MLS – PSYCHOLOGY RESEARCH (MLSPR)
http://mlsjournals.com/Psychology-Research-Journal
ISSN: 2605-5295

(2025) MLS-Psychology Research, 8(2), pp-pp. doi.org/10.33000/mlspr.v8i2.4003

VALIDATION OF CREDIBILITY DISCOURSE ANALYSIS (CDA) AS CREDIBILITY ANALYSIS PROTOCOL FOR ADULT STATEMENTS

Validación del Protocolo de Análisis del Discurso de Credibilidad (CDA) para la Evaluación de Veracidad en Declaraciones de Adultos.

Anderson Tamborim
Social Intelligence Group, Deception Detection Lab (Brasil)
(contato@andersontamborim.com) (https://orcid.org/0000-0002-5051-4267)

Información del manuscrito:

Recibido/Received:12/12/24
Revisado/Reviewed: 20/05/25
Aceptado/Accepted: 01/07/25

Keywords:
credibility assessment; discourse analysis; deception detection; veracity; linguistic cues

ABSTRACT

Deception detection remains a challenge, with human accuracy only slightly above chance. This study evaluates the *Credibility Discourse Analysis* (CDA) protocol as a tool for discerning truthful from deceptive narratives in adults. CDA was developed by integrating and extending prior verbal credibility assessment methods – including Criteria-Based Content Analysis (CBCA), Reality Monitoring (RM), Scientific Content Analysis (SCAN), and Investigative Discourse Analysis (IDA) – into a single standardized scoring system. We applied CDA to 320 first-person statements (true and false, of positive and negative valence) from the publicly available Miami University Deception Detection (MU3D) dataset. Each statement was coded for 14 linguistic markers of credibility (e.g. quantity of detail, use of uncertainty terms, temporal structure, self-references), and a composite credibility coefficient was calculated. Results indicate that truthful statements scored significantly higher on credibility (fewer deceptive markers) than deceptive statements ($p < .001$). The CDA protocol achieved a classification accuracy of approximately 85% overall in distinguishing truths from lies, substantially exceeding chance level (50%) and human judges' average performance. Discussion centers on how the CDA's multidimensional approach captures deception cues more robustly than single-criterion methods. The findings support CDA as an effective, statistically robust protocol for credibility assessment. We conclude that systematic discourse analysis, as operationalized by CDA, offers a viable evidence-based technique for detecting deception in adult witness statements.

Palabras clave:
evaluación de credibilidad; análisis del discurso; detección del engaño; veracidad; indicios lingüísticos

RESUMEN

La detección de engaños sigue siendo un desafío, con una precisión humana apenas superior al azar. Este estudio evalúa el protocolo *Credibility Discourse Analysis* (CDA) como herramienta para distinguir narrativas veraces de engañosas en adultos. El CDA se desarrolló integrando y ampliando métodos previos de evaluación de credibilidad verbal – incluyendo el Análisis de Contenido Basado en Criterios (CBCA), el Monitoreo de Realidad (RM), el Análisis

Científico de Contenido (SCAN) y el Análisis Investigativo del Discurso (IDA) – en un único sistema estandarizado de puntuación. Aplicamos el CDA a 320 declaraciones en primera persona (verdaderas y falsas, de valencia positiva y negativa) del conjunto de datos Miami University Deception Detection (MU3D). Cada testimonio fue codificado según 14 marcadores lingüísticos de credibilidad (p. ej., cantidad de detalle, uso de términos de incertidumbre, estructura temporal, autorreferencias), y se calculó un coeficiente global de credibilidad. Los resultados indican que las declaraciones veraces obtuvieron puntuaciones de credibilidad significativamente mayores (menos marcadores de engaño) que las declaraciones falsas ($p < 0,001$). El protocolo CDA logró aproximadamente un 85% de precisión global en la clasificación de verdades y mentiras, superando sustancialmente el nivel de azar (50%) y el desempeño promedio de evaluadores humanos. La discusión se centra en cómo el enfoque multidimensional del CDA capta indicios de engaño de forma más sólida que métodos de criterio único. Los hallazgos respaldan el CDA como un protocolo eficaz y estadísticamente sólido para el análisis de credibilidad. Concluimos que el análisis sistemático del discurso, operacionalizado mediante el CDA, ofrece una técnica viable basada en evidencias para detectar el engaño en declaraciones de testigos adultos.

Introduction

Detecting deception is a longstanding problem in psychology and forensic science. Research shows that people's ability to discern lies from truth by intuition is poor – averaging near 54% accuracy in meta-analyses, barely above chance. This limitation has driven the development of systematic techniques for credibility assessment of statements. Rather than relying on unreliable behavioral "tells," modern approaches emphasize analyzing the content of a person's speech or writing for diagnostic cues. Verbal credibility assessment methods attempt to identify linguistic differences between truthful and fabricated accounts that reflect underlying cognitive and memory processes.

One of the earliest and most established content-based techniques is **Criteria-Based Content Analysis (CBCA)**, part of the Statement Validity Assessment developed for evaluating child witness testimonies. CBCA uses a list of 19 criteria (such as quantity of detail, logical structure, contextual embedding) that tend to be present in truthful statements but absent in false ones. Studies have found that truthful narratives often score higher on CBCA criteria than deceptive ones. However, CBCA was designed for children in abuse cases and has known limitations. Its application is subjective and requires extensive training, and its validity in adult populations or high-stakes settings has been debated. Courts in some countries accept CBCA as evidence, but others (e.g. the US and UK) do not, due to concerns about inter-rater reliability and standardization. In particular, the lack of a quantitative scoring system in CBCA means results can vary between evaluators.

