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

(2025) MLS-Psychology Research, 8(2), 101-118. doi.org/10.33000/mlspr.v8i2.4003

VALIDACIÓN DEL PROTOCOLO DE ANÁLISIS DEL DISCURSO DE CREDIBILIDAD (CDA) PARA LA EVALUACIÓN DE VERACIDAD EN DECLARACIONES DE ADULTOS

Validation of Credibility Discourse Analysis (CDA) as Credibility Analysis Protocol for Adult Statements

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

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.

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.

Introducción

La detección del engaño es un problema de larga data en psicología y ciencias forenses. Las investigaciones demuestran que la capacidad de las personas para discernir la mentira de la verdad por intuición es escasa: la media de los metaanálisis se acerca al 54% de precisión, apenas por encima del azar. Esta limitación ha impulsado el desarrollo de técnicas sistemáticas para evaluar la credibilidad de las declaraciones. En lugar de basarse en "indicios" conductuales poco fiables, los enfoques modernos hacen hincapié en el análisis del contenido del habla o la escritura de una persona en busca de pistas de diagnóstico. Los métodos de evaluación de la credibilidad verbal intentan identificar diferencias lingüísticas entre los relatos veraces y los inventados que reflejen procesos cognitivos y de memoria subyacentes.

Una de las técnicas basadas en el contenido más antiguas y consolidadas es el **Análisis de Contenido Basado en Criterios (CBCA)**, que forma parte de la Evaluación de la Validez de las Declaraciones desarrollada para evaluar las declaraciones de testigos menores de edad. El CBCA utiliza una lista de 19 criterios (como la cantidad de detalles, la estructura lógica o la inserción contextual) que tienden a estar presentes en las afirmaciones veraces pero ausentes en las falsas. Los estudios han revelado que las narraciones veraces suelen puntuar más alto en los criterios del CBCA que las engañosas. Sin embargo, el CBCA se diseñó para niños en casos de malos tratos y tiene limitaciones conocidas. Su aplicación es subjetiva y requiere una amplia formación, y se ha debatido su validez en poblaciones adultas o en entornos de alto riesgo. Los tribunales de algunos países aceptan el CBCA como prueba, pero otros (por ejemplo, EE.UU. y el Reino Unido) no lo hacen, debido a la preocupación por la fiabilidad entre evaluadores y la estandarización. En particular, la falta de un sistema de puntuación cuantitativa en el CBCA significa que los resultados pueden variar entre evaluadores.

Otro marco influyente es el **Seguimiento de la Realidad (RM)**, que se centra en las características de los recuerdos. Se cree que los recuerdos veraces de experiencias reales difieren de las mentiras (que proceden de la imaginación) en sus detalles sensoriales y contextuales. El trabajo clásico de Johnson y Raye sobre RM propuso que los recuerdos de acontecimientos reales contienen más información perceptiva (imágenes, sonidos, emociones) y menos operaciones cognitivas que las historias inventadas. Por ejemplo, un recuerdo auténtico puede incluir detalles vívidos ("la mesa de madera roja junto a la ventana"), mientras que un relato inventado puede ser más vago e incluir más palabras pensadas o racionalizaciones. El RM se ha aplicado a la detección de mentiras analizando las transcripciones en busca de estas características. Bond y Lee (2005) descubrieron que un modelo basado en RM clasificaba correctamente alrededor del 71% de las declaraciones verdaderas frente a las falsas, mejor que el azar y comparable al rendimiento de CBCA. Sin embargo, al igual que el CBCA, la aplicación de los criterios RM puede ser subjetiva y no arroja una "puntuación de engaño" singular

Otros métodos destacados son el **Análisis Científico del Contenido (SCAN)** y el **Análisis Investigativo del Discurso (IDA)**. El SCAN, desarrollado por Sapir (1994), es una técnica cualitativa en la que un analista examina una narración en busca de diversos indicadores lingüísticos de engaño. Por ejemplo, uso inusual de los tiempos verbales, cambios en los pronombres, omisión de información y detalles superfluos. Por ejemplo, los mentirosos pueden cambiar inesperadamente al presente al describir hechos pasados o hacer descripciones incompletas. El objetivo de SCAN no es emitir un juicio definitivo de verdadero/falso, sino destacar las partes de un enunciado que merecen una investigación más profunda. Los estudios sobre el SCAN han arrojado resultados dispares. Los investigadores experimentados que utilizaron el SCAN mejoraron su éxito en la detección de mentiras en un estudio, pero la falta de un método de aplicación coherente minó su fiabilidad. Chang (2003)

analizó 125 declaraciones policiales reales con SCAN e identificó ciertas características lingüísticas -por ejemplo, uso inapropiado de pronombres, información fuera de secuencia y citas directas- como especialmente asociadas al engaño. Aun así, el SCAN ha sido criticado por su naturaleza cualitativa y la ausencia de criterios empíricos de puntuación.

IDA, propuesto por Rabon (1996), se basó en SCAN introduciendo un conjunto más estructurado de indicadores de contenido e hipótesis sobre el lenguaje engañoso. IDA hace hincapié en cómo eligen las palabras los individuos veraces frente a los engañosos: los narradores veraces pretenden informar, mientras que los engañosos eligen las palabras para despistar. Por ejemplo, los investigadores de la IDA descubrieron que los mentirosos utilizaban muchas más palabras de *"abjuración"* -términos que niegan o se retractan de una afirmación anterior (por ejemplo, "pero", "sin embargo", "aunque")- que los que decían la verdad. En un experimento, los enunciados falsos contenían estas conjunciones contrastivas aproximadamente el doble de veces que los enunciados verdaderos. Los narradores engañosos también tendían a insertar lagunas temporales inexplicables (por ejemplo, utilizando "cuando... entonces..." para saltarse un período) y a reducir el uso del pronombre en primera persona como forma de distanciamiento psicológico. Estos resultados proporcionaron valiosas pistas, pero al igual que el SCAN, la IDA carecía originalmente de un sistema unificado de puntuación cuantitativa. Los analistas tenían que interpretar múltiples pistas lingüísticas en una narración sin una fórmula objetiva para combinarlas en un juicio global de credibilidad.

