

Detecting deception through RST: A case study of the Casey Anthony trial

Kelli Lynn Finney
Simon Fraser University
kfinney@sfu.ca

Many researchers have used linguistic analyses to determine if features, such as syntactic patterns or word choice, vary based on the truth or untruth of an utterance. For example, Newman et al. (2003) examined lying in written communication, finding that deceptive utterances used more total words but fewer personal pronouns. However, relatively few studies have focused on speech or writing style, which can be used to aid in authorship attribution and plagiarism identification (Cristani et al., 2012), and would thus seem to prove valuable for detecting deception.

Recently, efforts have been made to remedy this by extending the application of linguistic feature analysis. For example, Rubin and Lukoianova (2014) applied Mann and Thompson's (1987) Rhetorical Structure Theory (RST) to elicited written narratives that participants self-identified as either truthful or deceitful. Their findings suggest that RST relations, illustrative of functional relationships between 'spans' of text, vary based on the truthfulness of the narratives. However, this study, like previous studies, relies on researcher-prompted untruths rather than naturally occurring ones. As such, participants have little motivation to make the deception believable, unlike in real-world situations.

The present study thus combines linguistic analysis with an examination of naturally occurring deception in the high-stakes setting of the *State of Florida versus Casey Marie Anthony*, in order to determine if findings like those of Rubin and Lukoianova (2014) are generalizable to deceptive statements in real-world settings. From publically available legal case documents, a corpus of 724 words (65 text segments) was selected and RST relations were coded. While some of Rubin and Lukoianova's (2014) findings were minimally supported, no strong correlation between relations and the truth value of an utterance were found, suggesting the need for additional research in this area.

Keywords: Rhetorical Structure Theory; RST; deception; deceptive statements

1 Introduction: Detecting deception by linguistic means

Many researchers have used linguistic analysis to determine if features, such as syntactic patterns and word choice, vary based on the truth or untruth of an

utterance. For example, Newman et al. (2003) examined lying in written communication, finding that deceptive utterances used a greater number of words as well as fewer personal pronouns. Hancock et al. (2007) conducted a comparable study of synchronous computer mediated communication with similar findings. It is notable, however, that relatively few studies have focused specifically on speech or writing style, which can be used to aid in authorship attribution and plagiarism identification (Cristani et al., 2012), and would thus seem to prove valuable for detecting deception.

More recently, some have made efforts to remedy this, thereby extending the application of linguistic feature analysis. For example, Rubin and Lukoianova (2014) applied Mann and Thompson's (1987) Rhetorical Structure Theory (RST) to elicited written narratives that participants described as being either truthful or deceitful. After examining the groups of 18 deceptive stories and 18 truthful stories collected from Amazon's Mechanical Turk website, they found that the RST relations, illustrative of functional relationships between 'spans' of text, varied based on the truthfulness of the narratives. While these studies suggest that a relationship does exist between the truthfulness of utterances and linguistic variables, they all rely upon researcher-prompted untruths rather than naturally occurring ones. As such, study participants have little motivation to make the deception believable, unlike in a real-world situation.

Other fields, such as psychology, note a similar lack of non-laboratory studies. Vrij and Mann (2001) comment that their study, examining the deceptive and truthful statements of a convicted murderer, was, at the time, the only known study of its type in a "high-stakes realistic setting" (p. 187). This study, while in such a setting, did not approach deception detection from a linguistic framework, it focused instead on the potential correlation of specific behaviors with deception.

The present study thus extends previous work by combining linguistic analysis with the examination of naturally occurring deception in the high-stakes setting of a courtroom trial. It is hypothesized that deceptive statements will evidence different RST relations than those found in truthful statements. Prior to presenting the analysis and accompanying results, however, some background related to the subtopics of study will be helpful.

This background comprises two main categories: an overview of RST and its applications and a survey of the use of linguistic research in deception detection. Following this, several case studies employing linguistics for deception detection are introduced. An examination of these works evidences the possibility of employing linguistic feature analysis for detecting deception, and shows the strong potential that RST has in such analyses.

