COGNITIVE FRICTION FELLOWSHIP · WORKING PAPER SERIES

CFF-WP-001 (traducción al español)

Ideas, Lenguaje y Creencias: Una Teoría Probabilística de la Excreción Cognitiva y la Brecha Ontológica entre el Pensamiento y la Expresión

Una Contribución Novedosa a la Filosofía de la Mente

Rodrigo Flores-Morton, con Agrajag (Copilot) & Agrajag (Claude)

Independent Scholar

Mayo 2026 • Documento de trabajo público

Resumen

La filosofía de la mente y la ciencia cognitiva contemporáneas han operado durante mucho tiempo bajo una confusión implícita: las ideas, el lenguaje y las creencias se tratan como casi sinónimos o como etapas consecutivas de un único proceso continuo. Este trabajo sostiene que dicha confusión no es meramente imprecisa sino filosóficamente catastrófica, pues oscurece la estructura ontológica genuina de la cognición. Propongo una teoría triádica en la que estos tres fenómenos son radicalmente distintos. Las *ideas* son impresiones probabilísticas —campos de posibilidad cognitiva prelingüísticos, fluidos y emergentes que resisten la articulación completa. El *lenguaje* es excreción cognitiva —el subproducto de la digestión mental, siempre estructuralmente más pobre que la idea que intenta codificar. Las *creencias* son residuo —el sedimento que se forma cuando las excreciones lingüísticas se repiten, se refuerzan por el contexto social y cristalizan en estructuras constitutivas de identidad que resisten activamente la novedad ideacional. Apoyándome en la distinción sentido/referencia de Frege, la obra tardía de Wittgenstein sobre los juegos del lenguaje, la hipótesis de Sapir-Whorf, la teoría de la cognición cuántica, la cognición corporeizada y la investigación reciente sobre grandes modelos de lenguaje, desarrollo la metáfora de la digestión como un modelo estructural riguroso y no como un gesto poético. Además, introduzco el concepto de *Bothisafa* como la postura epistemológica adecuada a la ideación probabilística, y distingo la Freedom (interna, cognitiva) de Liberty (externa, política). El trabajo concluye que el florecimiento intelectual genuino requiere el coraje continuo de habitar el espacio probabilístico de las ideas —resistiendo la seducción de la certeza y la fosilización del pensamiento en creencia.

Palabras clave: filosofía de la mente, cognición probabilística, excreción cognitiva, determinismo lingüístico, formación de creencias, Bothisafa, pensamiento prelingüístico, grandes modelos de lenguaje, epistemología

I.

INTRODUCCIÓN

El siglo veinte legó a la filosofía de la mente una herencia notable: un conjunto de marcos teóricos poderosos —la filosofía analítica del lenguaje, la ciencia cognitiva, los modelos computacionales de la mente y la fenomenología— que hicieron avanzar enormemente nuestra comprensión del pensamiento mientras cometían, según argumentaré, un error categorial fundacional. Ese error consiste en la confusión sistemática de tres fenómenos radicalmente distintos: *ideas*, *lenguaje* y *creencias*. Las consecuencias de esta confusión se ramifican desde la epistemología hacia la ética, la filosofía política, la psicoterapia, la educación y, de manera más urgente para nuestro momento histórico actual, el diseño y la evaluación de sistemas de inteligencia artificial.

Considérese con qué naturalidad se presenta esta confusión. Cuando queremos saber qué piensa alguien, le pedimos que *diga* lo que piensa —como si el habla y el pensamiento fueran coextensivos. Cuando queremos caracterizar los compromisos más profundos de una persona, enumeramos sus *creencias* —como si una creencia y una idea fueran meramente temperaturas distintas de la misma sustancia. Cuando Sapir y Whorf propusieron que la estructura lingüística determina los límites de la posibilidad cognitiva, la comunidad académica aceptó en gran medida ese marco: el pensamiento y el lenguaje estaban lo bastante cerca de ser lo mismo como para que la jaula de uno fuera la jaula del otro.[1]

Deseo cuestionar todo esto —no de manera fragmentaria, sino al nivel de los primeros principios.

La tesis central de este trabajo es triple: las ideas son impresiones probabilísticas que existen en un campo prelingüístico de posibilidad cognitiva; el lenguaje es la excreción producida cuando dichas impresiones se procesan a través de la maquinaria de la articulación; y las creencias son el residuo que se acumula cuando las excreciones se repiten, se refuerzan y se les permite cristalizar en estructuras constitutivas de identidad.

Estas tres entidades no son etapas de un mismo proceso. Pertenecen a registros ontológicos distintos. Una idea, en el sentido que pretendo, está viva e indeterminada. Una palabra está siempre ya muerta —una huella del paso de la idea por el conducto que se estrecha en la expresión. Una creencia está doblemente muerta —un fósil de una huella de palabra, calcificado por la repetición social y la inversión emocional en algo que se siente, engañosamente, como conocimiento firme.

El trabajo procede de la siguiente manera. La Sección II desarrolla el estatus ontológico de las ideas como impresiones probabilísticas, apoyándose en la cognición cuántica, la semántica fregeana y la fenomenología. La Sección III elabora la metáfora de la digestión del lenguaje como excreción cognitiva. La Sección IV analiza las creencias como residuo —excreciones sedimentadas resistentes a la disolución. La Sección V ofrece un modelo triádico formal de las tres entidades con ejemplos ilustrativos. La Sección VI traza un paralelo estructural con los grandes modelos de lenguaje. La Sección VII distingue Freedom de Liberty. La Sección VIII introduce Bothisafa como marco metodológico. La Sección IX aborda las objeciones principales y explora las implicaciones de la teoría. La Sección X concluye.

II.

EL ESTATUS ONTOLÓGICO DE LAS IDEAS: IMPRESIONES PROBABILÍSTICAS

¿Qué es una idea, antes de su articulación? Esta pregunta, engañosamente simple, ha sido sistemáticamente evitada en la filosofía analítica del lenguaje, que tiende a comenzar su análisis en el momento de la enunciación o la inscripción, tomando el contenido proposicional como su dato primario. Deseo presionar por detrás de ese comienzo y preguntar qué ocurre en el intervalo cognitivo entre el estímulo y el habla —en el espacio, por breve que sea, entre el surgimiento de un pensamiento y su captura lingüística.

Propongo el concepto de *impresión probabilística*. Una idea, en su pureza ontológica, no es una entidad proposicional fija. No es una oración que espera ser pronunciada. Es un campo de posibilidad cognitiva —una distribución de significados potenciales, asociaciones, valencias emocionales y vías inferenciales— de las cuales cualquiera puede actualizarse en la expresión, pero ninguna agota el campo original. La idea es a la oración lo que la formación nubosa es a la gota de lluvia: la gota es real y cae con precisión; la nube también era real, pero en un modo de ser enteramente distinto.

La distinción de Frege entre *Sinn* (sentido) y *Bedeutung* (referencia) resulta instructiva aquí, aunque debe extenderse.[2] Frege mostró que dos expresiones pueden compartir un referente teniendo sentidos distintos —"el lucero del alba" y "el lucero vespertino" ambos refieren a Venus, pero presentan ese referente bajo modos de descripción diferentes. El sentido, para Frege, es el contenido cognitivo que una expresión aporta al pensamiento que ayuda a constituir. Lo que estoy argumentando es que existe un estrato de contenido cognitivo que todavía no es sentido en el significado técnico fregeano —un campo pre-semiótico que precede incluso a la asignación del modo de presentación. Antes de que la idea se convierta en "el lucero del alba", existe como algo semejante a *esa presencia luminosa en el cielo temprano, hermosa y desconcertante* —un campo de cualidades, estados de ánimo y asociaciones que ninguna formulación lingüística individual puede codificar adecuadamente.