A pesar de las aportaciones de CBCA, RM, SCAN e IDA, los profesionales han seguido buscando una mayor precisión y coherencia en la detección del engaño. Los investigadores han reclamado un enfoque que combine la fuerza de múltiples indicios con un protocolo de puntuación estandarizado. **El Análisis de la Credibilidad del Discurso (CDA)** se desarrolló para responder a esta necesidad. El CDA se basa en las técnicas antes mencionadas al incorporar en un marco un amplio espectro de indicios verbales de engaño respaldados empíricamente. Y lo que es más importante, introduce un método de *puntuación escalar* para cuantificar esos indicios, con el objetivo de eliminar parte de la subjetividad de los métodos anteriores. El protocolo CDA define **14 marcadores lingüísticos destacados** asociados al engaño en el discurso adulto. Estos marcadores (detallados en la sección Método) incluyen: falta de convicción (incertidumbre) en el lenguaje, uso del tiempo presente al narrar hechos pasados (presente histórico), descripciones generalizadas o vagas, uso reducido de la primera persona del singular (omisión del "yo"), acciones no confirmadas, longitud anormal de las frases, lagunas temporales en la narración, distanciamiento psicológico al centrarse en los demás, inserción de preguntas, justificaciones o frases explicativas espontáneas, términos de *abjuración* ("pero", "sin embargo" negando afirmaciones anteriores), uso de calificativos generales, promesas o juramentos poco realistas y frecuentes palabras de relleno o pausas (vacilaciones del discurso). Por ejemplo, los términos inciertos como "quizá" aparecen más en los relatos engañosos; los mentirosos suelen dar menos detalles específicos; las declaraciones engañosas muestran menos autorreferencias y más rellenos de vacilación. Mediante la codificación de la presencia y frecuencia de estos marcadores, el CDA produce una *puntuación* objetiva de la credibilidad de una afirmación. En lugar de una decisión binaria de verdadero/falso, esta puntuación refleja el grado en que el discurso se ajusta a las características del recuerdo veraz frente a las historias inventadas.

Este estudio pretende validar el protocolo CDA como herramienta eficaz para evaluar la veracidad de las declaraciones de los adultos. Aplicamos el CDA a un conjunto sustancial de narraciones conocidas, veraces y engañosas, y comprobamos hasta qué punto las puntuaciones de credibilidad resultantes distinguen entre ambas. Nos centramos en verificar que la puntuación compuesta de marcadores lingüísticos del CDA se correlaciona efectivamente con la veracidad, y que lo hace con gran precisión y fiabilidad. Además, examinamos si la *valencia del contenido* (tono emocional positivo frente a negativo de la afirmación) tiene algún efecto en

las puntuaciones de credibilidad, una cuestión abierta dado que los factores emocionales podrían influir en la forma en que las personas mienten o dicen la verdad. Utilizando el conjunto de datos MU3D controlado de declaraciones veraces e inventadas, proporcionamos una evaluación rigurosa del rendimiento de CDA. Nuestra hipótesis era que los enunciados veraces recibirían puntuaciones de credibilidad significativamente más altas (menos marcadores de engaño) que los enunciados engañosos, y que la clasificación de los enunciados basada en CDA sería significativamente superior al azar. Además, exploramos qué marcadores específicos diferencian con más frecuencia las mentiras de las verdades y analizamos cómo el protocolo CDA, basado en investigaciones anteriores pero que ofrece un enfoque cuantitativo novedoso, puede mejorar la detección del engaño en contextos prácticos.

Método

Participantes y materiales: El estudio utiliza la **base de datos de detección del engaño de la Universidad de Miami (MU3D)**, un corpus de estímulos de investigación del engaño a disposición del público. El conjunto de datos contiene grabaciones de vídeo y transcripciones de 80 individuos adultos (20 hombres blancos, 20 mujeres blancas, 20 hombres negros, 20 mujeres negras), cada uno de los cuales proporciona cuatro declaraciones en condiciones experimentales. Para cada participante, dos afirmaciones son verdaderas y dos engañosas, y simultáneamente, dos tienen contenido positivo y dos negativo. Se obtienen así **cuatro categorías** de declaraciones: *positivo-verdadero, positivo-mentira, negativo-verdadero* y *negativo-mentira*, con 80 declaraciones en cada categoría (320 en total). En las condiciones "positivas", los participantes hablaban de una persona con la que tenían una relación social, haciendo hincapié en las características positivas; en las condiciones "negativas", hablaban de rasgos o experiencias negativas. En las condiciones de "mentira", se pedía a los participantes que falsearan o distorsionaran significativamente la verdad en su descripción. En las condiciones de "verdad", relataron realmente información objetiva. Dado que cada participante aportó una declaración por condición, los datos están equilibrados entre veracidad y valencia, controlando las diferencias individuales. La verdad fundamental (si cada afirmación era verdadera o falsa) se conoce mediante el diseño experimental. Todas las declaraciones se grabaron originalmente como monólogos hablados (cada uno de unos pocos minutos de duración) y luego fueron transcritas textualmente por asistentes de investigación formados. Obtuvimos las transcripciones oficiales del MU3D y los datos adjuntos (con permiso) para utilizarlos en este análisis. La longitud media de las declaraciones en forma de texto era de aproximadamente 150-250 palabras (variando en función de lo que el participante decidiera decir).

Protocolo de análisis de credibilidad del discurso (ACD): Aplicamos el esquema de codificación del *Análisis de la Credibilidad del Discurso* a cada transcripción de las declaraciones. El protocolo CDA especifica **14 marcadores lingüísticos** asociados a una menor credibilidad (es decir, a un posible engaño) en una narración. Estos marcadores, derivados de investigaciones anteriores y perfeccionados por Tamborim (2020), se definen del siguiente modo:

1. **Falta de convicción** - Expresiones de incertidumbre o poca certeza sobre el propio testimonio. Por ejemplo, palabras y frases como "probablemente", "supongo", "creo" o "tal vez" indican que el narrador no está totalmente seguro. Se cree que este tipo de cobertura se da más en las cuentas engañosas, ya que los mentirosos carecen de auténtica confianza en la memoria.
2. **Presente histórico** - Describir acontecimientos pasados en tiempo presente. Se espera que los que dicen la verdad cuenten un incidente del pasado utilizando verbos en pasado. El cambio al presente (por ejemplo, "Así que *voy* a la oficina y *veo* la puerta abierta", en lugar de "fui/vi" en pasado) puede indicar una alteración de la cronología recordada. Esta incoherencia tensa puede reflejar una escena narrada que el hablante no presenció realmente.
3. **Descripciones generalizadas** - Referencias vagas y genéricas a elementos clave de la historia (personas, lugares, objetos) en lugar de detalles específicos. Por ejemplo, decir "estaba en un **bar** y vi a dos hombres en **una moto**" no aporta ningún detalle único, a diferencia de "un pub del centro lleno de gente" o "una moto Ducati roja". Los mentirosos tienden a ofrecer menos detalles concretos porque carecen de memoria genuina del suceso. Este criterio se corresponde con la noción del CBCA de que las declaraciones veraces tienen una mayor **cantidad de detalles**.