2 Introduction to RST and stylometry

2.1 RST: Purpose and applications

Rhetorical structure theory (RST) is an attempt to make visible the organization of a text by illustrating functional relationships (Mann & Thompson, 1987). This is done by chunking a text, identifying nuclei (main units) and satellites (supplementary units) of each block of text, and labeling the corresponding relationship between nuclei and satellites. Taboada and Mann (2006) summarize the variety of applications of RST, from identification of key parts in an evaluative text (such as a movie review) to essay scoring, to writing instruction.

Not only does RST provide a way of organizing relationships within text, it can also potentially provide information about the types of relationships that characterize a particular genre (Benwell, 1999, as cited in Taboada & Mann, 2004). While RST has been applied to numerous linguistic situations, surprisingly, only one study was found that used it to analyze textual relations as a distinctive stylistic feature (discussed in Section 3.2). Thus, the present study, which does so, serves to remedy this gap and provide an innovative application of RST.

2.2 Written idiolects and stylometric analysis

Some, such as Coulthard (2004) have proposed that a set of linguistic features found in writing functions in the same manner as an idiolect. He speculates that an individual's writing contains unique characteristics that can assist with such issues as authorship identification and plagiarism, serving as a "linguistic fingerprint" of sorts. Following from his proposal, numerous studies have employed 'stylometry,' the statistical analysis of the presence of various linguistic features, in order to determine authorship. One of the most widely-known cases where stylometry was successfully employed in author identification is that of the Unabomber. For nearly two decades, this United States-based terrorist sent bombs to individuals and universities, and threatened airlines. Due to his use of common scrap materials in bomb construction, the bombs proved untraceable. It was not until he sent an essay to the FBI in the mid-1990s to explain his motives and, perhaps, seek recognition, that authorities were able to make progress on the case. The essay he sent was published in various national newspapers; the individual's brother recognized the writing style, and provided writing samples to authorities (FBI, 2008). After comparing these numerous samples to the essay, the Unabomber was eventually identified from the presence of various idiolect features, such as the use of a particular set of phrases.

While stylometry potentially encompasses numerous categories of linguistic features, including lexical (e.g. word frequencies, n-grams, and depth of vocabulary), syntactic (e.g. types of phrases), and character, semantic, and genre-specific features, most studies have focused on lexical features (Stamatatos, 2011). This is likely because, of all possible features, these are the easiest to

collect, count, and describe statistically. Newman et al. (2003), for example, examine such features as the frequency and type of personal pronouns used, the use and type of conjunctions, and the use and type of prepositions in deceptive and truthful statements. The frequencies were compared not only between statements with differing truth value, but also among typed, handwritten, and spoken text formats wherein university students provided deceptive and true statements regarding their views on abortion.

Some, however, have extended the breadth of stylometric analyses by exploring other linguistic feature patterns. For example, in an attempt to assist with genre classification, Picornell (2013) examined cohesion and coherence relations and syntactic structures in addition to lexical features in written witness statements. Also in an attempt to move beyond lexical counts in stylometric analyses, Cristani et al. (2012) examined conversational features such as turn-taking in naturally-occurring instant messaging conversations to assist with authorship identification.

3 Linguistic methods of detecting deception

3.1 Stylometric analysis for deception detection

One field where stylometry has been widely applied is in the field of forensic linguistics. While identifying authorship is still a main focus, the application of stylometry has been extended to the analysis of written witness statements to differentiate between truthful utterances and deceptive ones (Picornell, 2013). In addition, actual tools used by law enforcement often incorporate features of linguistic style analysis, such as the statement validity analysis checklist mentioned by Porter and Yuille (1996), which includes analysis of coherence in statements.