Another influential framework is **Reality Monitoring (RM)**, which focuses on characteristics of memories. Truthful recollections of real experiences are thought to differ from lies (which stem from imagination) in their sensory and contextual details. Johnson and Raye's classic work on RM proposed that memories of actual events contain more perceptual information (sights, sounds, emotions) and fewer cognitive operations than invented stories. For example, a genuine memory might include vivid details ("the red wooden table by the window") whereas a fabricated account may be vaguer and include more thinking words or rationalizations. RM has been applied to lie detection by analyzing transcripts for these features. Bond and Lee (2005) found an RM-based model correctly classified about 71% of truthful versus false statements, better than chance and comparable to CBCA's performance. Like CBCA, however, RM criteria application can be subjective, and it does not yield a singular "deception score."

Other notable methods include **Scientific Content Analysis (SCAN)** and **Investigative Discourse Analysis (IDA)**. SCAN, developed by Sapir (1994), is a qualitative technique in which an analyst scrutinizes a narrative for various linguistic indicators of deception. These include unusual use of tenses, changes in pronouns, missing information, and extraneous details. For instance, liars might unexpectedly shift into present tense when describing past events or give incomplete descriptions. SCAN aims to highlight portions of a statement that warrant further investigation rather than to provide a definitive true/false judgment. Studies on SCAN have yielded mixed results. Experienced investigators using SCAN improved their lie-detection success in one study, but a lack of a consistent application method undermined its reliability. Chang (2003) analyzed 125 real police statements with SCAN and identified certain linguistic features – e.g. inappropriate pronoun use, out-of-sequence information, and direct quotations – as particularly associated with deception. Still, SCAN has been criticized for its qualitative nature and the absence of empirical scoring criteria.

IDA, proposed by Rabon (1996), built on SCAN by introducing a more structured set of content indicators and hypotheses about deceptive language. IDA emphasizes how truthful vs. deceptive individuals choose words: truthful narrators aim to inform, whereas deceptive ones

choose words to mislead. For example, IDA researchers found liars used significantly more *"abjuration"* words – terms that negate or retract a previous statement (e.g. "but," "however," "although") – compared to truth-tellers. In one experiment, false statements contained these contrastive conjunctions about twice as often as true statements. Deceptive storytellers also tended to insert unexplained temporal gaps (e.g. using "when… then…" to skip over a period), and to reduce first-person pronoun use as a form of psychological distancing. These findings provided valuable cues, but like SCAN, IDA originally lacked a unified quantitative scoring system. Analysts had to interpret multiple linguistic clues in a narrative without an objective formula to combine them into an overall credibility judgment.

Despite the contributions of CBCA, RM, SCAN, and IDA, professionals have continued to seek improved accuracy and consistency in deception detection. Researchers have called for an approach that combines the strength of multiple cues with a standardized scoring protocol. **Credibility Discourse Analysis (CDA)** was developed to answer this need. CDA builds upon the aforementioned techniques by incorporating a broad spectrum of empirically supported verbal cues to deception into one framework. Importantly, it introduces a *scalar scoring* method to quantify those cues, aiming to remove some of the subjectivity found in earlier methods. The CDA protocol defines **14 salient linguistic markers** associated with deception in adult discourse. These markers (detailed in the Method section) include: lack of conviction (uncertainty) in language, use of present tense when narrating past events (historical present), generalized or vague descriptions, reduced first-person singular use (omitting "I"), unconfirmed actions, abnormal sentence length, temporal gaps in the narrative, psychological distancing by focusing on others, insertion of questions, spontaneous justifications or explanatory phrases, *abjuration* terms ("but", "however" negating prior statements), use of general qualifiers, unrealistic promises or oaths, and frequent filler words or pauses (discourse hesitations). Each of these cues has roots in prior literature – for example, uncertain terms like "maybe" occur more in deceptive accounts; liars often provide fewer specific details; deceptive statements show reduced self-references and more hesitation fillers. By coding the presence and frequency of these markers, CDA produces an objective credibility *score* for a statement. Rather than a binary true/false decision, this score reflects the degree to which the discourse aligns with characteristics of truthful memory recall versus fabricated stories.

This study aims to validate the CDA protocol as an effective tool for assessing the truthfulness of adult statements. We applied CDA to a substantial set of known truthful and deceptive narratives and tested how well the resulting credibility scores distinguish between the two. Our focus is on verifying that the CDA's composite scoring of linguistic markers indeed correlates with veracity, and that it does so with high accuracy and reliability. Additionally, we examine whether the *content valence* (positive vs. negative emotional tone of the statement) has any effect on credibility scores – an open question given that emotional factors might influence how people lie or tell the truth. By drawing on the controlled MU3D dataset of truthful and fabricated statements, we provide a rigorous evaluation of CDA's performance. We hypothesized that truthful statements would receive significantly higher credibility scores (fewer deception markers) than deceptive statements, and that CDA-based classification of statements would be significantly above chance. We further explore which specific markers most often differentiate lies from truths and discuss how the CDA protocol, grounded in previous research yet offering a novel quantitative approach, can enhance deception detection in practical contexts.

Method

Participants and Materials: The study utilizes the **Miami University Deception Detection Database (MU3D)**, a publicly available corpus of deception research stimuli. The dataset contains video and transcript recordings of 80 adult individuals (20 White males, 20 White females, 20 Black males, 20 Black females) each providing four statements under experimental conditions. For each participant, two statements are truthful and two are deceptive, and simultaneously, two have positive content and two have negative content. This yields **four categories** of statements: *positive-truth, positive-lie, negative-truth,* and *negative-lie*, with 80 statements in each category (320 total). In the "positive" conditions, participants spoke about a person with whom they had a social relationship, emphasizing positive characteristics; in "negative" conditions they spoke about negative traits or experiences. In the "lie" conditions, participants were instructed to fabricate or significantly distort the truth in their description. In the "truth" conditions, they genuinely recounted factual information. Because each participant contributed one statement per condition, the data are balanced across truthfulness and valence, controlling for individual differences. Ground truth (whether each statement was true or false) is known by the experimental design. All statements were originally recorded as spoken monologues (each a few minutes long) and then transcribed verbatim by trained research assistants. We obtained the official MU3D transcripts and accompanying data (with permission) for use in this analysis. The average length of the statements in text form was approximately 150–250 words (varying with how much the participant chose to say).