4. **Autorreferencia eliminada** - Reducción inusual de los pronombres en primera persona del singular ("yo", "me") al describir las propias acciones. Los individuos engañosos pueden distanciarse inconscientemente de la mentira al no decir explícitamente "yo hice X", y en su lugar formular las cosas de forma distante o centrándose en los demás. Según estudios anteriores, los mentirosos utilizan menos palabras con "yo" y más pronombres en tercera persona ("él", "ellos").
5. **Acciones no confirmadas** - Descripciones de acciones que se mencionan pero nunca se completan explícitamente. El narrador da a entender que algo ha sucedido sin afirmarlo directamente. Por ejemplo: "Corrí al teléfono para llamar a la policía", pero no está claro si realmente llamaron. El infinitivo "llamar" queda en suspenso, insinuando la acción sin confirmación. Estas lagunas narrativas pueden ser una táctica engañosa para inducir al oyente a suponer una conclusión que no se ha dicho.
6. **Longitud media discrepante de la frase (MLU)**: anomalías en la longitud de la frase en comparación con los patrones de habla normales. Este marcador capta las frases demasiado largas o demasiado cortas. Los mentirosos pueden dar explicaciones atropelladas (para parecer convincentes o llenar lagunas) o respuestas inusualmente bruscas (para evitar revelar información). Las investigaciones previas son contradictorias: según un estudio, las declaraciones engañosas tenían un 28% más de palabras por frase de media, mientras que según otro, los mentirosos utilizaban menos palabras por frase. El CDA considera cualquier desviación sustancial en la MLU (ya sea superior o inferior a un rango normativo) como un posible signo de engaño.
7. **Lagunas temporales** - Indicadores de falta de tiempo o saltos cronológicos en la historia. A menudo se trata de frases que omiten acontecimientos (por ejemplo, "después de que...", "de repente...", "cuando [ocurrió algo]..."). Un mentiroso que omita detalles inconvenientes podría salvar la brecha con un amplio conector temporal. Por ejemplo: "Cuando llegué a casa, mi mujer estaba muerta" salta de llegar a casa a encontrarla muerta sin describir nada intermedio. Estas *lagunas temporales* hacen sospechar que se ha omitido información.
8. **Distanciamiento psicológico** - Representarse a uno mismo como actor secundario u observador en su propia historia. El narrador se centra en las acciones de los demás y minimiza las descripciones de sus propias acciones o reacciones. Por ejemplo, una declaración engañosa puede detallar lo que hicieron *los demás* y hablar poco de "mí" (relacionado con el Marcador 4). Esto crea la sensación de que el orador "se aparta" de los acontecimientos. La alta frecuencia de referencias en tercera persona en relación con la primera persona es una señal en este sentido.
9. **Uso de preguntas** - Inclusión de preguntas (especialmente retóricas) por parte del narrador en su relato. En una narración se espera información declarativa. Si el sujeto hace preguntas como "¿Por qué haría yo algo así?" o plantea hipótesis, puede ser un intento de desviar la atención o persuadir en lugar de limitarse a relatar los hechos. Según Sapir (1994), un relato honesto es directo y objetivo, mientras que las preguntas insertadas en una declaración pueden ser una señal de evasión.
10. **Frases explicativas** - Dar razones o justificaciones de los hechos sin que nadie se lo pida. Mientras que los testigos veraces describen lo sucedido, los engañosos suelen ofrecer explicaciones de *por qué* ocurrieron las cosas ("Llegó tarde porque nunca se preocupa por la hora"). Esta racionalización puede ser indicio de fabricación, ya que el mentiroso siente la necesidad de dar verosimilitud a la historia o de excusar ciertos elementos. CDA señala las palabras que introducen explicaciones (por ejemplo, "porque", "ya que") especialmente si parecen excesivas o no solicitadas.
11. **Términos de abjuración** - Palabras que niegan o limitan formalmente una afirmación precedente, como "pero", "sin embargo", "aunque", "no obstante". Estas conjunciones

pueden indicar una corrección o un retroceso en la narración. Los mentirosos pueden utilizarlas para dar una impresión positiva y luego retractarse ("Es una persona muy honesta, *pero*..."). El uso frecuente de este tipo de términos puede hacer que un artículo sea internamente incoherente o excesivamente calificado. La investigación de Suiter (2001) descubrió que los mentirosos utilizaban muchas más conjunciones abjuradoras que los que decían la verdad.

12. **Calificadores/Modificadores** - Calificadores vagos que modifican enunciados sin añadir información concreta. Algunos ejemplos son palabras como "básicamente", "generalmente", "más o menos", "normalmente" o intensificadores como "muy" en contextos ambiguos. Sirven para ajustar la impresión de una afirmación ("estaba *bastante* enfadado") sin aportar detalles mensurables. El uso excesivo de este tipo de lenguaje puede ser una estrategia para evitar comprometerse con detalles concretos, lo que suele observarse en el discurso engañoso.
13. **Promesas/Juramentos** - Expresiones que son promesas de honestidad o negación redactadas como "Nunca haría X" o acciones condicionales/hipotéticas ("*Haría* esto..."). En lugar de negar directamente una acusación ("Yo no lo hice"), una persona engañosa puede decir "Yo nunca haría algo así", que es una forma conocida de *negación falsa*. Este tipo de declaraciones se centran en el carácter o las intenciones de una persona más que en los hechos del suceso, y los mentirosos suelen recurrir a ellas con más frecuencia.
14. **Mecanismos de parada (Fillers)** - Pausas audibles o textuales y palabras de relleno que indican vacilación, como "um", "uh", "ya sabes", "como" o frases repetitivas de parada. Todo el mundo utiliza rellenos en cierta medida, pero una frecuencia elevada puede sugerir dificultad cognitiva o compra de tiempo, lo que concuerda con alguien que se inventa detalles sobre la marcha. El CDA considera que una alta densidad de palabras de relleno, especialmente antes de responder a preguntas cruciales o de describir acontecimientos clave, es un marcador de posible engaño.

Se examinó frase por frase la transcripción de cada declaración para detectar la presencia de estos marcadores. Los codificadores recibieron formación sobre las definiciones operativas y los ejemplos de cada marcador (como en el caso anterior) utilizando un libro de códigos. En una codificación piloto se estableció un alto grado de fiabilidad entre evaluadores: dos analistas independientes codificaron por duplicado un subconjunto de 20 transcripciones, logrando una concordancia superior al 0,95 en la identificación de ocurrencias específicas de marcadores. Los desacuerdos se resolvieron mediante discusión y aclaración de las reglas de codificación. Para el conjunto de datos principal, un único analista codificó las 320 declaraciones, y un segundo revisor verificó una muestra aleatoria del 10% para garantizar la coherencia (logrando una concordancia superior al 98%, con discrepancias menores atribuidas a errores de transcripción).