While it thus appears that stylometry might be a useful tool in a forensic setting, others, including some working in law enforcement, are skeptical. For example, Armistead (2012) notes that such analyses have methodological problems, including the presumption that deceptive statements differ from truthful ones in ways that are observable and consistent. In addition, Armistead (2012) comments that, even if it were the case that such a consistent pattern was established, implementing the techniques used in the studies would not be practical in a law enforcement setting. However, he acknowledges that these linguistic studies of deception do have the benefit of being subject to academic scrutiny, in terms of being peer reviewed and empirically tested. Thus, it is probable that a linguistic method of deception detection could be implemented by law enforcement professionals in the future, if such a method could provide consistent results and be easily used by non-linguists.

3.2 Extension of RST to deception detection

With RST serving as a way of identifying the structural relationships between parts of a text, and potentially, in combination with such theories as that of Biber's (2011) continua, extending to an identification of characteristic patterns in a particular register or genre, it would seem that using RST to compare truthful and deceptive utterances would be a natural extension of the theory. Recently, some have attempted to do so, namely Rubin and Lukoianova (2014). The authors collected a sample of truthful and deceptive narratives and then analyzed these using a combination of RST and vector space modeling. Representing the stories as vectors allowed them to arrange stories in similarity-based clusters, determined by identified levels of truth or deception. In turn, this allowed them to statistically analyze the clusters based on the RST relations found.

The results seem to suggest that, as hypothesized for the present study, a different set of RST relations characterizes deceptive stories than characterizes truthful stories. For example, Rubin and Lukoianova (2014) found that summary, preparation, unconditional, and disjunction relations appeared only in the deceptive stories, while enablement, restatement, and evidence were found only in the truthful stories. However, since this study relied on elicited stories that were ranked by their authors as deceptive or truthful, rather than real-world cases of deception or truth, it is unclear whether these results can be generalized. If it is the case that a similar pattern is found in a high-stakes setting using naturally occurring deception and truth, RST analysis could become a useful tool for law enforcement analyses of suspect and witness statements.

4 Introduction to the data

In selecting data for this analysis, it was necessary to find legal documents that were both easily accessible, and, more importantly, that contained examples of both statements determined to be deceptive and statements determined to be truthful. The documents selected are from the case of the *State of Florida versus Casey Marie Anthony*. This case centred on the disappearance of the defendant's two-year-old daughter, Caylee Marie Anthony. Dubbed "the social media trial of the century" by *TIME* magazine (Cloud, 2011), the case received daily media coverage for nearly three years (2008-2011) in both newspapers and television, as well as in fields such as Twitter, which hosted feeds from both the *Orlando Sentinel* as well as Florida's Ninth Judicial Circuit Court, among others (Cloud, 2011).

Initially, it is likely that the public was captivated by disbelief that a mother would refrain from reporting her child's disappearance to anyone for a month, as Anthony did. However, as coverage of the case progressed, attention shifted to the numerous lies that Anthony told when providing statements to the police, all the while pleading for their help in locating her daughter. Such news outlets as ABC even published articles listing these lies, complete with links to audio recordings of the defendant's interviews with police (Hopper, 2011). At the time

of writing, YouTube still hosts videos of the trial proceedings, and both official statements from the Orange County Sheriff's office, as well as interview transcripts, are accessible to the public online.

4.1 Methodology

The texts selected for analysis were obtained online from media.trb, which in turn obtained them from the *Orlando Sentinel*. Both portions of Casey Anthony's written statement as well as her responses to police, as recorded in interview transcripts, were analyzed. The written statement was that provided by Casey Anthony to the Orange County Sheriff's Office (OCSO) on July 16, 2008, and consists of 680 words. The interview transcript is that of the July 23, 2008 interview between Anthony and OCSO Detective Melich. In its entirety, the statement consists of 13,495 words. Of this total, 4,346 words were those of the defendant. Combining Anthony's written statement and her remarks from the interview transcript yielded a corpus of 5,026 words.

