authors report that (i) four logistic regression indicators were identified (from a set of 18) pointed to truth (DISJUNCTION, PURPOSE, RESTATEMENT, SOLUTIONHOOD), while another predictor (CONDITION) pointed to deception. Finally, the proposed approach RST-SVM obtained 63% of accuracy.

[4] https://www.npr.org/

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

4.3. Fake Reviews Detection

Popoola (2017)

In this proposal, the author analyzes RST coherence relations on a forensic collection of authentic and fake Amazon book reviews. The author concludes that paid review writers deploy deceptive pragmatics (i.e., a coherent set of linguistic strategies) to support the intent to deceive. At aiming to analyze deceptive and true intentions from reviews, the author annotated fifty reviews classified equally in fake and real reviews from the DeRev corpus (Fornaciari and Poesio, 2014). This corpus was collected of 6,819 Amazon book reviews of 68 books written by 4,811 different reviewers. A complete corpus overview is shown in Table 3.

Table 3: DeRev-RST corpus (Popoola, 2017).

Description	All Reviews	Real	Fake
Number of words	4,931	2,222	2,709
Average number of words per review/stdev	98.6/40.7	88.9/43.2	108.4/36.3
Number of RST coherence relations annotated	490	239	251
Average number of relations per review/stdev	9.8/5.0	9.6/5.6	10.0/4.4
Average "words per relation"/stdev	10.7/2.6	10.0/2.9	11.3/2.1

In this proposal, the author provides a robust corpus study. Based on obtained results, fake reviews present more ELABORATION, JOINT and BACKGROUND coherence relations, while the true reviews have more EVALUATION, CONTRAST, EXPLANATION relations. Moreover, the coherence relation of COMPARISON was found only in the real reviews. In addition, the author qualitatively evaluated the content and location of most Nuclear Discourse Units (NDU) (Stede, 2008), which, in accordance with the author, were predictors of deception. Results are shown in Table 4.

Table 4: Comparative analysis of nuclear discourse units (NDUs).

Description	Fake	Real
NDU in first sentence of review	17	9
NDU mentions Title	18	3
NDU mentions Author	8	5
NDU describes content/plot	8	4
NDU contains appraisal/evaluation	18	22

As shown in Table 4, the NDUs were mainly located in the opening sentence, mentioned the book title, and often provided the author name with a brief plot/content description. Therefore, the author concluded that RST analysis provides rich qualitative data for the generation of a set of regulatory heuristics that might include consumer warnings such as: (i) fake reviews are more likely to mention book titles, and authors, as well as give details of a book's content; (ii) fake 5-star reviews tend to be all positive, whereas genuine 5-star reviews usually contain caveats.

5. Discourse-Aware Deception Detection: Main Challenges

Although the RST has been applied to a wide variety of successful applications, we should not simply see it without any criticism. For instance, there are several vague statements and definitions described in the RST original proposal. Indeed, it has been criticized by various authors mainly concerning the aspects related to the absence of a minimal text unit's granularity. Furthermore, rhetorical relations are also highly ambiguous. Since the author have suggested that a level of ambiguity is completely natural between the relations, there are not any instructions or enough scientific and methodology elements to address the relations ambiguity. According to Schauer and Hahn (2000), the number and nature of the rhetorical relations are faintly defined. For example, could any researcher propose and use a set of coherence relations that suits her purposes, and would be them really rhetorical relations? In spite of the disapproval from various authors, a couple of authors proposed to address these "open questions". For example, Maier and Hovy (1993) suggested a taxonomy of three different levels: *ideational*, *interpersonal*, and *textual* in order to group rhetorical relations. In the same setting, Stede et al. (2017), Carlson and Marcu (2001), Vargas et al. (2021) have proposed new rhetorical relations and updated the RST framework. Lastly, another relevant challenge consists of lack of RST-annotated corpus for the deception detection tasks, and low-performance of RST parsers.

6. Conclusions

This paper provides the first comprehensive review of discourse structure and the rhetorical structure theory framework for fake news and fake reviews detection. We also discuss the main challenges and opportunities in this area. Rhetorical structure approach is usually framed as a supervised learning problem, which provides a text representation using coherence relations and hierarchical nuclearity information in order to build automated classifiers. In addition, discourse features are applied with other stylistic-based linguistic features. The initial proposals applied classical machine learning (e.g., SVM and Logistic Regression), while recent proposals have used neural networks (e.g., LSTM and CNN, BERT and RuBERT embeddings). Finally, we conclude that regardless of the important weak points of RST, the relevance and potential of this framework for a wide variety of NLP areas are unquestionable. The RST provides relevant information about subjective aspects of a language, including semantic information and discourse structure, which brings cognitive and contextual insights and may be explored in automatic classification models.

