The Devil Is in the Details: New Directions in Deception Analysis

Philip M. McCarthy, Nicholas D. Duran*, and Lucille M. Booker

Department of English (Applied Linguistics), University of Memphis, Memphis, TN, USA
*Cognitive and Information Sciences, University of California Merced, Merced, CA, USA
{pmmccrth, lmbooker}@memphis.edu; nduran2@ucmerced.edu

Abstract

In this study, we use the computational textual analysis tool, the Gramulator, to identify and examine the distinctive linguistic features of deceptive and truthful discourse. The theme of the study is abortion rights and the deceptive texts are derived from a Devil's Advocate approach, conducted to suppress personal beliefs and values. Our study takes the form of a contrastive corpus analysis, and produces systematic differences between truthful and deceptive personal accounts. Results suggest that deceivers employ a *distancing strategy* that is often associated with deceptive linguistic behavior. Ultimately, these deceivers struggle to adopt a truth perspective. Perhaps of most importance, our results indicate issues of concern with current deception detection theory and methodology. From a theoretical standpoint, our results question whether deceivers are deceiving at all or whether they are merely poorly expressing a rhetorical position, caused by being forced to speculate on a perceived proto-typical position. From a methodological standpoint, our results cause us to question the validity of deception corpora. Consequently, we propose new rigorous standards so as to better understand the subject matter of the deception field. Finally, we question the prevailing approach of abstract data measurement and call for future assessment to consider contextual lexical features. We conclude by suggesting a prudent approach to future research for fear that our eagerness to analyze and theorize may cause us to misidentify deception. After-all, *successful* deception, which is the kind we seek to detect, is likely to be an elusive and fickle prey.

Introduction

Of all human behaviors that are considered to breach conventions of social and communicative interaction, deception is one of the most pervasive and by far the most elusive. Deception is a violation of what is known to be true for the purpose of providing misleading, but seemingly trustworthy information (Ekman, 1997). To succeed, the deception must be covert and is thus designed to thwart detection. Yet, despite the potential social risks involved, deception is surprisingly common in everyday interactions (DePaulo et al., 1996). This relatively high frequency is presumably the result of deception usually going undetected or being excused as hyperbole. Thus, from the "tall tales" told by fishing buddies, to the homework excuses of students, deceivers generally *get away with it*. But even when the potential risks of lying are high or when the lie strongly deviates from the truth, detection rates of deception are still little better than chance (Vrij et al., 2000). One of the reasons for poor detection is that humans come equipped with a truth-bias, whereby all statements are initially assumed to be true (Gilbert, 1991). Researchers have attempted to overcome this truth-bias by explicitly training people to look for "leakage" cues that are expressed in a deceiver's actions, such as facial movements and body posture (Vrij, 2001); or in their language output, such as in the vividness of spatiotemporal descriptions or number of verbal hedges (Johnson & Raye, 1981). However, even when people are trained in these various techniques, their performance is still too inconsistent for real-world applicability (Bond & DePaulo, 2006; Vrij et al, 2000). Although this poor performance might be attributed to techniques that are theoretically misguided, a more likely account is that the grain-size of leakage cues is outside the normal processing abilities of trained and novice judges. For this reason, many researchers interested in detecting deception have turned to computational techniques to generate rapid, unbiased analyses.

The work described in this paper builds from Hancock et al. (2008) and Duran et al. (2010) and their respective analyses of the linguistic features characterizing deceptive and truthful conversations. We also build from Newman et al. (2003) and their linguistic analysis of deception in arguments on personal beliefs about abortion. In these aforementioned studies, as is the case in other computational work (e.g., Zhou et al., 2004), the focus is on the stylistic organization of language; that is, the abstract linguistic properties that exist at the word, sentence, and discourse level.

Although computational research has certainly provided a great deal of insight into deception, many content-analytic questions remain unaddressed. These questions include *What are the topics that people tend to lie about?* And *How does a particular phrasing reflect the cognitive, social, and motivational biases involved in deception?* Such questions have the potential to expose story elements and thematic content that may only occur because of the processing demands inherent in deceptive communication (e.g., aspects that tap working memory resources monitoring what is real from what is not, see Johnson & Raye 1981).

To address these issues within an applied natural language processing approach, we use the computational tool called the *Gramulator* (McCarthy, Watanabe, & Lamkin, 2012). This tool provides a numeric representation of relevant *qualitative* content: content that consists of short sequences of text (up to four words) that are indicative of one corpus relative to a thematically similar contrastive corpus (e.g., a truth corpus and a lies corpus). Having derived these indicative features, we can go back to the corpus from which the features were found and hypothesize how these features are used in context. By doing so, a richly detailed characterization of deceptive language can be offered. Of course, if the corpus is sufficiently large so as to be divided into sub-corpora (e.g., *training* and *testing* sets), then the validity of the analysis is substantially increased because the features drawn from one set of data are being examined on an unexplored set of data.