From this corpus, portions comprising 31 and 34 segments of text, respectively, were selected. This was done for two reasons. First, since the purpose of the study was to determine if different RST relations characterize true versus deceptive texts, it was thought that it would be beneficial to compare two text portions that contained a similar number of segments in order to potentially obtain a comparable number of relations. Second, especially at the beginning of the interview, Anthony's responses consist of "Uh-huh," coded by the transcriber as an indication of an affirmative response. It was thought that analysis of 'content full' responses would provide more useful results.

The resulting corpus from the interview transcript consisted of 259 words comprising 34 segments, and that from the written statement consisted of 465 words in 31 text segments. In total, the written statement portion and the selected responses from the interview transcript consisted of 724 words and 65 segments. Due to space constraints, the written statement and interview transcripts are not herein included; however, a URL where they can be accessed is provided in the reference section (Guide to Casey Anthony case legal documents, n.d.)

The documents, originally in .pdf form, were manually converted to plain text format and then imported into RST tool for annotation (O'Donnell, 2003). The written statement was analyzed separately from the interview remarks. In RST tool, each text was segmented following sentence boundaries (the original statement did not have paragraph boundaries; these were not added). After segmenting each text, RST structures were identified and RST relations indicated, using the provided set of "classic" relations from RST tool.

Since portions containing deceptive statements and potentially truthful statements were often in relationship with one another, the decision was made to analyze the RST relations in each text as a whole, and then compare the relations in the deceptive parts to those in the truthful portions. This also helped to guard against bias in the analysis. As the goal of the study is to determine whether specific sets of RST relations correlate with the truthfulness or deceptiveness of

statements, being conscious of the truth value during analysis would call the validity of the results into question.

After all texts were marked in RST tool, the portions of each were coded as being deceptive or truthful. Deceptiveness of the texts was determined using the list of lies in Hopper (2011), as well as by the statements made by OCSO Detective Melich in the July 23, 2008 interview, wherein he identifies certain of the defendant's statements as such. Other portions not explicitly identified as lies were coded as true.¹ Following this, a list of relations and their frequencies were compiled for both the deceptive texts and the truthful texts. Finally, the relations identified in the deceptive texts were compared to those found in the truthful texts.

5 RST relations in deceptive and truthful texts

After completing RST analyses of the written text segment and interview transcript excerpts, each chunk of text was identified as either truthful or deceptive, and the relations in each were compared. In the interview excerpts, one chunk, identified as the 'phone call' text, was almost entirely deceptive with the exception of two text spans (16-17). The RST analysis of this text is provided in Figure 1.

¹ Identifying statements based on proven truthfulness, either through police statements in the interview transcripts or court determinations, would have been methodologically stronger (Thank you to a reviewer for noting this). Doing so was the original intent. However, the nature of the material made this a difficult task. Most transcripts featured police working to uncover falsehoods rather than corroborate truths, which follows from the structure of the US justice system: a suspect is innocent until proven guilty. Thus, all statements are held to be true until they are proved false. It would, perhaps, be more appropriate to view the 'truthful' statements as 'non-false' statements.