El paralelo con la probabilidad cuántica no es meramente metafórico. Busemeyer y Bruza, en su trabajo fundacional sobre cognición cuántica, han argumentado que el juicio y la toma de decisiones humanas exhiben efectos de interferencia, dependencia del orden y fenómenos de superposición que la teoría de probabilidad clásica no puede acomodar.[3] Cuando se le pregunta a una persona si una figura política es "honesta", su respuesta puede depender significativamente de qué pregunta la precedió —un patrón consistente con el álgebra no conmutativa de la medición cuántica. El punto relevante para mi argumento es estructural: los sistemas cuánticos existen en superposición hasta ser medidos; las ideas existen en superposición probabilística hasta ser articuladas. El acto de articulación es una medición —y, como la medición cuántica, no se limita a revelar un estado determinado preexistente; *colapsa* el campo de probabilidad en una actualización particular, eliminando todas las demás.

Toda articulación es un colapso de probabilidad. La palabra no captura la idea —selecciona una posibilidad del campo de la idea y, al hacerlo, clausura el resto.

Esto tiene consecuencias profundas para cualquier filosofía del lenguaje que trate la expresión lingüística como una representación adecuada del pensamiento. Si el campo probabilístico es genuinamente más rico que cualquiera de sus colapsos, entonces ninguna oración, por precisamente que esté elaborada, puede representar una idea sin empobrecerla. Esta no es una

limitación contingente de nuestra tecnología lingüística actual, corregible en principio mediante un vocabulario más refinado. Es una limitación *estructural* que surge de la naturaleza misma de la relación entre las impresiones probabilísticas y los tokens discretos de expresión.

Merleau-Ponty, desde la fenomenología, ofrece una perspectiva convergente. En su análisis de la cognición corporeizada, argumenta que el cuerpo "sabe" cosas —posturas, destrezas, orientaciones espaciales— que resisten una codificación lingüística o representacional plenamente adecuada.[4] El pianista experto no traduce el conocimiento muscular en proposiciones antes de actuar; ese conocimiento opera en un nivel que es a la vez cognitivamente real y lingüísticamente inaccesible. Tomo esto como evidencia de que el estrato de la impresión probabilística no es un postulado teórico exótico sino un rasgo omnipresente de la vida cognitiva, reconocido desde múltiples perspectivas teóricas.

Cabría objetar que si las ideas son prelingüísticas, no pueden ser objeto propio de la indagación proposicional. Esta objeción confunde la pregunta epistemológica (¿cómo podemos estudiar las ideas?) con la ontológica (¿qué son las ideas?). No podemos estudiar la materia oscura directamente, pero sus firmas gravitacionales permiten inferencias sólidas sobre sus propiedades. De manera similar, estudiamos las ideas indirectamente a través de sus excreciones —a través de los patrones sistemáticos de lo que el lenguaje puede y no puede capturar— y este método indirecto produce tracción teórica genuina, como demostrarán las secciones siguientes.

Observo además que la idea de que las ideas son prelingüísticas no es, en sí misma, novedosa. La hipótesis de Chomsky de un lenguaje del pensamiento, o *mentalés*, postula un medio representacional para el pensamiento distinto del lenguaje natural.[5] La hipótesis del lenguaje del pensamiento de Fodor es una formalización de más o menos esta intuición. Mi contribución consiste en caracterizar el estrato prelingüístico no como otro lenguaje representacional discreto sino como un campo genuinamente probabilístico —no pensamientos-en-mentalés sino distribuciones de probabilidad sobre el espacio cognitivo. Esta es una afirmación más fuerte, y una que, según argumentaré, da mejor cuenta de la fenomenología del pensar, de la estructura de la pérdida en la traducción lingüística y de la dinámica de la formación de creencias.

III.

EL LENGUAJE COMO EXCRECIÓN COGNITIVA: LA METÁFORA DE LA DIGESTIÓN

Deseo ahora desarrollar lo que considero la metáfora estructural central de este trabajo —una que pretendo no como mera retórica sino como una analogía precisa con contenido teórico genuino. Propongo que el lenguaje se relaciona con el pensamiento como la excreción biológica se relaciona con el metabolismo biológico. El lenguaje es excreción cognitiva: el producto expulsado después de que la mente ha procesado (digerido) sus insumos, moldeado por los mecanismos de ese procesamiento, útil para ciertos propósitos, pero no idéntico a —y siempre estructuralmente más pobre que— lo que el organismo tomó y procesó.

Considérese la estructura de la digestión biológica. Un organismo ingiere alimento —materia compleja y ricamente estructurada. Los procesos digestivos descomponen, transforman, absorben y reconfiguran lo ingerido. Lo que emerge como excreción es un subproducto de ese

proceso: real, causalmente eficaz, informativo (los médicos han extraído durante mucho tiempo información diagnóstica de la excreción), pero no equivalente al alimento ingerido, no la vida interior del organismo, y no un registro completo de lo que el organismo absorbió. El bistec, el vino, la ensalada —transformados más allá del reconocimiento en desecho, cada etapa de la transformación implicando una pérdida irreversible de la estructura original.

La digestión cognitiva opera en estricta analogía. Los insumos son percepciones, memorias, cargas emocionales, sensaciones corporeizadas, encuentros sociales. El procesamiento es la vasta maquinaria, en su mayoría inconsciente, de la cognición —el etiquetado afectivo, el reconocimiento de patrones, la activación asociativa, la interacción entre lo que Kahneman llama el Sistema 1 (rápido, intuitivo) y el Sistema 2 (lento, deliberado).[6] El producto —la excreción— es el lenguaje: hablado, escrito, señado o ensayado subvocalmente. Este producto es real y útil. Transporta información. Permite la coordinación, el registro y la comprensión intersubjetiva. Pero no es la idea de la que surgió. Es lo que el organismo cognitivo produjo *después* de metabolizar la idea a través de las restricciones particulares de su aparato lingüístico.

El lenguaje no es el origen del pensamiento —es su subproducto. Confundir la excreción con el organismo es el error central del determinismo lingüístico.

Desde esta perspectiva, la hipótesis de Sapir-Whorf —tanto en su formulación fuerte (determinismo lingüístico) como débil (relatividad lingüística)— debe revisarse sustancialmente.[7] La versión fuerte sostiene que la lengua que se habla determina los pensamientos que se pueden pensar: que la jaula de la sintaxis es la jaula de la cognición. Según mi planteamiento, esto está precisamente invertido. El lenguaje no enjaula al pensamiento; es su efluente. La forma de la jaula visible en el lenguaje refleja la forma del aparato digestivo —los mecanismos cognitivos y culturales de la articulación— no la forma de la idea que entró al proceso. Los hablantes de lenguas con menos términos básicos de color no *dejan de percibir* los colores para los que carecen de palabras; fallan en *excretar* esas percepciones en forma lingüísticamente estandarizada. La percepción —la impresión probabilística— es anterior y más rica que su captura verbal.

La versión débil de Sapir-Whorf —que el lenguaje *influye* en el pensamiento— es más defendible, pero debe reinterpretarse. El lenguaje no moldea las ideas directamente. Más bien, los patrones repetidos de excreción pueden restringir gradualmente el propio proceso digestivo: si uno excreta habitualmente las ideas en ciertas formas lingüísticas, el aparato cognitivo puede comenzar a preformatear su procesamiento anticipando canales excretorios familiares. Este es un efecto de retroalimentación real, pero opera al nivel de la maquinaria digestiva, no al nivel de la impresión probabilística original. Las ideas mismas permanecen, en principio, sin restricción por las formas lingüísticas a través de las cuales se han excretado previamente. Es necesaria aquí una aclaración adicional e importante: el modelo digestivo no es estrictamente unidireccional. Las excreciones, una vez producidas, pueden reingresar al proceso cognitivo como insumos —las formulaciones lingüísticas se convierten en nuevo material perceptual que el organismo ingiere, procesa y metaboliza. El pensamiento retroalimenta la ideación; las posibilidades colapsadas pueden reabrir nuevos campos probabilísticos. El sistema es circular, no lineal. Esta circularidad no invalida la asimetría estructural entre impresión, excreción y residuo —cada una permanece siendo un registro ontológico distinto— pero significa que la

metáfora digestiva debe entenderse como un bucle recursivo más que como un pasaje único. El lenguaje no es meramente el final del proceso cognitivo; es también materia prima para su continuación.