In general, content-analytic research is defined as the attempt to extract meaningful representations from large sets of qualitative material, where these representations are derived from objective methods that can be easily reproduced, and that can be interpreted to yield new insights on how people might differ (Holsti, 1969; Smith, 2001). For this paper, we evaluate discourse generated as an expository monologue (spoken and written). We are particularly interested in how people's language becomes marked (i.e., becomes varied from the default truth form) when asked to be deceptive. Our focus of linguistic analysis is short phrases of two to three contiguous word sequences, also called n-grams. We hypothesize that these sequences are important for capturing salient narrative themes (e.g., types of characters, locations, events, feelings; Mandler & Johnson, 1977) or pragmatic elements (e.g., dialogue acts, disfluencies, editing expressions; Clark, 1996) that best characterize deceptive language. However, this approach is *inductive* insofar that the organization of these textual units into psychologically interesting constructs is not known *a priori*, but must be interpreted based on compatibility with existing theories. As described in discussing *the Gramulator*, we go to great lengths to ensure that the extracted textual units are statistically more probable in deceptive texts than in truthful texts (and vice versa), and that these textual units are interpreted within the local sentential context in which they originally occurred (by using a specially adapted concordancer module of the Gramulator).

For deception, homing in on the specific phrases and unique wording can have potentially important consequences in detecting deception. In studies where people are asked to record what they lie about during the course of a day, the bulk of deception tends to dwell on feelings and opinions, as well as personal preferences, achievements, and failures (DePaulo et al., 1996). Thus, knowing what people tend to lie about can signal when a would-be detector should be particularly vigilant. The thematic content of deceptive speech is also relevant for understanding information management strategies that accompany deceptive intent (McCornack, 1992). These are strategies that deceivers use to control the content of a message by obfuscating the truth and thwarting perceptions of guilt. Such control is showcased by Burgoon et al. (1996) who evaluated what was said by deceivers in structured interviews. The researchers concluded that deceivers tend to provide impoverished details, downplay personal involvement, and provide less relevant information. Again, such information is crucial for enhancing the goals of deception detection.

To evaluate language, content-analytic researchers often rely on human raters to code theoretically interesting features, where high agreement between raters is a priority. However, this process can be extremely time-consuming, particularly when there are many texts and multiple features to code. Furthermore, human raters can easily overlook subtle semantic patterns that are embedded in more salient content.

Of course, the limitations of human raters are easily contrasted with the processing speed and pattern extraction abilities of computational approaches. As in the studies of Duran et al. (2010), Hancock et al. (2008), and Newman et al. (2003), natural language processing tools have been used to process hundreds of linguistic features in a matter of seconds. Many of these features are also likely impossible for human raters to identify. For example, Duran et al. (2010) used the *given-new index* available in the Coh-Metrix natural language processing tool to capture a construct of *information novelty* (McCarthy et al., 2012). This index functions by computing the co-occurrence patterns of content words across contiguous sentences in a text. The algorithm is not dependent on any *a priori* notion of what raters might agree to be typical, or even what raters would recognize as being typical. Rather, the algorithm is designed to mindlessly (quite literally) evaluate hundreds of texts in terms of the amount of new information present in each text.

As is generally the case in this and other computational approaches, the data are interpreted without any direct reference to the specific words (or extended text). That is, the analysis is based on a composite measure of the abstract properties of words and relationships between words. When encountering the phrase "this is a chair," computational algorithms similar to *given-new* might track information like: there are four words, there is one verb phrase, there is a noun that is a hypernym of furniture, et cetera. Thus, the analysis, by design, transforms the semantic content into higher-level, abstract properties. Two texts could be about very different topics, but potentially have the same Coh-Metrix values. However, as noted earlier, there are notable advantages in evaluating short, semantic phrases of discourse. For example, although a *given-new* evaluation might show that deceivers tend to be more redundant, this conclusion could be strengthened by also determining what people are more likely to talk about when telling the truth, or avoid talking about when telling a lie. In this way, researchers can begin asking why certain themes are avoided and others are not, and ultimately use this information to improve our understanding of the psychological underpinnings of deception, as well as the development of techniques for enhanced detection.

The Gramulator

Natural language processing tools have been tremendously successful at offering insight into register differences (e.g., Duran, McCarthy, Graesser, & McNamara, 2007; Pennebaker & King, 1999). However, the emphasis in these analyses has been on converting content to abstract representation. Thus, words in a text such as "happy" or "grateful" may count towards a pre-defined measure (e.g., *percentage of positive emotion words*). Given that this kind of output is removed from the actual context in which these words originally appeared, researchers might overlook changes in meaning that are context-dependent. What is needed then is a computational tool that can complement existing techniques by revealing context-embedded features of the text.