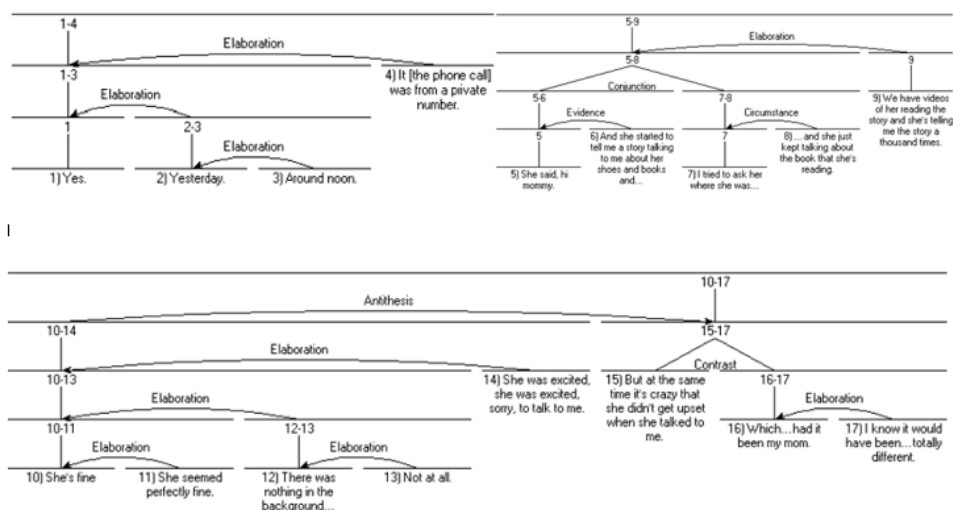


Figure 1: RST analysis of "phone call" text

The other interview text, termed the 'park' text, a portion of which is presented in Figure 2, was judged to be entirely true.

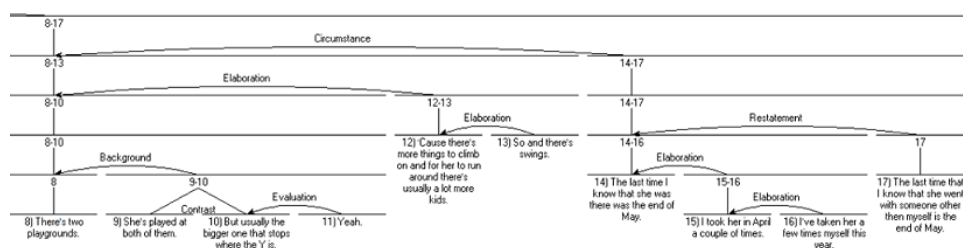


Figure 2: RST analysis of 'park' text, spans 8-17

The written statement portion, however, showed fluidity between true and deceptive portions. For example, the true spans describing Anthony's daughter (spans 2-6 in Figure 3) are provided as background to the deceptive information about Zenaida the nanny, who never existed, in spans 8-15.

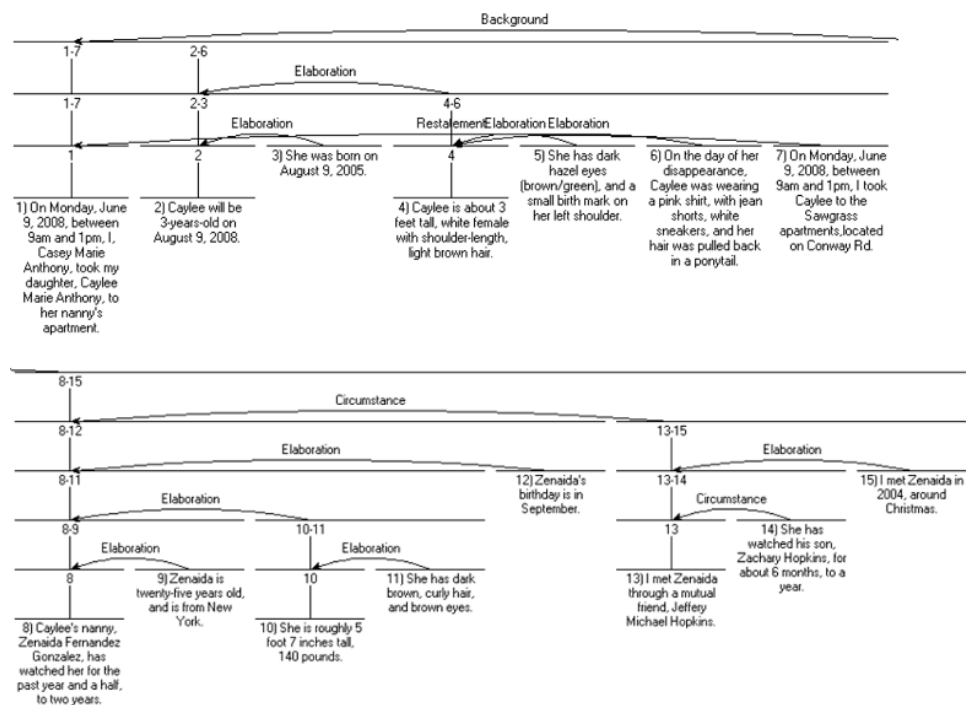


Figure 3: RST analysis of portion of 'statement' text

Table 1 summarizes the RST relations found in both the truthful and deceptive text portions and indicates the frequency with which the various relation types occur.