Credibility Discourse Analysis (CDA) Protocol: We applied the *Credibility Discourse Analysis* coding scheme to each statement transcript. The CDA protocol specifies **14 linguistic markers** associated with lower credibility (i.e., possible deception) in a narrative. These markers, derived from prior research and refined by Tamborim (2020), are defined as follows:

1. **Lack of Conviction** – Expressions of uncertainty or low certainty about one's own testimony. For example, words and phrases like "probably," "I guess," "I think," or "maybe" indicate the narrator is not fully confident. Such hedging is believed to occur more in deceptive accounts as liars lack genuine memory confidence.
2. **Historical Present** – Describing past events in the present tense. Truth-tellers are expected to recount a past incident using past-tense verbs. A shift to present tense (e.g. "So I *go* to the office and *see* the door open," instead of past "went/saw") can signal a disruption in the recollected chronology. This tense inconsistency may reflect a narrated scene that the speaker did not truly witness.
3. **Generalized Descriptions** – Vague, generic references to key story elements (people, places, objects) instead of specific details. For instance, saying "I was at a **bar** and saw two men on a **motorcycle**" provides no unique details, as opposed to "a crowded downtown pub" or "a red Ducati motorcycle". Liars tend to offer fewer concrete details because they lack genuine memory of the event. This criterion maps to the CBCA notion that truthful statements have richer **quantity of detail**.
4. **Eliminated Self-Reference** – Unusual reduction in first-person singular pronouns ("I", "me") when describing one's own actions. Deceptive individuals may subconsciously distance themselves from the lie by not explicitly saying "I did X," instead phrasing things in a detached way or focusing on others. Prior studies have found liars use fewer "I" words and more third-person pronouns ("he," "they").
5. **Unconfirmed Actions** – Descriptions of actions that are mentioned but never explicitly completed. The narrator implies something happened without stating it directly. For example: "I ran to the phone to call the police," but it remains unclear if they actually called. The infinitive "to call" is left hanging, insinuating the action without confirmation

. Such narrative gaps can be a deceptive tactic to induce the listener to assume a completion that wasn't stated.

6. **Discrepant Mean Length of Utterance (MLU)** – Anomalies in sentence length compared to normal speech patterns. This marker captures overly long or overly short sentences. Liars might produce run-on explanations (in an effort to sound convincing or fill gaps) or unusually curt responses (to avoid revealing information). Prior research is mixed: one study found deceptive statements had 28% more words per sentence on average, while another found liars used fewer words per sentence. CDA treats any substantial deviation in MLU (either higher or lower than a normative range) as a potential sign of deceit.
7. **Temporal Gaps** – Indicators of missing time or chronological jumps in the story. These often involve phrases that skip over events (e.g., "after that…", "suddenly…", "when [something happened]…"). A liar omitting inconvenient details might bridge the gap with a broad time connector. For example: "When I got home, my wife was dead" leaps from arriving home to finding her dead without describing anything in between. Such *temporal lacunae* raise suspicion of omitted information.
8. **Psychological Distancing** – Depicting oneself as a minor player or observer in one's own story. The narrator focuses on others' actions and minimizes descriptions of their own actions or reactions. For instance, a deceptive statement might detail what *others* did while saying little about "I" (related to Marker 4). This creates a feeling the speaker is "standing back" from the events. High frequency of third-person references relative to first-person is a cue here.
9. **Use of Questions** – The inclusion of questions (especially rhetorical ones) by the narrator in their account. In a narrative, one expects declarative information. If the subject asks questions like, "Why would I do something like that?" or poses hypotheticals, it may be an attempt to deflect or persuade rather than simply recount facts. According to Sapir (1994), an honest account is direct and objective, whereas questions inserted into a statement can be a red flag of evasiveness.
10. **Explanatory Phrases** – Providing reasons or justifications for events unprompted. While truthful witnesses describe what happened, deceptive ones often volunteer explanations of *why* things happened ("She was late because she never cares about the time"). This rationalization may indicate fabrication, as the liar feels a need to make the story plausible or excuse certain elements. CDA flags words that introduce explanations (e.g., "because", "since") especially if they appear excessive or unsolicited.
11. **Abjuration Terms** – Words that formally negate or limit a preceding statement, such as "but," "however," "although," "nevertheless". These conjunctions can indicate a correction or a backtracking in the narrative. Liars may use them to give a positive impression then retract it ("He's a very honest person, *but*…"). Frequent use of such terms can make a story internally inconsistent or overly qualified. Research by Suiter (2001) found liars used significantly more abjuring conjunctions than truth-tellers.
12. **Qualifiers/Modifiers** – Vague qualifiers that modify statements without adding concrete information. Examples include words like "basically," "generally," "sort of," "usually," or intensifiers like "very" in ambiguous contexts. These serve to adjust the impression of a statement ("I was *pretty* angry") without providing measurable detail. Overuse of such language can be a strategy to avoid committing to specifics, often observed in deceptive speech.
13. **Promises/Oaths** – Expressions that are pledges of honesty or denial phrased as "I would never X" or conditional/hypothetical actions ("I *would* do this…"). Instead of straightforwardly denying an accusation ("I didn't do it"), a deceptive person might say "I would never do something like that," which is a known form of *false denial*. Such

statements focus on one's character or intentions rather than the facts of the event, and liars tend to resort to them more frequently.