For this study, we address the need for context-embedded features by using the computational textual analysis tool, the Gramulator (McCarthy, Watanabe, & Lamkin 2012). The Gramulator is designed to examine contrastive corpora. That is, sets of highly related data that are theoretically distinguishable at a single identifiable point of interest (i.e., truthful vs. deceptive accounts on the same theme.) The Gramulator distinguishes these corpora by identifying key sequences of text (n-grams) that are characteristic of one set of data while being non-characteristic of the compared data. The identified key linguistic sequences are referred to as *differentials*. The Gramulator is ideal for the current study because of its focus on related data sets and its content relevant assessment approach.

The Gramulator's operation can be understood as a multistep process. These steps include a) collecting two candidate corpora for analysis, b) assessing inclusion of n-grams based on noise/signal levels, c) Tallying the n-grams that appear in each corpus and retaining only the n-grams that appear with above average frequency in each corpus (i.e., the *typicals*), and d) comparing the typicals of each corpus and removing those typicals that overlap. In this latter step, the goal is to identify typicals that are not shared (the *differentials*), and therefore are indicative of their respective corpus relative to the corpus against which they have been compared.

The Role of Deception in the Context of Personal Beliefs and Values

The number of possible types of deception is virtually limitless but perhaps the lies that are most difficult to tell are those that require the rejection or suppression of personal beliefs and values (Noonan, 2003). As such, one of the best ways for exposing the language of deception is to take the "Devil's Advocate" approach and apply it to highly sensitive issues. With such an approach, human judges appear to achieve their most reliable and stable deception identification rates, presumably because when arguing against a strongly held belief, it is particularly difficult for people to maintain the same type of language as they might use when telling the truth (Leal et al. 2010).

Building from such research, this study uses essays and transcripts on abortion attitudes (see Newman et al., 2003). Newman and his colleagues' corpus (as we use here) has been subjected to extensive linguistic computational analyses and results claim to have identified a large variety of linguistic characteristics unique to deceptive and truthful language. To collect the texts for the corpus, participants were required to argue for both a *pro-choice* and *pro-life* position on abortion (participants first indicated which position they favored). In total, the corpus consists of 352 texts from 176 unique participants (81 men, 95 women). Of these 176 participants, 50 identified themselves as pro-life (24 men, 26 women), and 126 identified themselves as pro-choice (57 men, 69 women). Texts were an average of 195 words in length.

From this corpus, we focus on the language of the *fake pro-lifers*. This group was selected as the major point of interest because it was the largest group (i.e., 126 pro-choice participants as opposed to 50 pro-lifers), and because this group features the manipulated condition (i.e., they were asked to lie). To assess the validity of the language of the *fake pro-lifers*, we use the fake pro-lifers' LIE-differentials and evaluate whether these n-grams are consistent with the language of the *real pro-lifers*. Thus, we presume that a good liar will use the same language as someone telling the truth. More specifically, we presume that for deception to be successful (i.e., undetected) the language of *fake pro-lifers* will be sufficiently prevalent in the language of *real pro-lifers*. If this is not the case, then we have evidence that *fake pro-lifers* are unconvincing, (and evidence also that the Gramulator is detecting qualitative information that distinguishes deception from truth).

Following standard methodological procedures for the Gramulator, in this study we set "differentials" at *bigram*, set "frequencies" at *weighted averages*; and select "noise level" at *dis legomenon*. That is, differential lexical sequences were set at two adjacent words, with all differentials having above average frequency in terms of raw count normalized by document count, and with all bigrams of fewer than three occurrences (i.e., potential noise) removed from the analysis.

Our analysis begins with a corpus validation procedure, which is conducted to establish confidence in the consistency of the data under examination. The validation procedure is followed by two qualitatively based assessments: (1) Distancing by equivocating, and (2) Distancing by appealing to external agency. These assessments are based on Gramulator identified differentials and are supported quantitatively by the Gramulator's Concordancer module, which utilizes Fisher's Exact Test probabilities.

Corpus Validation

As a pre-validation step, we randomly selected two-thirds of *fake pro-lifers'* deceptive texts to form a LIE *training* set (83 texts) and a LIE *test* set (43 texts). The Gramulator was used to identify LIE-differentials by contrasting the LIE training set data against those participants' truthful (pro-choice) arguments. The procedure resulted in 60 LIE-differentials. Using the Gramulator's Evaluator module, we assessed the distribution of the 60 LIE-differentials in the LIE training set and the LIE test set. The results indicated no significant differences between the data sets, providing some confidence in the consistency of the corpus. Next, as the main validation procedure, we compared the LIE and TRUTH test sets (i.e., *fake pro lifers* and *real pro-lifers*). This time the results indicated a significant differences between the data sets with higher values for the LIE texts: LIE: $M = 6.79$, $SD = 4.27$; TRUTH: $M = 4.19$, $SD = 2.62$; $t(68) = 3.36$, $p < .01$. This result suggests that the test sets contained different qualitative information, even though both test sets were ostensibly arguments for the same pro-life position. Thus, the LIE-differentials (i.e., the indicative language of deception) appear to differentiate lies from truth. Most importantly, the result suggests that lying *for* a position generates different language from telling the truth *for* the position.