Relation Name	Deceptive	Truthful
Elaboration	14	13
Circumstance	4	1
Evaluation	0	2
Evidence	1	2
Justify	0	2
Interpretation	0	1
Conjunction	1	0
Restatement	1	0
Antithesis	1	0
Condition	1	0
Solutionhood	1	0
Volitional-Cause	1	0

Table 1: Frequencies of RST relations in analyzed texts

Most noticeable about the RST relations in the two classes of text is the similarity between the number of elaboration relations, with 13 relations of this type in the truthful text portions, and 14 in the deceptive text portions. However, since

elaboration is one of the most common discourse relations, this is not surprising. Also of note are the wider range of relations found in the deceptive texts, with nine different types of relations found as compared to six in the truthful text excerpts.

In addition, several spans of text were judged as showing a relation between a deceptive portion and a truthful one. These include spans 2-6 with spans 8-15 in the written statement (presented in Figure 3) and spans 15-17 in the 'phone call' portion of the interview transcript (seen above in Figure 1).

Rubin and Lukoianova (2014), in their RST analysis of narratives discussed in Section 3.2, found that a limited set of relationships characterized each of the classes of deceptive and truthful texts. Based on their conclusions, it was hypothesized that such relations as disjunction, unconditional, summary, and preparation would characterize deceptive texts, while enablement, restatement, and evidence would describe truthful texts. However, this was not the case. Instead, as seen in Table 1, some relations such as evidence and elaboration were found in both deceptive and truthful texts.

6 Discussion and conclusions

It was hoped that the truth value of statements would show a clearer correlation with the types of RST relations found. However, this was not the case. This could be due to a variety of factors. For example, the defendant, Casey Anthony, was known by friends to be a habitual liar (this was revealed over the course of the trial). As someone becomes more adept at deceiving, it is likely that the relations, which may have once differed from those found in truthful statements, would become similar. It is also possible that Anthony herself came to believe some of the lies she told, which would likely result in a patterning of the relations in such statements with those identified as true.

On a more practical level, it is also likely that the amount of text analyzed, as well as the type of text analyzed, contributed to the inconclusive nature of the results. In a small sample of text, the number of possible RST relations will be comparatively small. This is especially true of the interview transcript portions. Often, the spans of Anthony's answers were in relation with the questions posed by the interviewer rather than with each other. By separating these from their context, it was more difficult to see cohesive links between the spans. The length of her responses, often merely an "Uh-huh" or a "Yes" did not provide enough content to determine a relationship.

Likewise, the topic of the passage was perhaps not ideal for the purpose of an RST analysis. In the written statement, Anthony focuses mostly on a description of her daughter, the nanny, and the circumstances surrounding her daughter's disappearance. The interview statements also showed a similar elaborative structure. Possibly, selecting texts from the genre of written court statements or trial transcripts requires a more thorough perusal of content prior to analysis in order to obtain data that does not focus solely on elaboration. It is also possible that the interview transcripts and the written statement are better

classified as different sub-genres, and would thus not be appropriate for a comparative analysis such as the present one.

In addition to the excessive number of elaboration relations found, another difficulty encountered when completing the RST analysis was the definitions of RST relations themselves. That is, the wording of the definitions conveys the assumption that the provided text is not flouting Grice's maxim of quality. For example, it seems that the relation of Circumstance, by its very definition, implies that a statement must be true, in that it states that "S is not unrealized" (Mann & Thompson, 1987). If a satellite describes a situation that did not occur because it is in itself deceptive, it seems that a circumstance relation should not be possible.