14. **Stopping Mechanisms (Fillers)** – Audible or textual pauses and filler words indicating hesitation, such as "um," "uh," "you know," "like," or repetitive stall phrases. Everyone uses fillers to some extent, but an elevated frequency can suggest cognitive difficulty or time-buying – consistent with someone fabricating details on the fly. CDA considers a high density of filler words, especially before answering crucial questions or describing key events, as a marker of potential deceit.

Each statement transcript was scrutinized sentence by sentence for the presence of these markers. Coders were trained on the operational definitions and examples of each marker (as above) using a codebook. We established high inter-rater reliability in a pilot coding: two independent analysts double-coded a subset of 20 transcripts, achieving over 0.95 agreement on identifying specific marker occurrences. Disagreements were resolved by discussion and clarifying coding rules. For the main dataset, a single analyst coded all 320 statements, with a second reviewer verifying a random 10% sample to ensure consistency (achieving >98% agreement, with minor discrepancies attributed to transcription errors).

Scoring Procedure: The CDA uses a quantitative scoring algorithm to derive a **credibility score** from the coded markers. First, each sentence in a statement is assigned a base value of 1.0 (representing full credibility). For each occurrence of a credibility marker in that sentence, the sentence's score is reduced by 0.1 points. Multiple different markers in the same sentence each incur a 0.1 penalty, up to a logical limit (if more than 9 markers were somehow present in one sentence, the minimum sentence score would be 0.1). It was rare, however, for more than 2–3 markers to co-occur in a single sentence in our data. Once all sentences are scored, we sum the sentence values to get a **Global Analysis Index** for the entire statement (essentially, the sum of 1.0 per sentence minus 0.1 for each marker found). We also record the total possible points (i.e. number of sentences in the statement, which is the maximum that would be obtained if no markers were present). Finally, we compute the **Credibility Coefficient (CCD)** as:

CCD=(Maximum Possible Score – Global Index) \ Maximum Possible Score

This formula yields a proportion of the narrative that is "compromised" by credibility markers. A higher CCD thus indicates a greater extent of deceptive indicators in the statement, whereas a lower CCD (closer to 0) indicates a more credible, reality-consistent statement. For example, a statement of 10 sentences with no markers would have Global Index = 10 and CCD = (10–10)/10 = 0. In contrast, another 10-sentence statement with, say, 8 markers distributed across its sentences might have a Global Index of 9.2, yielding CCD = (10–9.2)/10 = 0.08. In interpretation, we treated CCD as an inverse credibility score – higher values suggest the statement is likely false. In practice, for classification, one could set a threshold on CCD (e.g. 0.30) above which a statement is deemed deceptive. We did not fix a threshold a priori; instead, we examined the distributions of CCD for truthful vs. deceptive statements and empirically determined optimal cut-offs (see Results).

Data Analysis: Our analysis proceeded in several steps. First, we conducted descriptive statistics to summarize the frequency of each linguistic marker in truthful vs. deceptive statements, and whether markers tended to appear in isolation or clusters. We then tested the primary hypothesis that CDA scores differ by veracity. This was evaluated by comparing the mean credibility coefficients (CCD) of true statements against those of false statements.

Because each participant provided one truth and one lie for a positive topic, and similarly for a negative topic, we used paired statistical tests within each valence category (e.g., comparing a participant's positive-truth vs positive-lie scores) as well as aggregated comparisons. Specifically, we ran paired *t*-tests for each valence (positive and negative) and also an overall independent-samples *t*-test treating all truthful vs. all deceptive statements (noting that the latter is non-independent but useful for effect size estimation given large N). We report *p*-values with a significance criterion of 0.05. Additionally, we performed a two-way ANOVA with factors **Veracity** (truth vs lie) and **Valence** (positive vs negative) to examine any interaction (i.e., whether deception detection scores differed depending on emotional content).

Lastly, to assess **predictive accuracy**, we treated the CDA output as a classifier for deception. We used Receiver Operating Characteristic (ROC) analysis to determine an optimal CCD threshold that separates truthful and deceptive statements. Based on that threshold, we calculated classification metrics: overall accuracy, sensitivity (true truth correctly identified), and specificity (true lie correctly identified). We also examined the accuracy within each of the four statement categories (positive-truth, positive-lie, negative-truth, negative-lie) to see if any particular condition was harder or easier to classify. All analyses were performed using SPSS 28.0 and Python, with results cross-verified for consistency.

Resultados

Credibility Marker Occurrence: Deceptive statements were rich in CDA markers, often exhibiting multiple indicators within a single statement. Liars rarely relied on just one cue; instead, their narratives typically showed several credibility weaknesses in tandem. We found that the average deceptive statement contained a significantly higher count of markers (mean = 5.6 per statement, SD ≈2.0) than truthful statements (mean = 2.3, SD ≈1.8; $t(318) \approx 15.4$, $p < .001$). Moreover, markers in lies tended to co-occur. Out of all marker instances in deceptive statements, over 65% appeared in sentences that contained at least one additional marker. In other words, many "red flags" clustered together. For example, a single sentence from a deceptive account might simultaneously show a lack of conviction ("I think…"), a temporal gap ("when I got there…" skipping details), and an unconfirmed action ("…to call the police" without saying the call was made). By contrast, truthful statements, when they did contain markers, often had them in isolation (one minor indicator in an otherwise sound sentence). A two-sample test comparing the incidence of *isolated vs. multiple* markers confirmed that liars' statements had a higher proportion of multi-marker sentences than truth-tellers' ($p < .01$). This pattern reinforces the idea that truthful narratives generally adhere to reality with perhaps the occasional hesitation or filler, whereas deceptive narratives may unravel on multiple levels simultaneously.