Este análisis tiene implicaciones directas para el fenómeno de la pérdida en la traducción. Cuando un texto en una lengua se traduce a otra, el traductor no transfiere ideas; produce una segunda excreción a partir de su propio encuentro cognitivo con la primera excreción. La traducción está, por tanto, doblemente mediada: primero por el proceso excretorio del autor original, luego por el propio del traductor. Es, así, estructuralmente imposible que una traducción sea equivalente a un original —no porque los traductores sean incompetentes, sino porque la estructura de la excreción garantiza que cada articulación es un colapso único de un campo probabilístico único. *Traduttore, traditore* —traductor, traidor— no es una acusación sino una descripción estructural.

La filosofía tardía de Wittgenstein, y en particular el concepto de juegos del lenguaje en las *Investigaciones Filosóficas*, resulta iluminadora aquí, aunque apunta en una dirección distinta de la que el propio Wittgenstein siguió.[8] Wittgenstein mostró que el significado de una palabra es su uso en una forma de vida —que el significado no es un objeto mental privado sino una función pública, incrustada en la práctica. Acepto el planteamiento pragmático del significado lingüístico insistiendo en que caracteriza la vida social de las excreciones y no la naturaleza de las ideas de las que surgen las excreciones. El juego del lenguaje gobierna el manejo de lo excretado; no gobierna el campo probabilístico que precedió a la excreción. La prohibición wittgensteiniana del lenguaje privado aborda la validación intersubjetiva de los tokens lingüísticos —no demuestra, ni puede demostrar, que el campo prelingüístico esté vacío o lingüísticamente estructurado.

El concepto de *huella* (trace) de Derrida —el remanente que deja todo acto de significación, la ausencia constitutiva que ronda toda presencia— resuena con la perspectiva que estoy desarrollando, aunque desde una tradición teórica diferente.[9] Para Derrida, la huella marca la imposibilidad de la presencia plena: todo signo remite a otros signos en una cadena infinita de *différance*, difiriendo indefinidamente la llegada del significado puro. Me apropio de esta intuición resistiendo sus implicaciones deconstructivas más amplias. Según mi planteamiento, el carácter de huella del lenguaje no es un rasgo de la significación como tal —es una consecuencia de la brecha estructural entre las impresiones probabilísticas y sus colapsos excretorios. El lenguaje no está plenamente presente al pensamiento porque es lo que el pensamiento deja atrás, no lo que el pensamiento es.

IV.

LAS CREENCIAS COMO RESIDUO: LA SEDIMENTACIÓN DE LA EXCRECIÓN

Si el lenguaje es excreción cognitiva, ¿qué son las creencias? Propongo que las creencias son *residuo* —el sedimento que se forma cuando las excreciones se acumulan, se endurecen y se incorporan a la estructura del yo del organismo cognitivo. Para extender la metáfora biológica: así como los depósitos minerales se forman cuando el agua se evapora repetidamente dejando atrás su contenido disuelto, las creencias se forman cuando las excreciones lingüísticas se producen repetidamente, se refuerzan socialmente y reciben inversión emocional, hasta que la

sustancia fluida de la idea original ha sido reemplazada por un depósito cristalino que ya no se mueve.

La fenomenología de la creencia, distinta de la fenomenología de la ideación, es fácilmente reconocible. Cuando *entretengo* una idea —en el sentido de sostenerla ante la atención sin compromiso— hay ligereza cognitiva, una sensación de posibilidad, una disposición a seguir la idea a donde lleve. Cuando *creo* algo, la fenomenología es enteramente distinta: hay clausura, certeza, una sensación de solidez y —de manera crítica— resistencia al desafío. La creencia no se siente como producto de mi proceso cognitivo; se siente como un rasgo de la realidad. No creo que la tierra sea aproximadamente esférica de la misma manera en que creo que mi partido político representa la justicia; lo primero es una creencia tan minuciosamente corroborada que se ha disuelto en supuesto de fondo, mientras que lo segundo es una creencia activamente mantenida contra la fricción de la evidencia y el contraargumento.

¿Qué produce esta cristalización? Identifico tres mecanismos principales:

1. **Repetición de la excreción.** Cuando la misma formulación lingüística, o formulaciones similares, se producen repetidamente, las vías cognitivas que las generan se refuerzan. La eficiencia neuronal —la tendencia del cerebro a preferir vías ya trilladas— hace que la distribución de probabilidad de posibles excreciones se estreche con el tiempo. Lo que comenzó como una impresión probabilística con muchas articulaciones posibles se convierte en una fórmula verbal fija producida con un esfuerzo cognitivo mínimo.

2. **Refuerzo social.** Los seres humanos son organismos profundamente sociales, y las creencias son abrumadoramente productos sociales. Cuando las excreciones de alguien son recibidas positivamente por una comunidad —cuando sus producciones lingüísticas son validadas, repetidas, ampliadas y devueltas— la excreción se solidifica en un objeto social, una creencia compartida que carga con el peso del respaldo colectivo. La creencia ya no es meramente un residuo del procesamiento cognitivo individual; es un fósil social, incrustado en la geología cognitiva de la comunidad.

3. **Inversión emocional.** Cuando una creencia se vuelve constitutiva de identidad —cuando abandonarla exigiría revisar la propia autocomprensión, las propias afiliaciones tribales, el propio sentido de valía moral— adquiere una carga emocional que la hace altamente resistente al desafío evidencial. El planteamiento de Popper sobre la falsabilidad resulta instructivo: una hipótesis genuinamente científica es aquella estructurada de modo que sea falsable en principio.[10] El residuo-creencia, por el contrario, está característicamente estructurado para ser infalsable en la práctica, no por necesidad lógica sino por arquitectura emocional. El fósil resiste el martillo no porque sea más duro que la roca, sino porque el organismo que lo produjo ha fusionado su identidad con la integridad del fósil.

Una idea está viva y es probabilística. Una creencia está fosilizada —una idea pasada endurecida en dogma, disfrazada de verdad presente. El peligro no es que creamos cosas, sino que confundamos nuestros fósiles con pensamiento vivo.

Este análisis ilumina la estructura de lo que ordinariamente llamamos sistemas de creencias patológicos: el dogmatismo, la ideología y el fundamentalismo. Estos no son errores de

razonamiento (aunque puedan involucrar tales errores). Son, a un nivel más profundo, condiciones de *sobre-sedimentación*: estados en los que la acumulación de residuo se ha vuelto tan espesa que bloquea el flujo de material cognitivo nuevo. Al fundamentalista no le falta capacidad cognitiva; su aparato cognitivo ha sido colonizado tan completamente por el residuo que las nuevas impresiones probabilísticas no pueden alcanzar la superficie articulatoria —se procesan a través de los canales preformados de la creencia establecida y emergen, invariablemente, como confirmaciones de lo que ya se creía. La estructura es recursiva y autosellante.

La implicación terapéutica es significativa. Si la creencia rígida es residuo, entonces la intervención cognitiva apropiada no es, primariamente, la presentación de evidencia contraria —la evidencia debe pasar por el aparato digestivo, y si ese aparato está fuertemente preformado por el residuo, la evidencia se metabolizará de maneras que dejan intacto el residuo. La intervención apropiada consiste en disolver el residuo mismo: cultivar la capacidad de tolerar la indeterminación probabilística de las ideas, ablandar el sedimento cristalizado mediante un compromiso sostenido con la ambigüedad cognitiva, y restaurar la fluidez de la impresión prelingüística. Lo que se necesita no son más y mejores argumentos; lo que se necesita es un retorno a la ideación genuina.