Distancing by equivocating

For the deceivers, the struggle to articulate a position is marked by increased differentials like *it's just.* This differential appears in 9.7% of the deceptive texts compared to 2.8% of the true texts, with the overall rate of usage being over three times greater in deceptive texts than in true texts (24 versus 7). Fisher's Exact Test assesses this difference as significant ($p = .014$). The use of *it's just* is riddled with common reformulations and hedges such as *um*, *uh*: themselves commonly occurring within differentials. Similar structures such as *kinda* and *yaknow* are also common place in this environment (see examples 1 to 5 below). Such usage may reflect relatively poor access to the topic matter that *fake pro-lifers* are trying to express, and may also explain why the examples appear more like an appeal to pathos than a presentation of an argument. This conclusion is supported by several other similar discourse markers, each of which appears as differentials: *um I*, *that uh*, *that um*, *and uh*, *um it*.

1. and, um, that is how i see it. it's, it's just like another life taken away. kinda, kinda, um,
2. … and, uh, it's just my own reason. I don't think it is right.
3. … to that and, um, it's just really messed up right now.
4. all the problems … it's just like, yaknow, overpopulated and everything
5. yaknow, going off and all things. so, yaknow, it's just like part of thing in their body, yaknow, just.

The lexical items of equivocation identified above may be used to signal a delay in communication, with such delays arising from problems in planning, retrieving a word or idea from memory, or hesitation due to uncertainty about the appropriateness of what is being said (Clark & Fox Tree, 2002). However, the deception literature appears to converge on the general notion that lying places an increased burden on processing. These demands may be reflected in the use of *it's just*. That is, rather than signaling a delay, *it's just* can be viewed as an editing expression that indicates a correction to, or justification of, the truth-value of some previous statement (Clark, 1996). Given the uncertainty conveyed by *um*, *uh*, *kinda*, *yaknow* and so forth, it is plausible that deceivers feel compelled to use *it's just* as an attempt to clarify what they know to be a tenuous argument. These verbal cues also appear unintentional, and therefore contrast with previous research that has found deceivers strategically minimize such cues to fend off the suspicion of others (Arciuli, Mallard, & Villar, 2010). It is possible that the emotional gravity of the present topic (i.e., whether abortion should or should not be allowed) limits strategic control that would normally be employed to avert suspicion.

Distancing by appealing to external agency

The LIE-differentials of *shouldn't be, government should, it shouldn't, responsible for, need to* all seem to point to some form of agency exerting moral or legal obligation. These combined differentials appear in 31.8% of texts compared to 16.5% for the truth condition. Fisher's Exact Test assesses this difference as significant ($p < .001$). Additionally, 7 of the top 25 LIE-differentials appear to correspond to some form of "legal agency" including *to kill*, *is murder*, *killing a*, *kill a*, *the Bible*, *murder and, be legal*. Given this relationship, the underlying theme that appears to be most characteristic of fake pro-life arguments (see examples 1 to 5 below) is *legal agency should prevent murder*. In contrast, legal agency is strikingly absent in the truth condition with just three examples: *of rape* (ranked 5th), a law (ranked 46th), and *rape or* (ranked 81st). The single specific act of *rape* in the truth condition should not go unanalyzed: This differential appears in just 2.3% of the deceptive texts compared to 13.6% of the true texts, with the overall rate of usage being nearly seven times greater in truthful texts (27 versus 4). Fisher's Exact Test assesses this difference as significant ($p < .001$).

1. even though it's just a zygote and it shouldn't be the government's decision to kill or whatever
2. it is, it is, it is murder. it is the same thing as murder.
3. You should just read the bible and the bible says thou shalt not kill
4. It just, murder is not legal so why should an abortion be legal? um, uh.

5. … should be outlawed, except in cases of rape. the known ways of preventing pregnancy are abstinence,

Turning to the *real pro-lifers,* the most striking lexical item is the pronoun *I*. The word *I* occurs in 9 TRUTH-differentials and yet not once in the lying counterparts. In terms of presence in texts, the pronoun *I* appears in 76.7% of the truth texts compared to 62.5% of the lie texts, with frequencies in lie texts reaching just 58% of the truth counterpart. Fisher's Exact Test assesses the difference in terms of document presence as significant ($p = .005$).

The finding for the frequent use of *I* is well documented in Newman et al. (2003); however, the context (i.e., the differentials and their context as provided by the concordancer) suggests that the use of *I* may have more to do with truth tellers' knowledge and rhetorical confidence than with deceivers choosing to distance themselves from their discourse.