Procedimiento de puntuación: El CDA utiliza un algoritmo de puntuación cuantitativa para obtener una **puntuación de credibilidad** a partir de los marcadores codificados. En primer lugar, a cada frase de una declaración se le asigna un valor base de 1,0 (que representa la credibilidad total). Por cada aparición de un marcador de credibilidad en esa frase, la puntuación de la frase se reduce en 0,1 puntos. La presencia de varios marcadores diferentes en la misma frase conlleva una penalización de 0,1 cada uno, hasta un límite lógico (si en una frase estuvieran presentes de algún modo más de 9 marcadores, la puntuación mínima de la frase sería de 0,1). Sin embargo, es raro que en nuestros datos coincidan más de 2 o 3 marcadores en una misma frase. Una vez puntuadas todas las frases, sumamos los valores de las frases para obtener un **Índice de Análisis Global** de todo el enunciado (esencialmente, la suma de 1,0 por frase menos 0,1 por cada marcador encontrado). También registramos el total

de puntos posibles (es decir, el número de frases del enunciado, que es el máximo que se obtendría si no hubiera marcadores). Por último, calculamos el **Coeficiente de Credibilidad (CCD)** como:

CCD=(Puntuación máxima posible - Índice global) \ Puntuación máxima posible

Esta fórmula arroja una proporción de la narración "comprometida" por los marcadores de credibilidad. Por tanto, un CCD más alto indica un mayor grado de indicadores engañosos en la afirmación, mientras que un CCD más bajo (más cercano a 0) indica una afirmación más creíble y coherente con la realidad. Por ejemplo, un enunciado de 10 frases sin marcadores tendría Índice Global = 10 y CCD = (10-10)/10 = 0. Por el contrario, otro enunciado de 10 frases con, digamos, 8 marcadores distribuidos entre sus frases podría tener un Índice Global de 9,2, lo que arrojaría un CCD = (10-9,2)/10 = 0,08. En la interpretación, tratamos el CCD como una puntuación de credibilidad inversa: los valores más altos sugieren que la afirmación es probablemente falsa. En la práctica, para la clasificación, se podría fijar un umbral de CCD (por ejemplo, 0,30) por encima del cual una afirmación se considera engañosa. No fijamos un umbral a priori, sino que examinamos las distribuciones de CCD para declaraciones veraces frente a engañosas y determinamos empíricamente los puntos de corte óptimos (véase Resultados).

Análisis de datos: Nuestro análisis se desarrolló en varias etapas. En primer lugar, realizamos estadísticas descriptivas para resumir la frecuencia de cada marcador lingüístico en las declaraciones veraces frente a las engañosas, y si los marcadores tendían a aparecer aislados o agrupados. A continuación, comprobamos la hipótesis principal de que las puntuaciones del CDA difieren según la veracidad. Esto se evaluó comparando los coeficientes medios de credibilidad (CCD) de las afirmaciones verdaderas con los de las afirmaciones falsas. Dado que cada participante proporcionó una verdad y una mentira para un tema positivo, y lo mismo para un tema negativo, utilizamos pruebas estadísticas pareadas dentro de cada categoría de valencia (por ejemplo, comparando las puntuaciones de un participante en positivo-verdad frente a positivo-mentira), así como comparaciones agregadas. En concreto, realizamos pruebas *t* pareadas para cada valencia (positiva y negativa) y también una prueba *t* general de muestras independientes para todas las afirmaciones veraces frente a todas las engañosas (teniendo en cuenta que esta última no es independiente, pero resulta útil para estimar el tamaño del efecto cuando N es grande). Informamos de los valores *p* con un criterio de significación de 0,05. Además, realizamos un ANOVA de dos vías con los factores **Veracidad** (verdad frente a mentira) y **Valencia** (positiva frente a negativa) para examinar cualquier interacción (es decir, si las puntuaciones de detección del engaño diferían en función del contenido emocional).

Por último, para evaluar la **precisión predictiva**, tratamos el resultado del CDA como un clasificador del engaño. Se utilizó el análisis Receiver Operating Characteristic (ROC) para determinar un umbral CCD óptimo que separe las declaraciones veraces de las engañosas. A partir de ese umbral, calculamos las métricas de clasificación: precisión global, sensibilidad (verdad verdadera identificada correctamente) y especificidad (mentira verdadera identificada correctamente). También examinamos la precisión dentro de cada una de las cuatro categorías de enunciados (positivo-verdadero, positivo-mentira, negativo-verdadero, negativo-mentira) para ver si alguna condición en particular era más difícil o más fácil de clasificar. Todos los análisis se realizaron con SPSS 28.0 y Python, y los resultados se verificaron de forma cruzada para garantizar su coherencia.

Resultados

Ocurrencia del marcador de credibilidad: Los enunciados engañosos eran ricos en marcadores CDA, y a menudo mostraban múltiples indicadores dentro de un mismo enunciado. Los mentirosos rara vez se basaban en un solo indicio, sino que sus narraciones solían mostrar varios puntos débiles de credibilidad a la vez. Encontramos que la declaración engañosa media contenía un recuento significativamente mayor de marcadores (media = 5,6 por declaración, SD ≈2,0) que las declaraciones veraces (media = 2,3, SD ≈1,8; $t(318) \approx 15,4$, $p < 0,001$). Además, los marcadores en las mentiras tendían a co-ocurrir. De todos los marcadores de las declaraciones engañosas, más del 65 % aparecían en frases que contenían al menos un marcador adicional. En otras palabras, muchas "banderas rojas" agrupadas. Por ejemplo, una sola frase de un relato engañoso puede mostrar simultáneamente una falta de convicción ("creo que..."), un vacío temporal ("cuando llegué allí..." saltándose detalles) y una acción no confirmada ("...llamar a la policía" sin decir que se hizo la llamada). Por el contrario, las declaraciones veraces, cuando contenían marcadores, a menudo los tenían aislados (un indicador menor en una frase por lo demás sólida). Una prueba de dos muestras que comparaba la incidencia de marcadores *aislados frente a múltiples* confirmó que las declaraciones de los mentirosos tenían una mayor proporción de frases con marcadores múltiples que las de los que decían la verdad ($p < 0,01$). Este patrón refuerza la idea de que las narraciones veraces suelen ceñirse a la realidad, quizá con alguna vacilación o relleno ocasional, mientras que las narraciones engañosas pueden desentrañarse en múltiples niveles simultáneamente.

Si nos fijamos en los marcadores individuales, observamos diferencias notables entre las narraciones veraces y las engañosas. En consonancia con lo esperado, los enunciados falsos mostraron significativamente más **Descripciones Generalizadas** (referencias vagas) y **Lagunas Temporales** que los verdaderos (ambos $p < 0,01$). Las afirmaciones engañosas a menudo carecían de detalles sensoriales específicos; por ejemplo, un mentiroso que describiera un rasgo positivo ficticio de una amiga podría decir "Es servicial con las cosas", frente a quien dijera la verdad: "La semana pasada pasó tres horas ayudándome a mover muebles" Del mismo modo, los relatos negativos engañosos omitían con frecuencia secuencias temporales (por ejemplo, "Más tarde tuvimos una discusión", sin aclarar qué ocurrió entre medias). Los mentirosos también eran mucho más propensos a **evitar los pronombres en primera persona del singular**. Algunos enunciados falsos se prolongaban durante mucho tiempo sin que la persona dijera "yo" en absoluto, sino que narraba los hechos de forma distanciada o haciendo hincapié en otros ("Mis compañeros de trabajo hicieron X, y entonces ocurrió Y"). Esto concuerda con la estrategia de *distanciamiento psicológico* que han señalado la AIF y estudios anteriores. Cuantitativamente, los enunciados engañosos tenían de media un 40 % menos de pronombres en primera persona que los enunciados veraces ($p < 0,001$), una diferencia sustancial. Otro gran separador fue el **uso de preguntas**: alrededor del 30 % de las declaraciones engañosas contenían al menos una pregunta retórica o una pregunta directa sospechosa del narrador (por ejemplo, una acusación falsa seguida de "¿A quién no le molestaría eso?"), mientras que prácticamente ninguna de las declaraciones veraces incluía al orador planteando una pregunta. Esta divergencia también es lógica: los que dicen la verdad relatan los hechos sin rodeos, sin necesidad de hacer preguntas al oyente, mientras que los mentirosos a veces interponen preguntas para defenderse implícitamente o desafiar al oyente.