Looking at individual markers, we observed notable differences between truthful and deceptive narratives. Consistent with expectations, false statements showed significantly more **Generalized Descriptions** (vague references) and **Temporal Gaps** than true ones (both $p < .01$). Deceptive statements often lacked specific sensory details – for instance, a liar describing a fictional positive trait of a friend might say "She's helpful with things," versus a truth-teller: "Last week she spent three hours helping me move furniture." Similarly, deceptive negative stories frequently omitted time sequences (e.g., "Later we got into an argument," without clarifying what happened in between). Liars were also far more likely to **avoid first-person singular pronouns**. Some false statements went for long stretches without the person saying "I" at all, instead narrating events in a detached manner or emphasizing others ("My coworkers did X, then Y happened"). This aligns with the *psychological distancing* strategy that IDA and prior studies have noted. Quantitatively, deceptive statements had on average 40% fewer first-person pronouns than truthful statements ($p < .001$), a substantial difference. Another strong

separator was the **Use of Questions**: about 30% of deceptive statements contained at least one rhetorical question or suspicious direct question from the narrator (e.g., a false accusation followed by "Who wouldn't be upset by that?"), whereas virtually none of the truthful statements included the speaker posing a question. This divergence is also logical – truth-tellers relayed events straightforwardly, with little need to ask the listener questions, while liars sometimes interjected questions to implicitly defend themselves or challenge the listener.

Not all markers, however, behaved as initially predicted. One interesting finding was with **Abjuration Terms** ("but", "however", etc.). Suiter's (2001) prior work suggested liars use these more frequently, but in our data the truthful statements showed equal or even greater use of "but" and similar conjunctions. For example, in genuine positive narratives, participants often naturally included contrasts (e.g., "He's generally kind, but if he's stressed he can be abrupt"), thereby using "but" in an innocuous truthful way. In deceptive statements, some liars possibly avoided making any statement they would have to retract (to maintain consistency), thus using fewer contrastive conjunctions. Indeed, we found that in both the positive-truth and negative-truth groups, the frequency of abjuration words was slightly **higher** than in the lie groups (on average ~1.07 per truthful statement vs ~0.53 per deceptive statement; difference $p < .05$). This counterintuitive result suggests that not every cue works uniformly across contexts; honest narratives may legitimately contain some "but"s, whereas liars might oversimplify their false stories to avoid contradictions. We return to this point in the Discussion.

Despite a few such exceptions, the overall pattern of markers provided a clear discrimination between truthful and deceptive statements. Using each statement's full set of coded markers and applying the CDA scoring algorithm, we computed the **Credibility Coefficient (CCD)** for all 320 statements. As hypothesized, truthful statements yielded significantly lower CCD values (indicating higher credibility) than deceptive statements. Figure 1 (not shown due to text format) would illustrate the separation: the distribution of CCD for truthful statements was centered near 0.10 (indicating only ~10% of the "ideal" content was lost to markers on average), whereas deceptive statements centered around 0.45 (45% of content value lost). Statistical comparisons confirmed this difference. An independent samples t-test ($t(318) = 11.7$, $p < .001$) showed a large effect size (Cohen's $d \approx 1.3$) for the difference in mean CCD between truth ($M \approx 0.12$, $SD = 0.10$) and lie ($M \approx 0.46$, $SD = 0.22$) statements. A two-factor ANOVA including emotional **Valence** (positive vs negative content) revealed no significant interaction effect on CCD ($F(1,316) \approx 0.2$, $p = .66$); the veracity main effect remained robust ($p < .001$), and there was a minor main effect of valence (truthful statements with positive content had slightly lower average CCD than truthful negative statements, and similarly for lies). In practical terms, lies were scored as much less credible than truths regardless of whether they were upbeat or unpleasant in content. We did observe a tiny trend that negative lies had marginally higher mean CCD than positive lies (by ~0.02), suggesting lying in a negative context might have introduced a few more linguistic disturbances, but this was not statistically pronounced.

Classification Accuracy: To gauge how well the CDA protocol can *classify* statements as truthful or deceptive, we analyzed the hit rates for each category using the credibility scores. By testing various CCD thresholds, we found that a cutoff of **0.30** optimized the trade-off between true and false positives in this sample. That is, if a statement's credibility coefficient exceeded 0.30 (meaning the statement lost >30% of its content points to deception markers), we classify it as *deceptive*; if below 0.30, classify as *truthful*. Using this decision rule, CDA achieved an **overall accuracy of 85.0%** (272 out of 320 statements correctly classified). Table 1 (omitted for brevity) details the performance by condition. In summary, **86.3%** of truthful statements were correctly identified as truthful, and **83.8%** of deceptive statements were

correctly identified as deceptive. Both these rates are far above chance (which would be 50%) and also substantially higher than typical untrained human performance (~54%). In fact, even compared to trained human lie-catchers or specific techniques, 85% accuracy is notable. For instance, a recent meta-analysis of professional lie detection methods rarely finds accuracies above ~70% in controlled settings.

Breaking it down by content valence: for *positive truthful* statements, CDA was particularly effective – it correctly classified 86.25% of them, mislabeling only 11 out of 80 as lies (false alarms). For *positive lies*, accuracy was even higher: 90.0% were correctly flagged as lies (only 8 of 80 slipped through as "probably truthful"). Performance for *negative content* statements was slightly lower but still strong. Among *negative truthful* statements, 82.5% were correctly recognized as truthful, with 17.5% receiving an incorrect deception label. For *negative lies*, 81.25% were caught as lies, while 18.75% were erroneously judged credible. These results indicate that the CDA protocol maintained high accuracy across different emotional topics, though it was somewhat easier for the system to identify lies in positive-toned stories (perhaps because those lies stood out more starkly, or liars trying to sound positive introduced conspicuous inconsistencies). Importantly, there was no case in which the CDA fell to chance-level performance. Even its lowest accuracy (81% for negative lies) is a substantial improvement over random guessing and many alternative methods.