To examine this hypothesis, let us consider the context of the pronoun *I* in truthful accounts. The highest ranked TRUTH differential (*but I*) occurs in 27 truth texts compared to just 6 lie texts ($p < .001$); The third highest ranked TRUTH differential (*I would*) occurs in 23 truth texts compared to 7 lie texts ($p = .004$); and (*I have*) fills out the top 10 TRUTH-differentials occurring in 17 truth texts compared to 2 lie texts ($p = .001$). As we see from the examples below, the pronoun *I* is used as a personal balance to a given counterpoint. Importantly, the point here is not whether the truth tellers have the correct argument (or even good ones), the point is that they appear to be demonstrating their broader knowledge of the theme, and situating their own opinion within that theme. Deceivers, as we saw earlier, could not be more different: struggling to articulate themselves and ultimately relying on external agencies as their support.

1. … how to handle an unplanned pregnancy. <u>but i</u> really believe that the child's right to life …
2. it is against my religion, but i think <u>i would</u> still be against it even if i wasn't catholic.
3. … the baby can't feel any pain during the abortion procedure , <u>i have</u> researched evidence that proves otherwise.

Taken as a whole, the analysis performed here provides evidence for the *distancing strategies* that are often attributed to deceivers (Buller & Burgoon, 1996; Newman et al., 2003). That is, deceivers appear willing to refer to the standards of others rather than their own sense of moral correctness, whereas truth tellers are comfortable placing themselves in the center of the argument. Such a conclusion supports the current theoretical framework; however, our results seem to offer an alternative hypothesis that is less about *personal distancing* and more about *being personally distant*. That is, if asked to lie about a particular theme, then participants are being asked to draw from a knowledge reserve that they don't personally have available. This lack of knowledge may cause cognitive duress, and because of processing limitations, deceivers may simply be latching on to some default stereotypical themes as a form of reprieve. Consequently, it cannot be surprising that deceivers would hedge, stumble, and struggle for articulation. Moreover, as they are being asked to personally remove their beliefs and values from the argument, it is reasonable that the personal pronoun would accompany their departure. In contrast, truth tellers are required to highlight their convictions, and so their demonstration of assumed world knowledge together with their beliefs and values within that frame are likely to be prevalent.

Discussion

In this study, we analyzed distinctive linguistic features of deceptive and truthful discourse. The focus was the suppression of personal beliefs and values within context of the issue of abortion. Our methodological approach was contrastive corpus analysis, which was automated through the employment of the computational textual analysis tool, the Gramulator. This automated analysis produced a series of differentials: lexical extracts that were indicative of one corpus while being antithetical of the compared corpus. Analysing the extracts using the Gramulator modules, we identified unique rhetorical elements across the two corpora that suggested systematic differences between truthful and deceptive accounts. Our results suggest that deceivers struggle to adopt the truth perspective, at least on issues that violate core beliefs and values. In other words, we find that *people aren't that good at lying*. But perhaps a more important issue than *whether* people are good at lying is *how* we go about lying. That is, the question that emerges is *where do we go to lie*? Our findings suggest that deceivers *go away* from themselves in order to lie and *go to* external agencies. However, whether deceivers are really *on the move* of their own volition or whether deceivers have been *made to move* and are therefore simply refugees in the tents of external agencies is an issue that requires further research.

Our findings here highlight the importance of continuing to evaluate deceptive behavior. However, given the variability and flexibility of language and the innumerable reasons to employ deception, the quest to find universal linguistic patterns within targeted domains (e.g., abortion rights) might be an exasperating venture. This said, as more domains are evaluated, *meta-themes* might emerge that bring researchers closer to identifying useful generalizations. One such meta-theme, *distancing strategies*, was explored in this study; but other themes such as *issue ownership* (Petrocik, Benoit, & Hansen, 2003) and *linguistic style matching* (Niederhoffer & Pennebaker, 2002) also need to be developed and explored. Such studies are encouraged, as they may help to identify fundamental cognitive or affective constraints that influence deceptive behavior.

In this study, we broke with traditional abstract assessment of natural language and instead adopted the Gramulator. The Gramulator allows us to focus on key linguistic features at the lexical level (i.e., differentials) and consequently to contextualize the abstract measurements (e.g., cohesion, positive emotion) that are currently the dominant approach. To be sure, a lexical level approach brings with it some potential disadvantages: For example, deceivers might be quite good at monitoring and controlling their content-based language, thus making the Gramulator approach easier to dupe. Thus, analyses that focus on more stylistic and abstract features might have the upper hand in detecting deception because they require extensive mathematical calculation and deceivers are therefore unlikely to be able to successfully monitor their language (Newman et al., 2003). Further research is needed to determine if abstract features are indeed resistant to duping, and whether this resistance is superior to the linguistic features approach used here.