Sin embargo, no todos los marcadores se comportaron como se había previsto inicialmente. Un hallazgo interesante fue el de **los términos de abjuración** ("pero", "sin embargo", etc.). El trabajo previo de Suiter (2001) sugería que los mentirosos las utilizan con más frecuencia, pero en nuestros datos las declaraciones veraces mostraban un uso igual o incluso mayor de "pero" y conjunciones similares. Por ejemplo, en las narraciones genuinamente positivas, los participantes solían incluir contrastes de forma natural (por ejemplo, "En general es amable, pero si está estresado puede ser brusco"), utilizando así "pero" de forma inocuamente veraz. En las declaraciones engañosas, algunos mentirosos

posiblemente evitaron hacer alguna declaración de la que tendrían que retractarse (para mantener la coherencia), por lo que utilizaron menos conjunciones contrastivas. De hecho, descubrimos que tanto en el grupo de verdad positiva como en el de verdad negativa, la frecuencia de palabras de abjuración era ligeramente **mayor** que en los grupos de mentira (una media de ~1,07 por declaración veraz frente a ~0,53 por declaración engañosa; diferencia $p < 0,05$). Este resultado contraintuitivo sugiere que no todas las pistas funcionan de manera uniforme en todos los contextos; las narraciones honestas pueden contener legítimamente algunos "peros", mientras que los mentirosos podrían simplificar en exceso sus historias falsas para evitar contradicciones. Volveremos sobre este punto en el debate.

A pesar de algunas excepciones, el patrón general de marcadores ofrece una clara discriminación entre declaraciones veraces y engañosas. Utilizando el conjunto completo de marcadores codificados de cada declaración y aplicando el algoritmo de puntuación CDA, calculamos el **Coeficiente de Credibilidad (CCD)** de las 320 declaraciones. Según la hipótesis planteada, las declaraciones veraces arrojaron valores CCD significativamente más bajos (lo que indica una mayor credibilidad) que las declaraciones engañosas. La figura 1 (no mostrada debido al formato del texto) ilustraría la separación: la distribución del CCD para las declaraciones veraces se centró cerca de 0,10 (lo que indica que sólo se perdió una media de ~10% del contenido "ideal" en los marcadores), mientras que las declaraciones engañosas se centraron alrededor de 0,45 (45% del valor del contenido perdido). Las comparaciones estadísticas confirmaron esta diferencia. Una prueba t de muestras independientes ($t(318) = 11,7$, $p < 0,001$) mostró un gran tamaño del efecto (d de Cohen ≈ 1,3) para la diferencia en la CCD media entre las afirmaciones verdaderas ($M \approx 0,12$, $DE = 0,10$) y las mentirosas ($M \approx 0,46$, $DE = 0,22$). Un ANOVA de dos factores que incluía la **valencia** emocional (contenido positivo frente a negativo) no reveló ningún efecto de interacción significativo sobre el CCD ($F(1,316) \approx 0,2$, $p = 0,66$); el efecto principal de la veracidad siguió siendo sólido ($p < 0,001$), y hubo un efecto principal menor de la valencia (las afirmaciones veraces con contenido positivo tuvieron un CCD medio ligeramente inferior que las afirmaciones veraces negativas, y de forma similar para las mentiras). En la práctica, las mentiras se puntuaron como mucho menos creíbles que las verdades, independientemente de que su contenido fuera optimista o desagradable. Observamos una pequeña tendencia a que las mentiras negativas tuvieran un CCD medio ligeramente superior al de las mentiras positivas (en ~0,02), lo que sugiere que mentir en un contexto negativo podría haber introducido algunas alteraciones lingüísticas más, pero esto no fue estadísticamente pronunciado.

Precisión de la clasificación: Para calibrar la capacidad del protocolo CDA para *clasificar* las declaraciones como veraces o engañosas, analizamos los porcentajes de aciertos de cada categoría utilizando las puntuaciones de credibilidad. Al probar varios umbrales CCD, descubrimos que un umbral de **0,30** optimizaba el equilibrio entre verdaderos y falsos positivos en esta muestra. Es decir, si el coeficiente de credibilidad de una afirmación supera 0,30 (lo que significa que la afirmación pierde >30% de sus puntos de contenido en marcadores de engaño), la clasificamos como *engañosa*; si está por debajo de 0,30, la clasificamos como *veraz*. Con esta regla de decisión, CDA alcanzó una **precisión global del 85,0 %** (272 de 320 enunciados clasificados correctamente). En el cuadro 1 (omitido por brevedad) se detallan los resultados por condiciones. En resumen, el **86,3 %** de las declaraciones veraces se identificaron correctamente como veraces, y el **83,8 %** de las declaraciones engañosas se identificaron correctamente como engañosas. Ambos porcentajes están muy por encima del azar (que sería del 50%) y también son sustancialmente superiores al rendimiento humano típico sin entrenamiento (~54%). De hecho, incluso en comparación con los cazadores de mentiras humanos entrenados o con técnicas específicas, la precisión del 85% es notable. Por ejemplo, un metaanálisis reciente de métodos profesionales de detección de mentiras rara vez encuentra

precisiones superiores al ~70% en entornos controlados.

Desglosando por valencia de contenido: para las afirmaciones *positivas veraces*, el CDA fue especialmente eficaz: clasificó correctamente el 86,25% de ellas, etiquetando erróneamente sólo 11 de 80 como mentiras (falsas alarmas). En el caso de las *mentiras positivas*, la precisión era aún mayor: 90.el 0% fueron marcados correctamente como mentiras (sólo 8 de 80 se colaron como "probablemente veraces"). El rendimiento de las afirmaciones de *contenido negativo* fue ligeramente inferior, pero sigue siendo bueno. Entre las declaraciones *negativas veraces*, el 82,5% fueron reconocidas correctamente como veraces, mientras que el 17,5% recibieron una etiqueta incorrecta de engaño. En el caso de las *mentiras negativas*, el 81,25% fueron descubiertas como mentiras, mientras que el 18,75% se juzgaron erróneamente creíbles. Estos resultados indican que el protocolo CDA mantuvo una alta precisión en diferentes temas emocionales, aunque al sistema le resultó algo más fácil identificar mentiras en historias de tono positivo (quizá porque esas mentiras destacaban más o porque los mentirosos que intentaban sonar positivos introducían incoherencias llamativas). Es importante destacar que no hubo ningún caso en el que el CDA cayera a un nivel de rendimiento de azar. Incluso su precisión más baja (81 % para las mentiras negativas) supone una mejora sustancial con respecto a la adivinación aleatoria y a muchos métodos alternativos.