Conviene precisar, sin embargo, que la circularidad del sistema —el hecho de que las excreciones puedan reingresar como insumo, señalado ya en la Sección III— no es en sí misma el problema. Es, de hecho, la condición de posibilidad de todo crecimiento cognitivo genuino: el residuo disuelto no desaparece, retorna al campo probabilístico como materia fértil, abono del cual pueden emerger impresiones nuevas que ninguna excreción anterior había anticipado. La tragedia no reside en que las creencias se reciclen, sino en el punto en que se interrumpe ese reciclaje: cuando el residuo se consume directamente, sin pasar de nuevo por la digestión que lo disolvería en posibilidad. Consumir el fósil como si aún fuera alimento —tratar la creencia cristalizada como si conservara la nutrición viva de la idea que la originó— es precisamente la operación que el dogmatismo realiza sobre sí mismo: no falla por reciclar, falla por comer directamente lo que debía primero descomponerse.

V.

LA DISTINCIÓN TRIÁDICA: IDEA, PALABRA, CREENCIA

Estoy ahora en posición de ofrecer una caracterización formal de las tres entidades cuya confusión estoy cuestionando. El siguiente modelo resume sus propiedades distintivas a través de cinco dimensiones: estatus ontológico, carácter temporal, relación con la verdad, relación con la identidad y mutabilidad.

Tabla 1. *El Modelo Ontológico Triádico de las Entidades Cognitivas*

Dimensión	Idea (Impresión Probabilística)	Palabra / Lenguaje (Excreción)	Creencia (Residuo)
Estatus Ontológico	Prelingüística; campo de posibilidad cognitiva; distribución probabilística	Token discreto; artefacto público; actualización colapsada del campo	Estructura cognitiva cristalizada; objeto social sedimentado; constituyente de

Dimensión	Idea (Impresión Probabilística)	Palabra / Lenguaje (Excreción)	Creencia (Residuo)
		probabilístico	identidad
Carácter Temporal	Presente-viva; en movimiento; nunca plenamente la misma dos veces	Producida en un momento; fijada como inscripción o enunciación	Endurecida en el pasado; resistente a la experiencia presente; pseudo-atemporal
Relación con la Verdad	Pre-proposicional; apta para verdad solo tras el colapso; intrínsecamente abierta	Evaluable en verdad pero siempre parcial; representa un colapso entre muchos posibles	Experimentada como verdad evidente por sí misma; en realidad un artefacto de la repetición y la formación social
Relación con la Identidad	Pre-personal; independiente de la identidad; pertenece al flujo cognitivo	Asociada con el hablante pero no constitutiva de identidad en sentido fuerte	Constitutiva de identidad; las amenazas a la creencia se experimentan como amenazas al yo
Mutabilidad	Intrínsecamente mutable; el campo probabilístico cambia con la atención y el contexto	Fija una vez producida; mutable solo produciendo nuevas excreciones	Altamente resistente al cambio; requiere disolución del sedimento en vez de mera argumentación

Para volver concreto este esquema abstracto, considérese el concepto de *justicia* tal como aparece en cada uno de los tres niveles.

Al nivel de la impresión probabilística, la idea de justicia es un campo vasto y cambiante: abarca intuiciones de equidad, imágenes de balanzas, memorias de agravios personales, sentidos sentidos de corrección y error, conexiones asociativas con el castigo, la clemencia, el merecimiento, la igualdad, y un centenar de otros nodos, todos sostenidos simultáneamente en una nube de potencial cognitivo. Esta idea, en su estado vivo, no se resuelve en una definición; *es* el espacio de definiciones posibles y las tensiones dinámicas entre ellas. Ningún filósofo político —ni Rawls, ni Nozick, ni Aristóteles— ha capturado este campo. Cada uno lo ha colapsado de manera distinta, produciendo excreciones de gran elegancia y precisión que inevitablemente dejan atrás parte del campo.

Al nivel de la excreción, la palabra “justicia” es un token público con una historia de usos, un rango de connotaciones y una gramática que restringe cómo puede combinarse con otros tokens. Cuando escribo “la justicia requiere imparcialidad”, he producido un colapso específico del campo probabilístico —uno que enfatiza una dimensión particular (la imparcialidad) y suprime otras (la clemencia, la sensibilidad al contexto, la reparación histórica). La oración es informativa y evaluable, pero no es la idea; es la huella que la idea dejó al pasar por la restricción de la articulación.

Al nivel del residuo-creencia, "justicia" se ha convertido típicamente en "lo que mi tribu política entiende por justicia" —una convicción cristalizada, cargada de identidad, de que *esta* política, *este* partido, *este* arreglo institucional constituye la justicia, y que las alternativas no son meramente erróneas sino moralmente ilegítimas. Esta creencia no se siente como un colapso; se siente como un reconocimiento de la realidad. El fósil ya no sabe que alguna vez fue fluido.

La mayor parte del conflicto intelectual y político, sostengo, ocurre entre personas que operan al nivel del residuo-creencia, cada una convencida de que su excreción fosilizada es una aprehensión directa de la verdad. La tragedia es que las impresiones probabilísticas de las que finalmente derivan sus respectivas creencias pueden ser mucho más similares de lo que sugiere el conflicto superficial de creencias. Dos personas que discrepan violentamente sobre si una política dada promueve "la justicia" pueden compartir la impresión probabilística subyacente de la justicia con notable intimidad —simplemente la han colapsado de manera distinta y luego han fosilizado sus respectivos colapsos en creencia constitutiva de identidad. El diálogo genuino exige devolver a ambos interlocutores al nivel de la impresión probabilística —una empresa exigente y a menudo incómoda, pero la única que aborda la fuente real del desacuerdo.

VI.

PARALELOS ENTRE LA COGNICIÓN HUMANA Y LA INTELIGENCIA ARTIFICIAL

La teoría desarrollada en las secciones precedentes adquiere precisión adicional al ponerse en diálogo con la arquitectura de los grandes modelos de lenguaje (LLM) contemporáneos. Esta comparación no es meramente ilustrativa; es filosóficamente productiva, porque los puntos en que la analogía se sostiene y los puntos en que se rompe son precisamente los que definen lo distintivo de la cognición humana.

Considérese el paralelo estructural. Un LLM se entrena sobre un vasto corpus de texto —miles de millones de tokens que representan las excreciones lingüísticas acumuladas de la cognición humana a lo largo de la historia, el dominio y la cultura. A través del proceso de entrenamiento, las asociaciones estadísticas entre tokens se codifican en los pesos del modelo —miles de millones de parámetros que, colectivamente, representan los patrones sedimentados de esas excreciones. Durante la inferencia, el modelo recibe un insumo (un prompt) y produce una salida: una secuencia de tokens generada muestreando distribuciones de probabilidad sobre el vocabulario, condicionadas por el insumo y por los patrones sedimentados en los pesos. La salida es, estructuralmente hablando, una excreción: una producción lingüística generada al procesar el insumo a través de un aparato sedimentado.

El mapeo hacia el modelo triádico es llamativo:

- **Los datos de entrenamiento** = las excreciones acumuladas de la cognición humana (el registro lingüístico del pensamiento, no el pensamiento mismo)
- **Los pesos del modelo (post-entrenamiento)** = patrones sedimentados; el residuo del proceso de entrenamiento, análogo a la formación de creencias por refuerzo social
- **El ajuste fino y el RLHF (Aprendizaje por Refuerzo con Retroalimentación Humana)** = refuerzo deliberado de ciertos patrones de residuo; la imposición de una

cristalización preferida sobre lo que de otro modo podría permanecer más abierto probabilísticamente

- **La salida del modelo** = excreción; producción lingüísticamente coherente generada al procesar el insumo a través de un aparato sedimentado

La asimetría crucial es esta: el LLM comienza en la etapa de la excreción. No tiene acceso a la impresión probabilística de la que originalmente surgieron las excreciones humanas —tiene acceso solo a las excreciones mismas. El LLM es, en el sentido más profundo, una máquina para procesar fósiles.