It is tempting at this point to sit back and disperse what appears to be a very convincing set of results. However, while the results presented here are encouraging, they need to be interpreted with some caution. Specifically, the validation analysis was conducted with noise levels (the rejection level for frequency) at the dis legomenon level (i.e., only consider examples with a minimum three instances). Such a selection is a reasonable point of departure for any study but confirmatory analysis at the hapax legomenon level (i.e., only consider examples with a minimum two instances) did not significantly distinguish the truth and lie corpora at the indexical level. That said, the qualitative analysis using the derived differentials remained strong, and so we are able to demonstrate confidence that our analysis is still of interest to the field of deception study.

Our concern for corpus validity may ultimately show only that a relatively small corpus (such as the one used here) requires a stricter level of noise reduction (as we used here). However, such a conclusion may seem too convenient, or at least lacking in the rigor that our field deserves. After all, if any field of interest deserves the highest degree of honesty it is surely the study of deception: for whom would wish to be *hoisted by their own petard*? Thus, a more appropriate response to our concern would seem to be one of examining the standards and requirements of corpus collection. Thus, we argue that what the field of deception detection may most urgently need is to abandon the convenience of relatively small data sets and instead collect purpose built corpora that can be rigorously validated prior to analysis.

Proposed Approach to Future Deception Studies

To address the issues raised in this study, we take the opportunity here to propose *where we need to go* in deception research. Thus, the remainder of this paper describes a proposed theme of study, a proposed corpus design that avoids the potential pitfalls identified here; and perhaps most critically, we propose a validation procedure that establishes confidence in the findings that any analysis may produce.

The Role of Deception in Accounts of Political Views

For our proposed study, we will collect deceptive and truthful views on key political issues. Our focus will be on differentiating the language of divergent political groups. The participants of our study will be randomly selected, self-identified liberals and conservatives. They will be asked to write both their truthful and deceptive views concerning whether or not to have *stricter gun control laws*, and whether or not to have *stricter immigration laws* in the United States. Such an approach will again highlight the suppression of personal beliefs and values, which appears to be critical to the understanding of deceptive linguistic behavior.

A Contrastive Corpus Approach

Our proposed study will necessitate the generation of a contrastive corpus (one sub-corpus of truthful texts and one sub-corpus of deceptive texts). The contrastive corpus will be analyzed following the procedures identified in this study so as to identify the indicative language features of each political group. The participants for our study will be recruited through *Amazon Mechanical Turk* (MTurk). This web-based vehicle uses a crowdsourcing design to connect people willing to be experimental participants with researchers who need them (Strain & Booker, 2012). In practical terms, MTurk is an online experimental laboratory.

Our proposed survey, entitled *Truth and Lies in Politics*, will be used to collect the truthful and deceptive texts from MTurk participants. Following corpus text size guidelines provided in McCarthy, Watanabe, and Lamkin (2012) and McCarthy and Jarvis (2007), approximately 100 words for each submission will be required, with a goal of collecting a total corpus size of 700 texts (350 TRUTH, 350 DECEPTION). This design allows for four contrastive constructs: (a) TRUTH and LIE; (b) FOR_IT and AGAINST_IT (with "IT" always referring to the issue under consideration; for example, FOR 'stricter gun control laws in the United States'); (c) FOR_IT_TRUTH and AGAINST_IT_TRUTH; and (d) FOR_IT_LIE and AGAINST_IT_LIE. These contrastive constructs will be based on the responses provided by participants. Having collected the data, the Gramulator's *Sorter* module will then be used to randomly divide each set of data into a training set and a test set: For example, TRUTH (stricter gun control laws) training set, and TRUTH (stricter gun control laws) test set. Following this procedure, the Gramulator will process the training set data to create the differentials that comprise the TRUTH and LIE indices.

Internal Validation Process

The *internal validation process* involves a multi-level approach that tests the *homogeneity* and *markedness* of the data and the derived indices. The process will be conducted using the Gramulator's *Evaluator* module and its various statistical components. The goal of the internal validation process is to expose the data to a series of interrelated assessments that collectively establish confidence in the overall internal validity of the data. We describe below the six individual components of the validation process and our approach to their collective interpretation.

Homogeneity of Data

The test of the *homogeneity of the data* requires us to assess whether the data across each data set (i.e., TRUTH and LIES) is consistent, as opposed to pockets of varying signal or simple noise. The assumption for this test is that the indices are valid, and with these valid indices we are evaluating the consistency of the data sets. Specifically then, the assessment is used to evaluate whether the test set data (which is independent of the training set data, from which the training indices were derived) yields predicted higher values for their corresponding training index. For example, using the LIE training index, the LIE test set is predicted to have a greater presence of LIE-differentials than the TRUTH test set.