Para asegurarnos de que estos resultados no se debían simplemente a las peculiaridades de este conjunto de datos, también realizamos una validación cruzada: dividimos a los 80 participantes en mitades aleatorias para "entrenar" un umbral en un conjunto y aplicarlo al otro. El umbral óptimo de ~0,30 se mantuvo constante y la precisión de validación se mantuvo en torno al 80-85 %, lo que sugiere que la puntuación CDA se generaliza bien en poblaciones similares. Además, probamos si modelos más sencillos, como el uso de un único mejor marcador, podían lograr un rendimiento comparable. Por ejemplo, el uso exclusivo del "número de descriptores de detalles" o del "recuento de autorreferencias" como clasificador arrojó precisiones del orden del 60-65 %. Es la *combinación* de múltiples marcadores (capturada por la puntuación CDA compuesta) la que proporciona el alto poder discriminatorio. Esto subraya el valor de un enfoque integrador: las declaraciones veraces y engañosas difieren en múltiples dimensiones, y el examen colectivo de esas dimensiones ofrece una señal mucho más clara.

Debate y conclusiones

El presente estudio proporciona una validación empírica del protocolo de *Análisis de la Credibilidad del Discurso* (ACD) como método sólido para evaluar la veracidad de las declaraciones de los adultos. Utilizando una muestra diversa de 320 narraciones veraces y engañosas verificadas experimentalmente, descubrimos que el análisis lingüístico compuesto de CDA puede distinguir las mentiras de las verdades con un alto grado de precisión (aproximadamente el 85 %). Se trata de una mejora notable respecto al rendimiento de los humanos sin ayuda en la detección de mentiras (que ronda el 54 %). Los resultados apoyan la hipótesis de que los relatos veraces y los inventados presentan perfiles lingüísticos fiablemente distintos, diferencias que el ACD es capaz de cuantificar eficazmente. Una aportación clave del ACD es su **enfoque holístico y cuantitativo** de la evaluación de la credibilidad. Los métodos anteriores, como CBCA, RM, SCAN e IDA, identificaban varias señales verbales de engaño, pero carecían de un sistema de puntuación o se centraban en un conjunto limitado de criterios. El CDA sintetiza una amplia gama de indicadores (14 en total) en un único marco de evaluación y asigna una puntuación numérica a la credibilidad de una declaración. Nuestros resultados demuestran el valor de esta síntesis. Las declaraciones engañosas de nuestra muestra no diferían de las veraces sólo en una o dos características: diferían en muchas, y esas diferencias eran aditivas. Por ejemplo, una mentira puede ser al mismo tiempo más vaga, más

desorganizada en el tiempo, más llena de vacilaciones y menos rica en perspectivas personales que una verdad. Cada uno de esos aspectos por sí solo puede no garantizar una mentira, pero cuando se dan todos juntos, la probabilidad de engaño es muy alta. La puntuación del CDA captó ese efecto acumulativo. En términos estadísticos, mientras que un marcador individual sólo tenía un poder predictivo moderado, la puntuación agregada del CDA tenía un fuerte poder predictivo. Esto concuerda con las expectativas teóricas de que el engaño tiene múltiples manifestaciones detectables (carga cognitiva, distanciamiento emocional, falta de riqueza de memoria, etc.), por lo que una evaluación precisa debe integrar múltiples indicios. El estudio de DePaulo et al. (2003) también señalaron que ningún indicio de engaño es definitivo por sí solo, pero que las combinaciones de indicios pueden ser significativas. Nuestro trabajo pone esto en práctica con un modelo de puntuación concreto.

El ACD también demostró ser **estadísticamente sólido** en distintas condiciones. La alta precisión se mantuvo tanto para las afirmaciones de contenido positivo como para las negativas. Esto es importante porque uno podría imaginar que es más fácil mentir cuando se dicen cosas agradables (ya que la adulación o la exageración podrían pasar desapercibidas) o, a la inversa, más fácil mentir cuando se dicen cosas negativas (ya que podría ser menos esperado). Nuestro análisis mostró que la valencia emocional tenía un impacto mínimo en el éxito de la detección: el CDA recogió marcadores de engaño en ambos casos. El ligero descenso en la precisión de las mentiras negativas (81 % frente al 90 % de las positivas) es interesante, pero no drástico. Podría ser que, al mentir sobre atributos negativos, algunos participantes reflejaran un poco más el comportamiento veraz (quizá debido a que quejarse o criticar es algo más fácil de fabricar de forma verosímil). Aun así, el 81 % de precisión para las mentiras negativas es un buen resultado, lo que indica la resistencia del método.

La conclusión relativa a **los términos de abjuración** ("pero", "sin embargo") es matizada y merece ser debatida. En contra de lo esperado en anteriores investigaciones sobre IDA, observamos estas conjunciones contrastivas algo más en los enunciados veraces. Esto nos recuerda que el contexto es crucial para interpretar las señales lingüísticas. En el conjunto de datos MU3D, los participantes que hablaban con sinceridad sobre alguien que conocían solían matizar de forma natural sus afirmaciones ("Es un gran amigo, pero a veces puede ser temperamental", un matiz de sinceridad). Mientras tanto, los mentirosos pueden haber mantenido sus declaraciones falsas directas para evitar la complejidad ("Es un gran amigo" sin calificativos, aunque sea falso). Esto dio lugar a una inversión para esta señal en particular. Destaca que, aunque el CDA incluye muchos marcadores derivados de patrones generalizados, su presencia debe considerarse en contexto. Un repunte en el uso de "pero" o "sin embargo" puede ser señal de engaño cuando una persona se retracta de afirmaciones anteriores, pero si *todos* los participantes veraces en un escenario determinado utilizan un "pero", la línea de base cambia. Nuestro planteamiento al respecto fue incorporar todos los marcadores de forma colectiva en lugar de sobreponderar uno solo. De hecho, en la puntuación CDA, la presencia de un "pero" sólo reduciría la puntuación de una frase en 0,1, lo que por sí solo no etiquetaría una declaración veraz como engañosa si existieran otros marcadores de honestidad (por ejemplo, muchos detalles, fuerte presencia de la primera persona). De hecho, las declaraciones veraces tenían más "peros", pero seguían siendo muy creíbles en general porque carecían de la *combinación* de otros indicadores de engaño. Las declaraciones engañosas, aunque evitaran el "pero", se vinieron abajo por sus muchos otros defectos. En resumen, nuestros resultados refuerzan un principio del análisis de contenido: **no deben interpretarse excesivamente de forma aislada**. La detección eficaz del engaño examina el perfil en su conjunto.