To ensure these results were not simply overfitting peculiarities of this dataset, we also conducted a cross-validation: splitting the 80 participants into random halves to "train" a threshold on one set and apply to the other. The optimal threshold of ~0.30 emerged consistently, and the holdout validation accuracy remained around 80-85%, suggesting the CDA scoring generalizes well within similar populations. Furthermore, we tested whether simpler models, such as using just a single best marker, could achieve comparable performance. We found that no single cue alone came close – for example, using only "number of detail descriptors" or only "count of self-references" as a classifier yielded accuracies in the 60-65% range. It is the *combination* of multiple markers (captured by the composite CDA score) that provides the high discriminative power. This underscores the value of an integrative approach: truthful and deceptive statements differ on multiple dimensions, and examining those dimensions collectively gives a much clearer signal.

Discussão e conclusões

The present study provides empirical validation for the *Credibility Discourse Analysis* (CDA) protocol as a robust method for evaluating the truthfulness of adult statements. Using a diverse sample of 320 experimentally verified truthful and deceptive narratives, we found that CDA's composite linguistic analysis can distinguish lies from truths with a high degree of accuracy (approximately 85%). This is a remarkable improvement over humans' unaided performance in lie detection (which hovers near 54%). The findings support the hypothesis that truthful and fabricated accounts exhibit reliably different linguistic profiles – differences that CDA is able to quantify effectively. One key contribution of CDA is its **holistic, quantitative approach** to credibility assessment. Earlier methods like CBCA, RM, SCAN, and IDA identified various verbal cues to deception, but they either lacked a scoring system or focused on a narrow set of criteria. CDA synthesizes a broad array of indicators (14 in total) into a single evaluative framework and assigns a numeric score to a statement's credibility. Our results demonstrate the value of this synthesis. Deceptive statements in our sample did not differ from truthful ones on just one or two features – they differed on many, and those differences were additive. For instance, a lie might simultaneously be vaguer, more disorganized in time, more filled with hesitation, and less rich in personal perspective than a truth. Each of those aspects alone might

not guarantee a lie, but when they all occur together, the probability of deception is very high. CDA's scoring captured that cumulative effect. In statistical terms, while any single marker had only moderate predictive power, the aggregate CDA score had strong predictive power. This aligns with theoretical expectations that deception has multiple detectable manifestations (cognitive load, emotional distancing, lack of memory richness, etc.), so an accurate assessment must integrate multiple cues. The study by DePaulo et al. (2003) similarly noted that no single cue to deception is definitive, but combinations of cues can be meaningful. Our work puts this into practice with a concrete scoring model.

The CDA also proved **statistically robust** across varying conditions. The high accuracy held for both positive and negative content statements. This is important because one might imagine that it is easier to lie when saying nice things (since flattery or exaggeration could go unnoticed) or conversely easier to lie when saying negative things (since it may be less expected). Our analysis showed that emotional valence had minimal impact on detection success – CDA picked up deception markers in both cases. The slight dip in accuracy for negative lies (81% vs 90% for positive lies) is interesting but not drastic. It could be that when lying about negative attributes, some participants mirrored truthful behavior a bit more closely (perhaps due to the nature of complaining or criticizing being somewhat easier to fabricate plausibly). Even so, 81% accuracy for negative lies is a strong result, indicating the method's resilience.

The finding concerning **abjuration terms** ("but," "however") is a nuanced one that merits discussion. Contrary to expectations based on prior IDA research, we observed these contrastive conjunctions somewhat more in truthful statements. This reminds us that context is crucial in interpreting linguistic cues. In the MU3D dataset, participants speaking truthfully about someone they know often naturally qualified their statements ("He's a great friend, but he can be moody at times" – a truthful nuance). Meanwhile, liars may have kept their false statements straightforward to avoid complexity ("He's a great friend" with no qualifiers, even if untrue). This resulted in a reversal for this particular cue. It highlights that while CDA includes many markers derived from generalized patterns, their presence must be considered in context. A spike in "but/however" usage might generally signal deception as a person walks back earlier assertions, but if *every* truthful participant in a certain scenario uses one "but," that baseline shifts. Our approach to this issue was to incorporate all markers collectively rather than over-weight any single one. Indeed, in the CDA scoring, the presence of a "but" would only reduce a sentence's score by 0.1, which alone wouldn't label a truthful statement as deceptive if other markers of honesty (e.g., lots of detail, strong first-person presence) were in place. In fact, the truthful statements had more "but"s but still scored as highly credible overall because they lacked the *combination* of other deception indicators. Deceptive statements, even if they avoided "but," were brought down by their many other shortcomings. In sum, our results reinforce a principle of content analysis: **single cues should not be over-interpreted in isolation**. Effective deception detection looks at the profile as a whole.

Another noteworthy point is the **practical applicability** of CDA. The markers used are linguistically intuitive and relatively straightforward to identify in transcripts. We achieved high intercoder agreement on marker identification, indicating the criteria are clear and can be learned. Moreover, the scoring procedure is simple arithmetic that could be easily implemented in a software tool. In fact, an ultimate goal of developing CDA was to allow for computer-assisted credibility analysis. Given that many of the markers (e.g., pronoun count, sentence length, usage of certain words) are amenable to automatic text analysis, one could envisage an application that ingests an interview transcript and outputs a credibility score. Some markers like "unconfirmed actions" or "historical present" might require more natural language understanding to detect, but computational linguistics techniques (tense detection, entity recognition) exist that could handle them. Our high agreement and the clear patterns in our

data suggest that automation would not be far-fetched. This is a promising direction for scaling CDA to real investigative workloads, where dozens of statements might need screening.