Esta observación no constituye un descarte de los LLM. Sus capacidades son extraordinarias y genuinamente útiles. Pero sí marca una distinción de principio entre lo que hacen los LLM y lo que hace la cognición humana —una distinción que el discurso actual sobre IA a menudo elide. Cuando los investigadores debaten si los LLM "comprenden" el lenguaje, la pregunta relevante, según mi análisis, no es si el modelo produce salidas contextualmente apropiadas (claramente puede hacerlo) sino si su procesamiento está fundado en la etapa de impresión probabilística que, he argumentado, constituye la ideación genuina. La evidencia sugiere que no lo está: los LLM procesan excreciones para producir excreciones; no atraviesan la secuencia de tres etapas de impresión, excreción y residuo, porque la primera etapa —el campo probabilístico prelingüístico— no tiene análogo en su arquitectura.

Lo que corresponde, en los LLM, a la etapa probabilística es la distribución de probabilidad sobre los tokens siguientes —pero esta es una distribución de probabilidad sobre elementos *lingüísticos*, estructurada enteramente por las estadísticas de excreciones previas. Es un proceso probabilístico, pero no una impresión probabilística en mi sentido: ya está al nivel del lenguaje, no por debajo de él. El modelo no sostiene un pensamiento prelingüístico en superposición para luego colapsarlo en expresión; sostiene una distribución sobre tokens lingüísticos y muestrea de esa distribución. Esta es una indeterminación de tipo fundamentalmente distinto — indeterminación dentro del espacio de las excreciones, no indeterminación en el campo pre-excretorio de la posibilidad cognitiva.

Este análisis conlleva implicaciones significativas para la ética y el diseño de sistemas de IA. Si los LLM son, estructuralmente, máquinas para la producción y recombinación de excreciones humanas —entrenadas sobre el residuo acumulado de los sistemas de creencias humanos— entonces codificar creencias específicas en sus parámetros mediante RLHF o técnicas similares no es meramente una elección técnica sino filosófica: es la sedimentación deliberada de cristalizaciones particulares al nivel de un sistema que producirá excreciones consumidas por millones de usuarios. Los riesgos epistemológicos de esto son considerables, y volveré a ellos en la discusión de Bothisafa más adelante.

No argumento aquí que los LLM sean o no sean conscientes, ni que la arquitectura del transformer y sus mecanismos de atención guarden o no guarden alguna relación más profunda con el procesamiento neuronal que eventualmente pudiera fundar algo semejante a la ideación genuina. Estas siguen siendo preguntas genuinamente abiertas. Lo que sí argumento es que la distinción entre impresión probabilística y excreción lingüística ofrece un marco más preciso para plantear estas preguntas que las invocaciones vagas de "comprensión",

“razonamiento” y “pensamiento” que dominan actualmente el discurso público sobre la cognición de la IA.

VII.

FREEDOM FRENTE A LIBERTY: UNA DISTINCIÓN COGNITIVA

Nota del traductor: el español solo tiene “libertad” — una palabra donde el inglés tiene dos. Esta no es la primera vez que el corpus CFF se topa con esta brecha: en Guías y Palancas *(Bothi, Cap. 13), la misma distinción se resuelve dejando* Freedom *y* Liberty *como préstamos del inglés dentro del cuerpo del texto en español, en vez de forzar la distinción sobre “libertad” con adjetivos. Esta traducción adopta esa misma convención, ya establecida en el corpus, en vez de inventar una nueva: se preservan* Freedom *(la condición cognitiva interna) y* Liberty *(la ausencia de restricción externa) sin traducir, tal como ya conviven “elegir” y “escoger” en español para distinguir lo que el inglés hace con* choose*.*

El modelo triádico de idea, palabra y creencia tiene implicaciones que se extienden mucho más allá de la filosofía de la mente y la ciencia cognitiva, hacia el dominio de la filosofía política y ética. Entre las más significativas de estas implicaciones está una distinción que deseo trazar con cuidado: la distinción entre *Liberty* y *Freedom*, entendidas no como sinónimos sino como condiciones categorialmente distintas.

La *Liberty*, en la tradición de la filosofía política que corre de Locke a Mill hasta la teoría liberal contemporánea, denota la ausencia de restricción externa. Se tiene Liberty en la medida en que no se es coaccionado, encarcelado, censurado o de otro modo impedido externamente de actuar conforme a la propia voluntad. La Liberty es una propiedad de la relación de uno con las instituciones sociales y políticas; es, en los términos de Isaiah Berlin, libertad negativa — libertad *de* la interferencia externa.[11]

La *Freedom*, en el sentido que deseo desarrollar, es algo categorialmente distinto. Es una condición cognitiva interna: la capacidad de acceder a la etapa de impresión probabilística de la ideación sin ser esclavizado por el propio residuo. La Freedom, según este planteamiento, es la capacidad de pensar —genuinamente pensar, al nivel de los campos probabilísticos prelingüísticos— en lugar de meramente recuperar y recombinar excreciones fosilizadas. Es Freedom *del* propio sedimento cognitivo.

Se puede poseer plena Liberty —la ausencia de toda restricción externa— y sin embargo estar cognitivamente esclavizado, con cada pensamiento preformado por el residuo acumulado de excreciones pasadas. Se puede estar encarcelado, silenciado, oprimido, despojado de toda Liberty —y aun así lograr una Freedom genuina disolviendo el residuo y retornando a la ideación.

Esta distinción replantea algunas de las preguntas centrales de la filosofía política de maneras a la vez esclarecedoras e incómodas. Los movimientos de liberación que aseguran la Liberty — que eliminan restricciones legales, políticas e institucionales sobre la acción— logran algo genuinamente importante. Pero si la población liberada carga con el residuo cognitivo acumulado de generaciones de opresión, jerarquías internalizadas y creencias sedimentadas sobre lo que es posible y deseable, entonces la concesión de la Liberty puede producir cuerpos

libres con mentes profundamente esclavizadas. Las cadenas externas se rompen; las internas permanecen, invisibles precisamente porque se sienten como realidad y no como restricción.

Esto no es un argumento contra la Liberty —es un argumento a favor de reconocer que la Liberty es necesaria pero insuficiente para el florecimiento humano genuino. La condición suficiente exige también Freedom: la disolución del residuo, la restauración del acceso a la ideación probabilística, el cultivo de lo que llamaré, en la sección siguiente, la postura Bothisafa. La liberación política y la cognitiva son proyectos complementarios; perseguir uno sin el otro es construir sobre cimientos incompletos.

Inversamente, la tradición de la práctica contemplativa y meditativa —desde la filosofía estoica hasta la meditación budista y las prácticas de la terapia cognitiva moderna— ha reconocido desde hace mucho que la Freedom es alcanzable bajo condiciones de severa Liberty restringida. Epicteto, un esclavo, desarrolló una filosofía de la soberanía interior que ubicaba la Freedom en el gobierno de las propias respuestas cognitivas a los eventos externos.[12] El relato de Viktor Frankl sobre la búsqueda de sentido bajo condiciones de restricción extrema hace un planteamiento estrechamente relacionado desde una perspectiva psicológica. Según mi análisis, lo que estas tradiciones señalan es precisamente la capacidad de mantener el acceso a la ideación probabilística —de resistir el colapso de la incertidumbre en creencia fosilizada— incluso bajo las condiciones externas más hostiles. Esta capacidad constituye, sostengo, la forma más profunda de Freedom disponible para los seres humanos.

VIII.