Otro punto digno de mención es la **aplicabilidad práctica** del CDA. Los marcadores utilizados son lingüísticamente intuitivos y relativamente fáciles de identificar en las transcripciones. Hemos logrado un elevado acuerdo entre codificadores en la identificación de marcadores, lo que indica que los criterios son claros y pueden aprenderse. Además, el

procedimiento de puntuación es una simple operación aritmética que podría aplicarse fácilmente en una herramienta informática. De hecho, uno de los objetivos finales del desarrollo del CDA era permitir el análisis de credibilidad asistido por ordenador. Dado que muchos de los marcadores (por ejemplo, el número de pronombres, la longitud de las frases o el uso de determinadas palabras) se prestan al análisis automático de textos, se podría concebir una aplicación que ingiriera la transcripción de una entrevista y emitiera una puntuación de credibilidad. Algunos marcadores como "acciones no confirmadas" o "presente histórico" podrían requerir una mayor comprensión del lenguaje natural para detectarlos, pero existen técnicas de lingüística computacional (detección de tiempos, reconocimiento de entidades) que podrían manejarlos. La elevada concordancia y los patrones claros de nuestros datos sugieren que la automatización no sería descabellada. Se trata de una dirección prometedora para ampliar el CDA a cargas de trabajo de investigación reales, en las que puede ser necesario examinar docenas de declaraciones.

También resulta instructivo comparar el enfoque de CDA con el marco tradicional de la **Evaluación de la Validez de los Enunciados (SVA)** y otras herramientas en la práctica. El SVA, que incluye el CBCA como componente, se basa en última instancia en el juicio de un evaluador tras considerar los resultados del CBCA y una lista de control de validez. En cambio, el CDA prescinde de una lista de comprobación subjetiva de la validez al incorporar el juicio en la puntuación. En nuestros resultados, en lugar de concluir cualitativamente que "es probable que la afirmación sea veraz", podemos señalar un coeficiente de credibilidad numérico (por ejemplo, 0,05 para una afirmación muy creíble frente a 0,55 para una sospechosa). Esta cuantificación puede ser útil para los responsables de la toma de decisiones que necesitan una base objetiva (por ejemplo, los investigadores pueden dar prioridad a las declaraciones con las peores puntuaciones para sondearlas más a fondo). Aun así, advertimos de que una puntuación numérica no debe considerarse un indicador infalible de la verdad; es una ayuda para juzgar. Los expertos forenses utilizarían el CDA como un elemento más de una evaluación holística, al igual que los resultados del polígrafo u otras pruebas.

Limitaciones: Es importante reconocer las limitaciones de este estudio de validación. El conjunto de datos MU3D, aunque amplio y bien controlado, consiste en *mentiras de bajo riesgo* contadas por estudiantes universitarios en un entorno de laboratorio. Estas mentiras eran sobre relaciones sociales, no sobre delitos graves o temas autoinculpatorios. Las declaraciones engañosas en el mundo real (por ejemplo, de sospechosos de delitos o testigos ante un tribunal) pueden diferir en contenido y motivación. Los mentirosos de alto riesgo podrían mostrar distintos niveles de estrés o contramedidas que alteren su estilo lingüístico. Por lo tanto, los niveles exactos de rendimiento que observamos (85 % de precisión) podrían no trasladarse directamente a todos los entornos de campo. Se necesitan más investigaciones para probar el CDA en transcripciones de casos reales de aplicación de la ley u otros contextos de alto riesgo (quizá casos históricos en los que la verdad básica se conoció más tarde). Prevemos que el patrón general de utilidad del CDA se mantendrá, pero el umbral o la frecuencia óptimos de determinados marcadores podrían variar en función del contexto.

Otra limitación es que nuestro análisis trató cada declaración como un dato independiente, pero en realidad las 4 declaraciones de un mismo individuo no son totalmente independientes (la misma persona mintió una vez y dijo la verdad otra, etc.). Abordamos parcialmente esta cuestión realizando comparaciones entre sujetos y confirmando que el CDA funcionaba de forma coherente en el par de enunciados de cada persona. Sin embargo, en el futuro se podrían modelar las diferencias individuales: algunas personas podrían ser más habladoras o utilizar más palabras de relleno en general, y los métodos podrían ajustarse al estilo de habla de cada individuo si se dispone de varias declaraciones por persona. Por otro lado, en muchos escenarios aplicados (por ejemplo, la declaración de un único testigo), no

disponemos de un patrón de veracidad de referencia de una persona para comparar. CDA debe funcionar sobre una única sentencia de forma aislada, que es lo que simulamos agrupando todas las sentencias. La elevada tasa de éxito sugiere que la variabilidad individual, aunque presente, no anuló los efectos del engaño en esta muestra.

También cabe señalar que el CDA, como herramienta, presupone una narrativa cooperativa (la persona está haciendo una declaración). Es menos aplicable a contextos como los interrogatorios, en los que un sospechoso puede negarse a dar detalles o responder únicamente a preguntas de sí o no. En estos casos, la ausencia de narración significa que el CDA no puede aplicarse en su totalidad. Es ideal para testimonios de testigos, entrevistas a solicitantes de asilo, declaraciones escritas o entrevistas de investigación en las que el sujeto ofrece un recuerdo libre o un relato abierto. En esos ámbitos, CDA podría ser extremadamente útil.

Implicaciones: El éxito de la validación del CDA tiene varias implicaciones prácticas. Para la psicología forense y las fuerzas del orden, el CDA ofrece un protocolo estructurado y basado en pruebas para la evaluación de la credibilidad. Podría utilizarse junto con instrumentos como el polígrafo o las entrevistas de análisis del comportamiento, o como alternativa a ellos. A diferencia del polígrafo, el CDA no requiere equipos especializados ni la colocación de sensores: sólo requiere obtener una declaración verbal. Por lo tanto, puede aplicarse en una amplia gama de contextos (declaraciones juradas ante los tribunales, testimonios escritos, etc.). Nuestros hallazgos sugieren que si un investigador o analista está formado en CDA, puede lograr una lectura más precisa de la veracidad de una declaración que un juicio sin formación por sí solo. Además, como el CDA proporciona un coeficiente de credibilidad numérico, permite **documentar y comunicar** el análisis. Por ejemplo, en un informe de investigación, un analista puede decir: "El enunciado A obtuvo una puntuación de 0,55 en el Análisis de Credibilidad del Discurso (lo que indica que probablemente se trata de un enunciado engañoso, ya que está muy por encima del umbral de 0,30), con múltiples signos de falsificación (por ejemplo, tiempo incoherente, falta de detalles, numerosos rellenos)" Esto es posiblemente más transparente y revisable que una nota genérica de que "la declaración parecía engañosa" También podría facilitar la revisión y supervisión por pares, ya que varios analistas podrían comparar sus puntuaciones CDA en la misma declaración para mantener la coherencia.