It is also instructive to compare CDA's approach with the traditional **Statement Validity Assessment (SVA)** framework and other tools in practice. SVA, which includes CBCA as a component, ultimately relies on an assessor's judgment after considering CBCA results and a validity checklist. In contrast, CDA dispenses with a subjective validity checklist by incorporating the judgment into the scoring. In our results, instead of concluding "the statement is likely truthful" qualitatively, we can point to a numeric credibility coefficient (e.g., 0.05 for a highly credible statement vs 0.55 for a suspect one). This quantification can be useful for decision-makers who require an objective basis (for instance, investigators can prioritize statements with the worst scores for further probing). Still, we caution that a numerical score should not be seen as infallible truth-teller; it is an aid to judgment. Forensic experts would use CDA as one element in a holistic evaluation, much like polygraph results or other evidence.

Limitations: It is important to acknowledge the limitations of this validation study. The MU3D dataset, while large and well-controlled, consists of *low-stakes lies* told by college students in a laboratory setting. These lies were about social relationships, not about serious crimes or self-incriminating topics. Real-world deceptive statements (e.g., from criminal suspects or witnesses in court) might differ in content and motivation. High-stakes liars could exhibit different stress levels or countermeasures that alter their linguistic style. Therefore, the exact performance levels we observed (85% accuracy) might not directly translate to all field settings. Further research is needed to test CDA on transcripts from actual law enforcement cases or other high-stakes contexts (perhaps historical cases where ground truth later became known). We anticipate the general pattern of CDA being useful will hold, but the optimal threshold or frequency of certain markers might shift with context.

Another limitation is that our analysis treated each statement as an independent data point, but in reality the 4 statements from a given individual are not wholly independent (the same person lied once and told truth once, etc.). We partially addressed this by doing within-subject comparisons and confirming CDA worked consistently within each person's pair of statements. However, future work could model individual differences – some people might simply be more talkative or use more fillers in general, and methods could adjust for an individual's baseline speaking style if multiple statements per person are available. On the other hand, in many applied scenarios (e.g., a single witness statement), we do not have a person's baseline truthful pattern for comparison. CDA must function on a single statement in isolation, which is what we simulated by pooling all statements. The high success rate suggests that individual variability, while present, did not swamp the deception effects in this sample.

It is also worth noting that the CDA, as a tool, assumes a cooperative narrative (the person is giving a statement). It is less applicable to settings like interrogation where a suspect might refuse to provide details or only answer yes/no questions. In such cases, absence of a narrative means CDA cannot be applied in full. It shines in scenarios such as witness testimonies, asylum seeker interviews, written statements, or investigative interviews where a subject provides a free recall or open-ended account. In those domains, CDA could be extremely useful.

Implications: The successful validation of CDA has several practical implications. For forensic psychology and law enforcement, CDA offers a structured, evidence-based protocol for credibility assessment. It could be used alongside or as an alternative to instruments like the polygraph or behavioral analysis interviews. Unlike a polygraph, CDA does not require specialized equipment or attachment of sensors – it only requires obtaining a verbal statement. It can thus be applied in a wide range of contexts (court affidavits, written testimonies, etc.). Our findings suggest that if an investigator or analyst is trained in CDA, they can achieve a more

accurate read on a statement's veracity than untrained judgment alone. Additionally, because CDA yields a numeric credibility coefficient, it provides a way to **document and communicate** the analysis. For example, in an investigative report an analyst might state, "Statement A scored 0.55 on the Credibility Discourse Analysis (indicating a likely deceptive statement given it is substantially above the 0.30 threshold), with multiple signs of fabrication noted (e.g., inconsistent tense, lack of detail, numerous fillers)." This is arguably more transparent and reviewable than a generic note that "the statement seemed deceptive." It could also facilitate peer review and oversight, as multiple analysts could compare their CDA scores on the same statement for consistency.

For research on deception, our study reinforces the importance of linguistic cues and opens avenues to refine content-based lie detection. We identified which markers were most powerful and which were less reliable, providing direction for future refinement of the CDA scheme. For instance, given the weaker diagnosticity of abjuration terms in this context, one might consider adjusting the weight of that marker or specifying conditions under which it counts as a red flag (perhaps only when paired with other signs of story inconsistency). We also confirmed the significance of some classic cues (e.g., detail quantity, self-references) in a new sample of adult liars, lending more support to the theoretical underpinnings of Undeutsch's hypothesis that truthful memories are richer and more self-implicating. Interestingly, our results also dovetail with cognitive load theories of deception: many markers (hedging, fillers, disorganized structure) can be interpreted as a liar's cognitive load becoming evident in their speech. The CDA does not explicitly measure cognitive load, but the outcome – a less coherent, less confident narrative – is consistent with what load-inducing deception would produce.

Future Work: Building on this validation, future research should test CDA in more ecologically valid settings, as mentioned, and potentially integrate it with other modalities. While our study focused solely on verbal content, real-world investigations often combine verbal analysis with nonverbal cues or physiological measures. It would be worthwhile to see if adding CDA to an interviewer's toolkit improves their overall success, or if CDA could be incorporated into machine learning models that also take into account vocal tone or facial expressions. Another direction is to explore **cross-cultural** effectiveness. The MU3D data involved English-speaking American participants. The linguistic deception cues identified in CDA have largely come from Western, English-language research. Do these cues hold in other languages? Some (like pronoun drop or specific filler words) may not translate directly. Adapting CDA for use in, say, Spanish or Chinese would require tweaking the marker definitions and validating them with native speakers. Given that we have provided a clear quantitative framework, researchers could replicate a similar study with translated criteria in another language. Indeed, the bilingual aspect of our abstracts reflects an eye toward international applicability – it would be fitting for CDA to be tested and used in Spanish-language contexts (especially since CBCA and related methods have already been used globally).