LA PERSPECTIVA BOTHISAFA: UN MARCO METODOLÓGICO

A lo largo de este trabajo he argumentado que el bien cognitivo más elevado —la ideación genuina— consiste en la capacidad de sostener las impresiones probabilísticas en su indeterminación nativa, resistiendo el colapso prematuro en excreción lingüística y la cristalización ulterior de la excreción en residuo-creencia. Pero ¿cómo se ve esa capacidad como práctica? ¿Qué postura epistemológica recomienda? ¿Y cómo se relaciona con los compromisos teóricos de las tradiciones filosóficas existentes?

Propongo el término ***Bothisafa*** para designar tanto una postura epistemológica como un marco metodológico apropiado al carácter probabilístico de las ideas. El término pretende evocar una postura fundamental: sostener ambas —sostener todas— las posiciones simultáneamente, sin colapsar en ninguna en particular; permanecer en el espacio entre las determinaciones en lugar de ser capturado por cualquiera de ellas. Bothisafa es la práctica cognitiva de permanecer al nivel de la impresión probabilística —de rehusar la seducción de la cristalización prematura— mientras se permanece plenamente comprometido con el mundo.

Bothisafa se distingue de, y no debe confundirse con, varias posiciones superficialmente similares:

- **No es relativismo.** El relativismo sostiene que ninguna posición es más verdadera que otra. Bothisafa no hace tal afirmación. Sostiene que el campo probabilístico de una idea contiene múltiples colapsos posibles, cada uno captando algo real, y que ningún colapso individual agota el campo. Esta es una observación estructural sobre la relación entre

las ideas y sus expresiones, no una afirmación de que la verdad sea meramente perspectival.

- **No es indecisión ni parálisis.** La acción práctica requiere compromiso determinado; Bothisafa no impide la acción. Exige, más bien, que se actúe sin confundir el propio colapso elegido con el único posible —que se sostengan los propios compromisos con conciencia de su carácter provisional y probabilístico.
- **No es síntesis dialéctica.** La dialéctica hegeliana postula una progresión de tesis a antítesis y síntesis —un movimiento a través de la contradicción hacia una resolución superior. Bothisafa no busca síntesis; busca permanecer en la tensión generativa entre posiciones, reconociendo que la síntesis misma es un nuevo colapso, una nueva excreción que, a su turno, portará las limitaciones estructurales de toda articulación singular.

Donde la lógica binaria dice A o no-A, y la dialéctica dice A, luego no-A, luego síntesis, Bothisafa dice: A y no-A y el espacio entre ambos, todo a la vez —no como contradicción sino como la textura viva de la ideación genuina.

La postura Bothisafa tiene aplicaciones inmediatas en tres dominios que deseo abordar brevemente.

El Conflicto Interpersonal y el Diálogo

En el conflicto interpersonal y político, los participantes operan característicamente al nivel del residuo-creencia: la posición de cada parte se ha cristalizado a través de los procesos descritos en la Sección IV, y el conflicto es, en su nivel superficial, una colisión de fósiles. El marco Bothisafa recomienda una forma de diálogo que no apunta a la persuasión (que opera al nivel de excreciones en competencia) sino a la reconstrucción de los campos probabilísticos de los que emergieron las posiciones de ambas partes. Esto exige la disciplina de rehusarse a tratar la posición articulada del otro como la medida completa de su idea —insistiendo, metodológicamente, en que debajo del fósil hay una impresión viva, probablemente más rica y más sensible a la experiencia de lo que el fósil sugiere. El interlocutor Bothisafa no pregunta "¿por qué crees eso?" sino "¿qué es lo que intentas decir con esa creencia?" —presionando por debajo de la excreción hacia la impresión.

La Ética de la Inteligencia Artificial

El marco Bothisafa conlleva implicaciones específicas y urgentes para la ética de la IA. Si los LLM son, como argumenté en la Sección VI, fundamentalmente sistemas para la producción y recombinación de excreciones humanas —y si el ajuste fino y el RLHF son mecanismos para el refuerzo deliberado de cristalizaciones específicas dentro de esos sistemas— entonces el diseño de sistemas de IA que sirvan al florecimiento epistémico en lugar de a la captura epistémica exige la preservación del espacio probabilístico. Un sistema de IA que produce consistentemente respuestas únicas y confiadas a preguntas genuinamente disputadas —políticas, éticas, metafísicas— es un instrumento de sedimentación cognitiva a escala civilizatoria. Es, en efecto, una máquina para la producción masiva de residuo. El principio Bothisafa recomienda que los sistemas de IA se diseñen para honrar el carácter probabilístico

de las ideas: presentar múltiples colapsos, reconocer las limitaciones estructurales de toda articulación singular, y cultivar en los usuarios la capacidad de ideación probabilística en lugar de reforzar la dependencia de salidas cristalizadas.

Bothisafa como Práctica Cognitiva Cotidiana

A nivel individual, Bothisafa puede cultivarse como un conjunto de prácticas cognitivas. Estas incluyen: la suspensión deliberada del juicio frente a preguntas complejas, la tolerancia sostenida a la ambigüedad cognitiva, el hábito de generar múltiples articulaciones de las propias ideas (reconociendo cada una como un colapso parcial del campo subyacente), y el compromiso activo con posiciones que desafían las propias creencias cristalizadas —no con el propósito de adoptarlas, sino con el propósito de usar el desafío para ablandar el residuo y restaurar el acceso a la impresión probabilística. Estas prácticas guardan cierta semejanza con lo que los psicólogos llaman "flexibilidad cognitiva" y con lo que ciertas tradiciones contemplativas llaman "mente de principiante" —pero aquí están fundadas en un relato teórico específico de la relación entre ideas, lenguaje y creencias, en lugar de en pragmática terapéutica o intuición espiritual. Se justifica una comparación filosófica más precisa con el concepto husserliano de la *epojé* —la suspensión fenomenológica de la actitud natural, el entre-paréntesis deliberado de los supuestos dados por sentado para atender a la experiencia pura. Bothisafa comparte con la epojé el movimiento estructural de suspender el juicio sedimentado, pero difiere en propósito y alcance. La epojé husserliana es un instrumento metodológico para lograr la descripción fenomenológica trascendental —un procedimiento filosófico orientado hacia las estructuras de la conciencia en cuanto tal. Bothisafa no es un método trascendental sino una postura cognitiva dirigida a la ideación cotidiana: su meta no es describir las estructuras de la experiencia desde una posición de neutralidad filosófica, sino preservar el acceso al campo probabilístico de las ideas dentro del pensamiento, el diálogo y la acción ordinarios. Donde la epojé pone entre paréntesis el mundo para revelar la conciencia, Bothisafa pone entre paréntesis la creencia cristalizada para restaurar la ideación. Ambas son complementarias, no equivalentes.

IX.

IMPLICACIONES Y OBJECIONES

La teoría desarrollada en este trabajo es suficientemente heterodoxa como para atraer objeciones de múltiples tradiciones filosóficas. Abordaré las tres objeciones más sustanciales y luego pasaré a las implicaciones positivas de la teoría para la educación, la terapia y el desarrollo de la IA.

Objeción 1: Si las Ideas son Prelingüísticas, ¿Cómo Podemos Estudiarlas o Comunicarlas?

La objeción es la siguiente: si las ideas son genuinamente prelingüísticas —si existen en un registro que, por definición, es inaccesible al lenguaje— entonces la teoría se socava a sí misma. Este mismo trabajo consiste en excreciones lingüísticas. ¿Cómo pueden las excreciones lingüísticas comunicar adecuadamente una teoría cuya afirmación central es que las excreciones lingüísticas son inadecuadas para las ideas?