Acknowledgments

The authors are grateful to SINCH, CAPES, CNPq, FAPEMIG, and FAPESP for supporting this project.

7. Bibliographical References

Atanasova, P., Nakov, P., Màrquez, L., Barrón-Cedeño, A., Karadzhov, G., Mihaylova, T., Mohtarami, M., and Glass, J. (2019). Automatic fact-checking using context and discourse information. *J. Data and Information Quality*, 11(3).

Breiman, L. (2001). Random forests. *Machine Learning*, 45:5–32.

Burtsev, M., Seliverstov, A., Airapetyan, R., Arkhipov, M., Baymurzina, D., Bushkov, N., Gureenkova, O., Khakhulin, T., Kuratov, Y., Kuznetsov, D., Litinsky, A., Logacheva, V., Lymar, A., Malykh, V., Petrov, M., Polulyakh, V., Pugachev, L., Sorokin, A., Vikhreva, M., and Zaynutdinov, M. (2018). DeepPavlov: Open-source library for dialogue systems. In *Proceedings of ACL 2018, System Demonstrations*, pages 122–127, Melbourne, Australia.

Carlson, L. and Marcu, D. (2001). Discourse tagging manual. *Tech. rep. ISI-TR-545*, pages 01–87.

Chen, C., Wu, K., Srinivasan, V., and Zhang, X. (2013). Battling the internet water army: Detection of hidden paid posters. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, page 116–120, New York, USA.

DePaulo, B., Lindsay, J. J., Malone, B., Muhlenbruck, L., Charlton, K., and Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, 129:74–118.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.*, pages 4171–4186, Minneapolis, Minnesota.

Fornaciari, T. and Poesio, M. (2014). Identifying fake Amazon reviews as learning from crowds. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 279–287, Gothenburg, Sweden.

Galasiňki, D. (2000). *The language of deception: a discourse analytical study*. SAGE Publications.

Gencheva, P., Nakov, P., Màrquez, L., Barrón-Cedeño, A., and Koychev, I. (2017). A context-aware approach for detecting worth-checking claims in political debates. In *Proceedings of the International Conference Recent Advances in Natural Language Processing*, pages 267–276, Varna, Bulgaria.

Glorot, X., Bordes, A., and Bengio, Y. (2011). Deep sparse rectifier neural networks. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 315–323, Florida, USA.

Hancock, J. T. and Guillory, J. (2015). Deception with technology. In S. Shyam Sundar, editor, *The Handbook of the Psychology of Communication Technology*, volume 16, pages 270–289. John Wiley Sons, Inc.

Hassan, N., Li, C., and Tremayne, M. (2015). Detecting check-worthy factual claims in presidential debates. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, page 1835–1838, New York, USA.

Ho, S. M. and Hancock, J. T. (2019). Context in a bottle: Language-action cues in spontaneous computer-mediated deception. *Computers in Human Behavior*, 91:33–41.

Joty, S., Carenini, G., and Ng, R. T. (2015). CODRA: A novel discriminative framework for rhetorical analysis. *Computational Linguistics*, 41(3):385–435.

Karimi, H. and Tang, J. (2019). Learning hierarchical discourse-level structure for fake news detection. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3432–3442, Minnesota, USA.

Kumar, K. K. and Geethakumari, G. (2014). Detecting misinformation in online social networks using cognitive psychology. *Human-centric Computing and Information Sciences*, 4(14).

Kuzmin, G., Larionov, D., Pisarevskaya, D., and Smirnov, I. (2020). Fake news detection for the Russian language. In *Proceedings of the 3rd International Workshop on Rumours and Deception in Social Media*, pages 45–57, Held Online.

Larcker, D. F. and Zakolyukina, A. A. (2012). Detecting deceptive discussions in conference calls. *Journal of Accounting Research*, 50(2):495–540.

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., and Zittrain, J. L. (2018). The science of fake news. *Science*, 359(6380):1094–1096.

Li, J., Ott, M., Cardie, C., and Hovy, E. (2014). Towards a general rule for identifying deceptive opinion spam. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1566–1576, Maryland, USA.

Maier, E. and Hovy, E. (1993). Organising discourse structure relations using metafunctions. In Helmut Horacek et al., editors, *New Concepts in Natural Language Generation*, pages 69–86. Pinter, London.

Mann, W. C. and Thompson, S. A. (1987). *Rhetorical structure theory: a theory of text organization*. University of Southern California, Information Sciences Institute, California, USA.

Meibauer, J. and Dynel, M. (2016). Empirical approaches to lying and deception. *International Review of Pragmatics*, 8(3).

Meibauer, J. (2018). The linguistics of lying. *Annual Review of Linguistics*, 4(1):357–375.