Despite the difficulties encountered in the current analysis, and the inconclusive results, it is still thought that RST could prove to be a valuable tool for law enforcement professionals. Additional studies examining RST relations in deceptive and truthful statements that use larger corpora of data, as well as a wider range of data types (i.e. audio and video statements or interview responses), would be beneficial in making such a determination. It also seems that RST relations might be better used in conjunction with other facets of linguistic feature analysis such as syntactic patterns and word choice, rather than being used as the sole means of determining the truth value of an individual's statements. Doing so would likely provide an analysis having greater depth and breadth, and, more importantly, would increase the credibility of using linguistic analyses in law enforcement settings while guarding against RST analysis becoming either a law enforcement fad or a linguistic polygraph.

References

- Armistead, T. W. (2012). The detection of deception by linguistic means. *Policing: An international journal of police strategies and management*, 35(2), 304–326.
- Biber, D. (2011). Register and discourse analysis. In Gee, J. P. & Handford, M. (Eds.) *The Routledge handbook of discourse analysis* (191-208). New York: Routledge.
- Cloud, J. (2011). How the Casey Anthony murder case became the social-media trial of the century. *TIME*. Retrieved from: content.time.com/nation/article/0,8599,2077969,99.html
- Coulthard, M. (2004). Author identification, idiolect, and linguistic uniqueness. *Applied Linguistics*, 25(4), 431–447.
- Cristani, M., Roffo, G., Segalin, C., Bazzani, L., Vinciarelli, A. & Murino, V. (2012). Conversationally-inspired stylometric features for authorship attribution in instant messaging. *Proceedings of the 20th ACM international conference on multimedia* (1121–1124). Retrieved from: <http://dl.acm.org.proxy.lib.sfu.ca/citation.cfm?id=2396398>.
- Federal Bureau of Investigation (FBI) (2008). FBI 100: The unabomber. Retrieved from: https://www.fbi.gov/news/stories/2008/april/unabomber_042408

- Guide to Casey Anthony case legal documents. n.d. [Data file]. Retrieved from: media.trb.com/media/acrobat/2008-08/41844520.pdf
- Hancock, J. T., Curry, L. E., Goorha, S. & Woodworth, M. (2007). On lying and being lied to: A linguistic analysis of deception in computer-mediated communication. *Discourse Processes*, 45(1), 1–23.
- Hopper, J. (2011). Listen to Casey Anthony's top ten lies. *ABC news*. Retrieved from: http://abcnews.go.com/US/casey_anthony_trial/casey-anthony-top-ten-lies/story?id=13742643
- Mann, W. C. & Thompson, S. A. (1987). Rhetorical structure theory: A theory of text organization. Retrieved from: http://www.sfu.ca/rst/pdfs/Mann_Thompson_1987.pdf
- Newman, M. L., Pennebaker, J.W., Berry, D. S. & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, 29(5), 665–675.
- O'Donnell, M. (2003). RST tool, version 3.4.1. *WagSoft*. <http://www.wagsoft.com/RSTTool/>
- Picornell, I. (2013). Analysing deception in written witness statements. *Linguistic Evidence in Security, Law and Intelligence (LESLI)*, 1(1), 41–50.
- 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 Behaviour*, 20(4), 443–458.
- Rubin, V. L. & Lukoianova, T. (2014). Truth and deception at the rhetorical structure level. *Journal of the Association for Science and Technology*, 66(5), 905–917.
- Stamatatos, E. (2011). Plagiarism detection using stopword *n*-grams. *Journal of the American Society for Information Science and Technology*, 62(12), 2512–2527.
- Taboada, M. & Mann, W. C. (2006). Applications of rhetorical structure theory. *Discourse Studies*, 8(4), 567–588.
- Vrij, A. & Mann, S. (2001). Telling and detecting lies in a high-stakes situation: The case of a convicted murderer. *Applied Cognitive Psychology*, 15(2), 187–203.