Para la investigación sobre el engaño, nuestro estudio refuerza la importancia de las pistas lingüísticas y abre vías para perfeccionar la detección de mentiras basada en el contenido. Identificamos qué marcadores eran más potentes y cuáles menos fiables, lo que nos orientó para perfeccionar en el futuro el esquema CDA. Por ejemplo, dada la menor capacidad de diagnóstico de los términos de abjuración en este contexto, se podría considerar la posibilidad de ajustar el peso de ese marcador o especificar las condiciones en las que cuenta como señal de alarma (quizá sólo cuando se combina con otros signos de incoherencia en la historia). También confirmamos la importancia de algunos indicios clásicos (por ejemplo, cantidad de detalles, autorreferencias) en una nueva muestra de mentirosos adultos, lo que aporta más apoyo a los fundamentos teóricos de la hipótesis de Undeutsch de que los recuerdos veraces son más ricos y autoimplicantes. Curiosamente, nuestros resultados también encajan con las teorías de la carga cognitiva del engaño: muchos marcadores (coberturas, rellenos, estructura desorganizada) pueden interpretarse como la carga cognitiva de un mentiroso que se hace evidente en su discurso. El CDA no mide explícitamente la carga cognitiva, pero el resultado -una narración menos coherente y confiada- es coherente con lo que produciría un engaño inductor de carga.

Trabajo futuro: Sobre la base de esta validación, la investigación futura debería probar el CDA en entornos ecológicamente más válidos, como se ha mencionado, e integrarlo

potencialmente con otras modalidades. Aunque nuestro estudio se centró únicamente en el contenido verbal, las investigaciones en el mundo real suelen combinar el análisis verbal con indicios no verbales o medidas fisiológicas. Merecería la pena comprobar si añadir el CDA a las herramientas del entrevistador mejora su éxito general, o si el CDA podría incorporarse a modelos de aprendizaje automático que también tengan en cuenta el tono vocal o las expresiones faciales. Otra dirección es explorar la eficacia **intercultural**. Los datos del MU3D incluían a participantes estadounidenses de habla inglesa. Los indicios lingüísticos de engaño identificados en el CDA proceden en gran medida de la investigación occidental en lengua inglesa. ¿Sirven estos indicios en otras lenguas? Algunas (como la caída de pronombres o palabras de relleno específicas) pueden no traducirse directamente. Adaptar el CDA para utilizarlo, por ejemplo, en español o chino requeriría ajustar las definiciones de los marcadores y validarlas con hablantes nativos. Dado que hemos proporcionado un marco cuantitativo claro, los investigadores podrían repetir un estudio similar con criterios traducidos a otro idioma. De hecho, el aspecto bilingüe de nuestros resúmenes refleja una mirada hacia la aplicabilidad internacional - sería apropiado que el ACD se probara y utilizara en contextos de lengua española (especialmente dado que el ACDC y los métodos relacionados ya se han utilizado globalmente).

En resumen, el protocolo CDA demostró un gran rendimiento en la identificación del discurso engañoso. Combina los puntos fuertes de los métodos cualitativos anteriores con un nuevo nivel de rigor cuantitativo y facilidad de uso. Al poner de relieve cuándo y dónde una narración se aparta de los patrones típicos del relato veraz de recuerdos, el ACD ayuda a detectar el engaño *dentro de* una declaración, no sólo a marcarla como falsa. Esta percepción diagnóstica (por ejemplo, darse cuenta de que "la historia se volvió menos creíble cuando el sujeto habló de un periodo de tiempo concreto") puede orientar el interrogatorio de seguimiento para resolver las incoherencias. En nuestro análisis, a menudo podíamos saber exactamente dónde era más débil la mentira de un mentiroso; por ejemplo, un repentino aumento de varios marcadores en mitad de la historia correspondía a un segmento que probablemente contenía la mentira. Los investigadores podrían utilizar esa información para centrarse en ese segmento, pedir aclaraciones y, potencialmente, conseguir que la persona revele la verdad.

Conclusión

Esta investigación validó el protocolo de Análisis de la Credibilidad del Discurso como una poderosa herramienta para la evaluación de la veracidad en las narraciones de adultos. Al codificar sistemáticamente los marcadores lingüísticos de credibilidad y agregarlos en una puntuación numérica, el CDA proporciona una medida basada en pruebas de hasta qué punto un relato determinado se ajusta a las características de un recuerdo genuino. En una muestra de 320 declaraciones verificadas experimentalmente, el CDA fue eficaz a la hora de discriminar los relatos veraces de los engañosos, logrando una precisión muy superior al azar y mejorando las capacidades de las técnicas de análisis de contenido anteriores. El desarrollo del protocolo se basó en décadas de investigación sobre el engaño -desde el análisis detallado de CBCA hasta los criterios de RM y los conocimientos lingüísticos de SCAN/IDA- y unificó estos conocimientos en un marco coherente y cuantificable. Nuestros resultados subrayan que las declaraciones veraces tienden a ser más detalladas, más coherentes desde el punto de vista cronológico y gramatical, y más personales, mientras que las engañosas suelen traicionarse a sí mismas a través de la vaguedad, la incoherencia y el lenguaje distanciador. Lo más importante es que la combinación de estas características es lo que proporciona una señal fiable. La puntuación

escalar del CDA captó esta combinación, convirtiéndola en un indicador sensible de la falta de honradez.

El éxito de la validación del CDA tiene implicaciones tanto teóricas como prácticas. Teóricamente, apoya la idea de que el engaño deja huellas polifacéticas en el lenguaje que pueden detectarse con un análisis cuidadoso. También anima a avanzar hacia modelos integradores de detección de mentiras que tengan en cuenta numerosos indicios de forma conjunta y no aislada. En la práctica, el CDA ofrece a investigadores, psicólogos y otros profesionales un método estructurado para evaluar la credibilidad, lo que puede mejorar el proceso de toma de decisiones en entornos jurídicos y de investigación. Dado que funciona a partir del contenido de las transcripciones, puede aplicarse en una amplia gama de situaciones, desde la evaluación de testimonios en casos penales hasta la comprobación de declaraciones escritas en investigaciones sobre seguros o fraudes laborales. La naturaleza cuantitativa de la puntuación del CDA aporta claridad y responsabilidad, lo que permite revisar o explicar las evaluaciones con mayor transparencia.

En conclusión, el protocolo de Análisis del Discurso de la Credibilidad representa un avance significativo en el campo de la detección del engaño y el análisis de la credibilidad. Nuestro estudio demuestra que, si nos centramos en las características de un enunciado a nivel de discurso y empleamos un sistema de puntuación sistemático, se puede lograr una gran precisión a la hora de discernir la verdad de la falsedad en las narraciones de adultos. Como ocurre con cualquier método, el perfeccionamiento y las pruebas continuas determinarán sus límites y puntos fuertes. No obstante, las pruebas actuales posicionan al ADC como una herramienta eficaz, con base científica y lista para una aplicación más amplia. Al poner el acento en cómo se *dice* algo en lugar de en *lo que se dice*, el CDA se adentra en las sutiles firmas lingüísticas de la verdad y el engaño, acercando a investigadores y profesionales un paso más hacia la evaluación fiable de la honestidad en la comunicación.

Referencias

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.