Mihaylova, T., Nakov, P., Màrquez, L., Barrón-Cedeño, A., Mohtarami, M., Karadzhov, G., and Glass, J. (2018). Fact checking in community forums. In *AAAI Conference on Artificial Intelligence*, pages 5309–5316.

Mubarak, H., Abdelali, A., Hassan, S., and Darwish, K. (2020). Spam detection on arabic twitter. In *International Conference on Social Informatics*, pages 237–251. Springer.

Nahari, G., Ashkenazi, T., Fisher, R., Granhag, P., Hershkowitz, I., Masip, J., Meijer, E., Nisin, Z., Sarid, N., Taylor, P., Verschuere, B., and Vrij, A. (2019). 'language of lies': Urgent issues and prospects in verbal lie detection research. *Legal and Criminological Psychology*, 24:1–23, 01.

Nakov, P., Màrquez, L., Moschitti, A., Magdy, W., Mubarak, H., Freihat, A. A., Glass, J., and Randeree, B. (2016). SemEval-2016 task 3: Community question answering. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 525–545, California, USA.

Newman, M. L., Pennebaker, J. W., Berry, D. S., and Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, 5(29):665–675.

Ott, M., Choi, Y., Cardie, C., and Hancock, J. T. (2011). Finding deceptive opinion spam by any stretch of the imagination. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 309–319, Portland, Oregon, USA.

Pennebaker, J., Boyd, R., Jordan, K., and Blackburn, K. (2015). The development and psychometric properties of liwc2015, 09.

Pérez-Rosas, V., Davenport, Q., Dai, A. M., Abouelenien, M., and Mihalcea, R. (2017). Identity deception detection. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing*, pages 885–894, Taipei, Taiwan.

Pisarevskaya, D. (2017). Rhetorical structure theory as a feature for deception detection in news reports in the russian language. In *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialogue 2017"*. ACM.

Popoola, O. (2017). Using Rhetorical Structure Theory for detection of fake online reviews. In *Proceedings of the 6th Workshop on Recent Advances in RST and Related Formalisms*, pages 58–63, Santiago de Compostela, Spain, September. Association for Computational Linguistics.

Ren, Y. and Zhang, Y. (2016). Deceptive opinion spam detection using neural network. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 140–150, Osaka, Japan, December. The COLING 2016 Organizing Committee.

Rubin, V. L. and Lukoianova, T. (2015). Truth and deception at the rhetorical structure level. *Journal of the association for information science and technology*, 66(5):905–917.

Rubin, V. L., Conroy, N. J., and Chen, Y. (2015). Towards news verification: Deception detection methods for news discourse. In *Proceedings of the The Rapid Screening Technologies, Deception Detection and Credibility Assessment Symposium, 48th Hawaii International Conference on System Sciences (HICSS48)*, page 01–11, Hawaii, United States.

Schauer, H. and Hahn, U. (2000). Phrases as carriers of coherence relations. In *Proceedings of 22nd Annual Conference of the Cognitive Science Society*, pages 429–434, Philadelphia, Pennsylvania.

Scholkopf, B. and Smola, A. J. (2001). *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, Cambridge.

Stede, M., Taboada, M., and Das, D. (2017). Annotation guidelines for rhetorical structure. pages 1–31. Linguistics Department at The University of Potsdam.

Stede, M. (2008). Rst revisited: Disentangling nuclearity. *"Subordination" versus "Coordination" in Sentence and Text*, (1):33–59.

Thompson, S. A. and Mann, W. C. (1988). Rhetorical structure theory: A fremework for the analysis of texts. *IPRA Papers in Pragmatics*, 1:79–105.

Vargas, F., Benevenuto, F., and Pardo, T. (2021). Toward discourse-aware models for multilingual fake news detection. In *Proceedings of the Student Research Workshop Associated with 13th Recent Advances in Natural Language Processing*, pages 210–218, Held Online.

Vosoughi, S., Roy, D., and Aral, S. (2018). The spread of true and false news online. *Science*, 359:1146–1151, 03.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., Drame, M., Lhoest, Q., and Rush, A. (2020). Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Held Online.

Zaynutdinova, A., Pisarevskaya, D., Zubov, M., and Makarov, I. (2019). Deception detection in online media. In *Proceedings of the 15th International Workshop on Experimental Economics and Machine Learning*, page 121–127, Perm, Russia.

Zhou, L., Twitchell, D. P., Qin, T., Burgoon, J. K., and Nunamaker, J. F. (2003). An exploratory study into deception detection in text-based computer-mediated communication. In *Proceedings of the 36th Annual Hawaii International Conference on System Sciences*, page 44.2, Hawaii, USA.