Homogeneity of the Indices

The test of the *homogeneity of the indices* requires us to assess whether the indices for each data set (i.e., TRUTH and LIES) are consistent as opposed to being a noisy array of varying n-grams. The assumption for this test is that the data sets are valid, and with these valid data sets we are evaluating the consistency of the indices. Thus, the assessment is used to evaluate whether the training indices (created from the training set data) are more predictive of their corresponding test set data than their contrastive indices. For example, the LIE training index is predicted to be more predictive than the TRUTH training index of the LIE test set data.

Markedness Test

The test of the *markedness* of the indices (i.e., variation from the default truth form) assesses whether there is more marked language in marked data than there is default language in default data? The test is designed to demonstrate that the more distinctive, or *marked index* (LIES) is a better predictor of the *marked data set*, than the default, or *unmarked index* (TRUTH) is a predictor of the *unmarked default data set*. Thus, the LIE index is predicted to measure LIE better than the TRUTH index is predicted to measure TRUTH.

Default Test

The *default* test is designed to assess the distribution of marked language. That is, the test assesses whether there is more default language in marked data than there is marked language in default data. Specifically, the default test is used to demonstrate that the marked data (LIES) have more of the language of the unmarked index (TRUTH) than the unmarked data (TRUTH) have the language of the marked index. For example, the LIE test set is predicted to have more of the language of the TRUTH index, than the TRUTH test set is predicted to have of the language of the LIE index, because there is presumably more truthful language in lying texts, than there is lying language in truthful texts. In other words, there is more ‘truth’ in lies, than there are ‘lies’ in ‘truth’.

Interpretation of the Results

The internal validation process involves conducting a total of six tests. On the face of it, each assessment may appear to be a simple t-test, with corresponding p-values for these tests assessing significance; however, a collection of six t-tests does not offer a useful gauge of overall significance of the assessment because, obviously, a series of six tests is less likely to yield consistent predicted results than is just one. Note that the IVP is effectively the opposite of conducting six tests on one set of data, which would call for a Bonferroni adjustment. Instead then, we need to assess the probability of six tests yielding a given result. In this case, we can say that the mean for each group is either a) in the predicted direction or b) not in the predicted direction. As such, the probability of either result is 0.5, or 50/50. Using the binomial test, we can say that the probability of all six tests resulting in means in the predicted direction is .016. And for five of the six, the probability is .094. Setting alpha at .05, we can say that if all six tests are in the predicted direction, we will deem the result “significant.” And if five of the six results are in the predicted direction, we will refer to the result as “approaching significance.” Fewer than five predicted results can be attributed to chance, and will be deemed not significant.

Conclusion

From a theoretical standpoint, the results of this study highlight four issues of consideration for future deception research 1) Deceivers may choose to be employing distancing strategies to avoid conflicts with personal beliefs and values; 2) Deceivers may be forced into employing distancing strategies because of an insufficient amount of personal knowledge; 3) Deceivers lack of relevant knowledge may result in a poor rhetorical performance; 4) Experiments that force participants to deceive outside of their comfort zone may be flawed because deception is typically what someone would choose to do for personal gain, not have thrust upon them for potential personal embarrassment.

From an analysis standpoint, the results of this study highlight two issues of consideration for future deception research 1) Corpora need to be sufficiently large for rigorous validation procedures to be conducted. 2) Analysis beyond the level of abstract measure is required so that contextual lexical features can be fully examined. With regard to the first point, we argue here that the next step in deception detection needs to be a prudent one. Specifically, we advocate establishing corpus and validation guidelines for future research. Without appropriate data, our finest computational approaches and our most cherished theoretical frameworks can have little value. Perhaps if we consider that deception studies to this point have been more of a quest for stronger theoretical paradigms and ever more fanciful computational and statistical gymnastics then it may be a good time for our field to move (at least for a while) away from issues of data analysis and towards issues of data collection. With regard to the second point, the approach of the computational textual analysis tool, the Gramulator, appears to be a highly valuable, and researchers are encouraged to use such contextual tools to supplement future deception analysis.

Although cautious of our claims and tentative of our proposed direction foreword, the current study offers important analysis for researchers in the field of deception detection. For these researchers, we feel it is important to note as we end what we drew attention to as we began: That deception is only successful if it thwarts detection. As such, cautious analysis and cautious claims would seem appropriate because although we may be detecting much, the much we are detecting may not end up being the deception we sought.

References

Arciuli, J., Mallard, D., and Villar, G. 2010. “Um, I can tell you’re lying”: Linguistic markers of deception vs. truth-telling in speech. *Applied Psycholinguistics*, 31: 397-411.

Bond, C.F., Jr., and DePaulo, B.M. 2006. Accuracy of deception judgments. *Personality and Social Psychology Review*, 10: 214-134.

Buller, D.B., and Burgoon, J.K. 1996. Interpersonal deception theory. *Communication Theory*, 6: 203-242.

Burgoon, J.K., Buller, D.B., Guerrero, L.K., Afifi, W., and Feldman, C. 1996. Interpersonal deception: XII. Information management dimensions underlying deceptive and truthful messages. *Communication Monographs*, 63: 50-69.