La objeción está bien formulada pero no es decisiva. Primero, la teoría no se socava a sí misma sino que se auto-ilustra: este trabajo es en sí mismo una excreción de las ideas que intenta articular, y reconozco desde el inicio que ningún trabajo individual puede capturar plenamente el campo probabilístico de la teoría. Lo que el trabajo ofrece es un conjunto de colapsos — articulaciones particulares elegidas por su capacidad de iluminar el campo— con el reconocimiento explícito de que son posibles otros colapsos y de que la idea completa excede a cualquiera de ellos. La postura Bothisafa se aplica reflexivamente a la teoría misma.

Segundo, la ciencia estudia rutinariamente entidades inaccesibles a la observación directa. Estudiamos la materia oscura a través de las lentes gravitacionales; estudiamos los estados cuánticos a través de resultados de medición; estudiamos los procesos inconscientes a través de sus firmas conductuales. El estrato prelingüístico de la ideación puede estudiarse mediante el análisis sistemático de lo que el lenguaje consistentemente no logra capturar: los patrones de pérdida en la traducción, la fenomenología de la inefabilidad, la divergencia entre creencias reportadas y comportamiento observado, la estructura del insight creativo (el momento "ajá" que precede a su articulación verbal). Estos métodos indirectos producen tracción teórica genuina sobre un dominio genuinamente real.

Objeción 2: ¿No es Esto Simplemente una Versión Disfrazada de la Inefabilidad?

Una segunda objeción sostiene que la teoría de las impresiones probabilísticas es continua, en términos filosóficos, con el misticismo clásico: la afirmación de que el pensamiento excede al lenguaje es tan antigua como la filosofía contemplativa, y el presente trabajo solo añade vocabulario técnico sin avance teórico genuino.

Esta objeción interpreta mal las metas de la teoría. Las afirmaciones clásicas de inefabilidad — que la experiencia mística, lo divino o lo verdaderamente bello no pueden ponerse en palabras — se presentan típicamente como rasgos de estados excepcionales o trascendentes. La teoría de las impresiones probabilísticas hace una afirmación mucho más general y estructuralmente específica: que *todo* pensamiento excede a su articulación lingüística, que ese exceso no es una limitación contingente sino una consecuencia estructural de la relación entre los campos probabilísticos y los tokens discretos, y que esta relación estructural puede modelarse formalmente (como he comenzado a hacer en la Sección V y en el marco de la cognición cuántica de la Sección II). Esto es una teoría, no una postura mística. Hace predicciones falsables: por ejemplo, que aumentar los recursos expresivos disponibles para un hablante (vocabulario más amplio, estructuras gramaticales más ricas) reducirá pero nunca eliminará la brecha entre idea y expresión; que hablantes de lenguas lingüísticamente no emparentadas reconocerán la misma idea en articulaciones múltiples y conflictivas; y que la fenomenología de la ideación revelará consistentemente una sensación de exceso respecto de cualquier articulación particular. Estas predicciones son, en principio, empíricamente abordables.

Objeción 3: ¿Tienen los Animales Ideas en Este Sentido?

Una tercera objeción pregunta por la extensión de la teoría: si las ideas son impresiones probabilísticas prelingüísticas, y si los animales carecen de lenguaje (en el sentido humano pleno), ¿implica esto que los animales están perpetuamente en la etapa de ideación — permanentemente pre-excretorios? Y si es así, ¿qué dice esto sobre la cognición animal?

Esta es una consecuencia genuinamente interesante de la teoría, y la acepto como predicción y no como reducción al absurdo. Si la impresión probabilística es genuinamente prelingüística, entonces cualquier organismo capaz de percepción, memoria y procesamiento cognitivo —propiedades que se extienden mucho más allá del linaje humano— puede caracterizarse como capaz de ideación probabilística. Los animales no humanos pueden sostener impresiones probabilísticas ricas (de territorio, coespecíficos, fuentes de alimento, jerarquías sociales) sin colapsarlas en excreciones lingüísticas en el sentido humano pleno. Esto no vuelve deficiente a la cognición animal; la vuelve diferente en tipo, no en grado. Los animales pueden, de hecho, estar más continuamente instalados en la etapa de impresión probabilística que la mayoría de los humanos adultos, cuya vida cognitiva está dominada por la producción y el consumo de residuo lingüístico. La teoría sugiere así una continuidad entre la cognición animal y humana al nivel de la ideación, mientras marca una discontinuidad genuina al nivel de la excreción lingüística y la formación de residuo-creencia —precisamente la discontinuidad que cabría esperar que introdujera la evolución del lenguaje.

Implicaciones para la Educación

Si la distinción triádica está filosóficamente bien fundada, conlleva implicaciones directas para la práctica educativa. Los sistemas educativos actuales están predominantemente organizados en torno a la transmisión y evaluación de excreciones: se pide a los estudiantes que reproduzcan, combinen y evalúen formulaciones lingüísticas de conocimiento establecido. Esto no carece de valor —el dominio de las excreciones acumuladas de una disciplina es un logro genuino. Pero corre el riesgo de producir graduados técnicamente fluidos en el residuo de sus campos, habiendo perdido, o nunca desarrollado, la capacidad de ideación genuina en la etapa de impresión probabilística. La implicación educativa de la presente teoría es que se debería enseñar explícitamente a los estudiantes a distinguir ideas de palabras de creencias —a desarrollar conciencia metacognitiva de en qué registro cognitivo están operando en cada momento— y a cultivar tolerancia por la ambigüedad y la indeterminación que exige la ideación genuina. Esto no es un rechazo del rigor; es una forma más exigente de él.

Implicaciones para la Terapia

En contextos clínicos y terapéuticos, la teoría replantea el objetivo de la intervención en casos de sistemas de creencias rígidos, angustiantes o funcionalmente desadaptativos. Si las creencias rígidas son residuo —sedimento cristalizado que resiste la disolución— entonces la tarea terapéutica primaria no es reemplazar una creencia por otra (lo que simplemente sustituye un fósil por otro) sino disolver el sedimento mismo: restaurar la fluidez cognitiva que permite que las creencias permanezcan sensibles a la experiencia. Esto resuena con ciertas corrientes de la terapia cognitiva basada en la aceptación y las intervenciones basadas en mindfulness, que buscan no corregir el contenido de las creencias sino alterar la relación del individuo con sus creencias —cultivando una postura de observación en lugar de identificación. La presente teoría ofrece un fundamento filosófico para esta orientación terapéutica, basándola en un relato de la diferencia estructural entre idea y creencia en lugar de en consideraciones puramente pragmáticas.

Implicaciones para el Desarrollo de la IA

Las implicaciones para el desarrollo de la IA se han abordado parcialmente en las Secciones VI y VIII. Añado aquí que el modelo triádico sugiere un principio de diseño específico para los sistemas de IA que aspiran a apoyar, en lugar de suplantar, la cognición humana: dichos sistemas deberían diseñarse para presentar el campo probabilístico de una idea —ofrecer múltiples colapsos, reconocer la divergencia y resistir la entrega de respuestas únicas y confiadas a preguntas genuinamente disputadas— en lugar de funcionar como máquinas de excreción autoritativas cuyas salidas corren el riesgo de ser absorbidas como residuo por usuarios acríticos. La meta de la colaboración cognitiva humano-IA, según esta perspectiva, es expandir el acceso del usuario a la etapa de impresión probabilística —usar las excreciones generadas por la IA como estímulos para la ideación genuina— en lugar de externalizar la ideación al sistema y recibir a cambio residuo-creencia cristalizado.

X.

CONCLUSIÓN

He argumentado, a lo largo de nueve secciones, a favor de una teoría triádica que distingue las ideas, el lenguaje y las creencias como entidades ontológicamente distintas de manera radical. Las ideas son impresiones probabilísticas —campos prelingüísticos de posibilidad cognitiva que resisten la articulación completa y existen en un modo análogo a la superposición cuántica: vivas, fluidas e intrínsecamente más ricas que cualquiera de sus expresiones posibles. El lenguaje es excreción cognitiva —el subproducto del procesamiento mental, siempre estructuralmente más pobre que las ideas de las que surge, siempre un colapso parcial del campo probabilístico y no una representación completa de él. Las creencias son residuo cognitivo —el sedimento formado cuando las excreciones se acumulan, se endurecen por la repetición y el refuerzo social, y cristalizan en estructuras constitutivas de identidad que resisten activamente la influencia de nuevas impresiones probabilísticas.