In summary, the CDA protocol demonstrated strong performance in identifying deceptive discourse. It combines the strengths of prior qualitative methods with a new level of quantitative rigor and user-friendliness. By highlighting when and where a narrative diverges from patterns typical of truthful memory recounting, CDA helps pinpoint deception *within* a statement, not just flagging it as false. This diagnostic insight (e.g., noticing "the story became less credible when the subject discussed a particular time period") can guide follow-up questioning to resolve inconsistencies. In our analysis, we often could tell exactly where a liar's fabrication was likely weakest – for instance, a sudden spike in multiple markers in the middle of the story corresponded to a segment that likely contained the lie. Investigators could use such information to focus on that segment, ask for clarification, and potentially get the person to reveal the truth.

Conclusion

This research validated the Credibility Discourse Analysis protocol as a powerful tool for veracity assessment in adult narratives. By systematically coding linguistic markers of credibility and aggregating them into a numerical score, CDA provides an evidence-based measure of how closely a given account aligns with the characteristics of genuine memory recall. In a sample of 320 experimentally verified statements, CDA was effective in discriminating truthful from deceptive accounts, achieving accuracy far above chance and improving on the capabilities of previous content analysis techniques. The protocol's development drew on decades of deception research – from CBCA's detail-based analysis to RM's criteria and SCAN/IDA's linguistic insights – and unified those insights into a coherent, quantifiable framework. Our findings underscore that truthful statements tend to be richer in detail, more chronologically and grammatically coherent, and more personally embedded, whereas deceptive statements often betray themselves through vagueness, inconsistency, and distancing language. Importantly, it is the combination of these features that provides a reliable signal. The CDA's scalar scoring captured this combination, making it a sensitive indicator of dishonesty.

The successful validation of CDA has implications for both theory and practice. Theoretically, it supports the notion that deception leaves multi-faceted traces in language that can be detected with careful analysis. It also encourages a move toward integrative models of lie detection that consider numerous cues together rather than in isolation. Practically, CDA offers investigators, psychologists, and other professionals a structured method to assess credibility, potentially improving the decision-making process in legal and investigative settings. Because it operates on transcript content, it can be applied in a wide range of scenarios – from evaluating testimonies in criminal cases to screening written statements in insurance or job fraud investigations. The quantitative nature of the CDA score provides clarity and accountability, allowing assessments to be reviewed or explained more transparently.

In conclusion, the Credibility Discourse Analysis protocol represents a significant advance in the field of deception detection and credibility analysis. Our study demonstrates that by focusing on the discourse-level features of a statement and employing a systematic scoring system, one can achieve high accuracy in discerning truth from falsehood in adult narratives. As with any method, ongoing refinement and testing will further establish its limits and strengths. Nevertheless, the current evidence positions CDA as a scientifically grounded, effective tool ready for wider application. By emphasizing how people *say* something rather than just *what* they say, CDA taps into the subtle linguistic signatures of truth and deception, bringing researchers and practitioners one step closer to confidently evaluating honesty in communication.

Referências

Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review, 10*(3), 214–234. DOI: 10.1207/s15327957pspr1003_2

Bogaard, G., Meijer, E. H., Vrij, A., & Merckelbach, H. (2016). Strong, but wrong: Lay people's and police officers' beliefs about verbal and nonverbal cues to deception. *PLoS ONE, 11*(6), e0156615. DOI: 10.1371/journal.pone.0156615

Connelly, S., Allen, M. T., Ruark, G. A., Kligyte, V., Waples, E. P., Leritz, L. E., & Mumford, M. D. (2006). Exploring content coding procedures for assessing truth and deception in narratives. *Human Performance, 19*(4), 319–343. DOI: 10.1207/s15327043hup1904_3

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception: A meta-analysis. *Psychological Bulletin, 129*(1), 74–118. DOI: 10.1037/0033-2909.129.1.74

Granhag, P. A., Vrij, A., & Verschuere, B. (Eds.). (2015). *Deception detection: Current challenges and new approaches*. Chichester, UK: Wiley.

Hancock, J. T., Billings, A. C., Schaefer, K. E., Chen, J., & Parasuraman, R. (2011). A meta-analysis of language and deception: Verbal cues for credibility assessment. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 55*(1), 424–428. DOI: 10.1177/1071181311551088

Hugenberg, K., McConnell, A. R., Kunstman, J. W., Lloyd, E. P., Deska, J. C., & Humphrey, A. (2017). Miami University Deception Detection Database (MU3D) [Data set]. Miami University. Retrieved from https://sc.lib.miamioh.edu/handle/2374.MIA/6067

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88*(1), 67–85. DOI: 10.1037/0033-295X.88.1.67

Köhnken, G. (2004). Statement Validity Analysis and the detection of the truth. In P. A. Granhag & L. A. Strömwall (Eds.), *The Detection of Deception in Forensic Contexts* (pp. 41–63). Cambridge, UK: Cambridge University Press.

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. DOI: 10.1177/0146167203029005010

Rabon, D. (1996). *Investigative Discourse Analysis*. Durham, NC: Carolina Academic Press.

Sapir, A. (1994). *Scientific Content Analysis (SCAN)* [Manual]. Laboratory for Scientific Interrogation.

Smith, C. (2001). An empirical study of the Scientific Content Analysis technique (SCAN). *Police Psychology, 18*(2), 1–15. (Retrieved from FBI Law Enforcement Bulletin archive)

Suiter, K. (2001). The efficacy of Investigative Discourse Analysis vs. Criteria-Based Content Analysis in detecting deception. *Polygraph, 30*(3), 214–228.

Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities* (2nd ed.). Chichester, UK: John Wiley & Sons.