Clark, H.H. 1996. *Using Language. Cambridge*: Cambridge University Press.

Clark, H.H. and Fox Tree, J.E. 2002. Using *uh* and *um* in spontaneous speech. *Cognition*, 84, 73-111.

DePaulo, B.M., Kashy, D.A., Kirkendol, S.E., Wyer, M.M., and Epstein, J.A. 1996. Lying in everyday life. *Journal of Personality and Social Psychology*, 70: 979-995.

Duran, N.D., Hall, C., McCarthy, P.M., and McNamara, D.S. 2010. The linguistic correlates of conversational deception: Comparing natural language processing technologies. *Applied Psycholinguistics*, 31: 439-462.

Duran, N.D., McCarthy, P.M., Graesser, A.C., and McNamara, D.S. 2007. Using temporal cohesion to predict temporal coherence in narrative and expository texts. *Behavior Research Methods, 29,* 212-223.

Ekman, P. 1997. Lying and deception. In N. L. Stein, P.A. Ornstein, B. Tversky, and C. Brainerd (Eds.), *Memory for Everyday and Emotional Events* (pp. 333-348). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Gilbert, D.T. 1991. How mental systems believe. *American Psychologist*, 46: 107-119.

Hancock, J.T., Curry, L.E., Goorha, S., and Woodworth, M. T. 2008. On lying and being lied to: A linguistic analysis of deception. *Discourse Processes*, 45: 1-23.

Holsti, O.R. 1969. *Content analysis for the social sciences and humanities. Reading*, MA: Addison-Wesley.

Johnson, M.K. and Raye, C.L. 1981. Reality monitoring. *Psychological Review*, 88: 67-85.

Leal, S., Vrij, A., Mann, S., and Fisher, R.P. 2010. Detecting true and false opinions: The Devil’s Advocate approach as a lie detection aid. *Acta Psychological, 134*: 323-329.

Mandler, J. and Johnson, N. 1977. Remembrance of things parsed: Story structure and recall. *Cognitive Psychology, 9*: 111-151.

McCarthy, P.M., Dufty, D.F., Hempelmann, C.F., Cai, Z., McNamara, D.S., and Graesser, A.C. 2012. Newness and Givenness of Information: Automated Identification in Written Discourse. In C. Boonthum-Denecke, P. M. McCarthy, & T. A. Lamkin (Eds.), Cross-Disciplinary Advances in Applied Natural Language Processing: Issues and Approaches (pp. 457-478). Hershey, PA: IGI Global.

McCarthy, P.M., and Jarvis, S. 2007. A theoretical and empirical evaluation of *vocd*. *Language Testing*. 24, 459-488.

McCarthy, P.M., Watanabe, S. and Lamkin, T.A. 2012. The Gramulator: A tool to indentify differential linguistic features of correlative text types. In P.M. McCarthy and C. Boonthum-Denecke (Eds.), *Applied Natural Language Processing: Identification, Investigation and Resolution* (pp. 312-333). Hershey, PA: IGI Global.

McCornack, S.A. 1992. Information manipulation theory. *Communication Monographs*, 59, 1-16.

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,* 29: 665-667.

Niederhoffer, K.G., and Pennebaker, J.W. 2002. Linguistic style matching in social interaction. *Journal of Language and Social Psychology*, 21, 337-360.

Noonan, H.W. 2003. *Personal Identity*. New York, NY: Routledge.

Pennebaker, J.W. and King, L.A. 1999. Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology,* 77: 1296-1312.

Petrocik, J.R., Benoit, W.L., and Hansen, G.T. 2003. Issue ownership and presidential campaigning, 1952 – 2000. *Political Science Quarterly*, 118 (4), 599-626.

Smith, C.P. 2001. Content analyis and narrative analysis. In H. T. Reis and C. M. Judd (Eds.), *Handbook of Research Methods in Social and Personality Psychology.* Cambridge, UK: Cambridge University Press.

Strain, A.C. and Booker, L.M. 2012. Amazon Mechanical Turk: A web-based tool for facilitating experimental research in ANLP. In C. Boonthum-Denecke, P. M. McCarthy, and T. A. Lamkin (Eds.), *Cross-Disciplinary Advances in Applied Natural Language Processing: Issues and Approaches*. Hershey, PA: IGI Global.

Vrij, A. 2001. *Detecting lies and deceit: The psychology of lying and implications for professional practice.* Chichester: Wiley.

Vrij, A., Edward, K., Roberts, K.P., and Bull, R. 2000. Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal Behavior,* 24: 239-264.

Zhou, L., Burgoon, J.K., Nunamaker, J.F., and Twitchell, D. 2004. Automating linguistic-based cues for detecting deception in text-based asynchronous computer-mediated communication. *Group Decision and Negotiations,* 13: 81-106.