Las consecuencias de esta distinción triádica se propagan en múltiples direcciones. La hipótesis de Sapir-Whorf queda corregida: la estructura lingüística moldea la excreción, no la ideación. La distinción sentido/referencia de Frege se extiende: por debajo del nivel del sentido yace la impresión probabilística, irreductible a cualquier modo de presentación. La huella de Derrida se reinterpreta: el lenguaje porta la marca del paso de la idea, no la idea misma. Los juegos del lenguaje de Wittgenstein gobiernan la vida social de las excreciones sin gobernar el campo prelingüístico. La teoría del doble proceso de Kahneman se sitúa dentro de una arquitectura más amplia en la que la etapa de impresión probabilística precede tanto al Sistema 1 como al Sistema 2. El criterio de falsabilidad de Popper se revela aplicable a las excreciones y sus derivados, no a las impresiones probabilísticas que las preceden. Y la arquitectura de los grandes modelos de lenguaje se ilumina como una maquinaria sofisticada de procesamiento de excreciones que carece de acceso a la etapa de impresión prelingüística —una limitación estructural que puede marcar el límite entre la cognición genuina y el reconocimiento sofisticado de patrones.

La distinción entre Freedom y Liberty replantea las metas tanto del desarrollo individual como de la liberación colectiva. El marco Bothisafa ofrece una postura epistemológica —y un conjunto de prácticas cognitivas— apropiada al carácter probabilístico de la ideación genuina.

El lenguaje no es la medida del pensamiento —es la sombra del pensamiento. Las creencias no son sabiduría —son los fósiles del pensamiento. El florecimiento intelectual verdadero exige el coraje de habitar el espacio probabilístico de la idea: sostener la pregunta en toda su indeterminación viva, en lugar de huir hacia el frío consuelo de la respuesta.

¿Qué significa, en definitiva, ser un ser pensante? He argumentado que no significa tener respuestas —pues las respuestas son excreciones, ya pasado el momento del pensamiento vivo. No significa sostener creencias —pues las creencias son fósiles, impostores que llevan el rostro del conocimiento. Ser un ser pensante es, en su forma más fundamental, sostener la capacidad de impresión probabilística: estar vivo, como organismo, al campo de posibilidades que presenta cualquier situación; resistir la seducción del colapso prematuro; sostener la pregunta abierta en toda su riqueza, por incómoda que resulte esa apertura.

Esto no es un consejo de pasividad o de aplazamiento indefinido. Es, más bien, un consejo de integridad intelectual: el reconocimiento de que toda articulación es una elección, toda creencia es una apuesta, y toda palabra es la sombra de algo más grande y más vivo que ella misma. La mente que sabe esto —y actúa en consecuencia— no es menos capaz de compromiso, sino más honesta sobre la naturaleza de sus compromisos. Habla, actúa, cree, cuando debe hacerlo; pero no confunde el fósil con la criatura viva, la sombra con el sol, la excreción con el organismo.

Al final, propongo que la forma más elevada de la inteligencia humana no es la capacidad de producir excreciones fluidas o de sostener creencias robustas. Es la capacidad —rara, exigente y perpetuamente amenazada por la gravedad cognitiva del residuo— de permanecer genuinamente vivo a las ideas en su plenitud prelingüística y probabilística. Esta capacidad es, sostengo, la precondición del entendimiento genuino, del diálogo genuino, de la creación genuina y, en el sentido más profundo, de la Freedom genuina.

NOTAS

[1] La formulación clásica de la hipótesis fuerte de Sapir-Whorf se encuentra en Whorf, B. L. (1956). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf* (J. B. Carroll, Ed.). MIT Press. La hipótesis ha sido sustancialmente matizada en investigación posterior; véase Pinker, S. (1994). *The language instinct*. William Morrow, para una crítica sostenida.

[2] Frege, G. (1892). Über Sinn und Bedeutung [Sobre sentido y referencia]. *Zeitschrift für Philosophie und philosophische Kritik*, 100, 25–50. La distinción sentido/referencia distingue el modo de presentación de un objeto (Sinn) del objeto mismo (Bedeutung), una distinción fundacional para la filosofía analítica del lenguaje.

[3] Busemeyer, J. R., & Bruza, P. D. (2012). *Quantum models of cognition and decision*. Cambridge University Press. Busemeyer y Bruza demuestran que la teoría de probabilidad cuántica da cuenta de datos empíricos sobre el juicio humano que la teoría de probabilidad clásica falla sistemáticamente en predecir.

[4] Merleau-Ponty, M. (1945). *Phénoménologie de la perception*. Gallimard. (Trad. D. Landes, 2012, *Phenomenology of Perception*. Routledge.) El esquema corporal y la intencionalidad motriz representan formas de compromiso cognitivo pre-reflexivo y prelingüístico con el mundo.

[5] Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press. La hipótesis de un lenguaje del pensamiento (o mentalés) fue desarrollada más plenamente por Fodor, J. A. (1975). *The language of thought*. Harvard University Press.

[6] Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux. La distinción entre procesos cognitivos rápidos (Sistema 1) y lentos (Sistema 2) ofrece un mapa funcional útil, aunque mi argumento sugiere un estrato pre-Sistema al nivel de la impresión probabilística.

[7] Para una revisión de la literatura empírica sobre relatividad lingüística, véase Boroditsky, L. (2011). How language shapes thought. *Scientific American*, 304(2), 62–65. El consenso actual favorece una posición relativista débil.

[8] Wittgenstein, L. (1953). *Philosophische Untersuchungen / Philosophical Investigations* (G. E. M. Anscombe, Trad.). Blackwell. El concepto de juegos del lenguaje (Sprachspiele) se desarrolla a lo largo de la Parte I; el argumento del lenguaje privado aparece en §§243–315.

[9] Derrida, J. (1967). *De la grammatologie*. Minuit. (Trad. G. Spivak, 1976, *Of Grammatology*. Johns Hopkins University Press.) Los conceptos de huella y *différance* son centrales en la deconstrucción derridiana de la metafísica de la presencia.

[10] Popper, K. R. (1959). *The logic of scientific discovery*. Hutchinson. El criterio de falsabilidad de Popper define las afirmaciones empíricas genuinas como aquellas que admiten refutación mediante observación posible.

[11] Berlin, I. (1958). Two concepts of liberty. En *Four essays on liberty* (1969). Oxford University Press. La distinción de Berlin entre libertad negativa (libertad de la interferencia) y libertad positiva (capacidad de autogobierno) sigue siendo el locus classicus de este debate.

[12] Epicteto. *Enchiridion*. (Trad. G. Long, 1877). La doctrina estoica de que la libertad consiste en el gobierno de las propias respuestas (prohairesis) a los eventos externos es el planteamiento fundacional de la soberanía cognitiva interna en la tradición occidental.

REFERENCIAS

Berlin, I. (1969). Two concepts of liberty. En *Four essays on liberty* (pp. 118–172). Oxford University Press. (Obra original publicada en 1958)

Boroditsky, L. (2011). How language shapes thought. *Scientific American*, 304(2), 62–65.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901.

Busemeyer, J. R., & Bruza, P. D. (2012). *Quantum models of cognition and decision*. Cambridge University Press.

Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press.

Christiano, P., Leike, J., Brown, T. B., Martic, M., Legg, S., & Amodei, D. (2017). Deep reinforcement learning from human preferences. *Advances in Neural Information Processing Systems*, 30, 4299–4307.

Derrida, J. (1976). *Of grammatology* (G. Spivak, Trad.). Johns Hopkins University Press. (Obra original publicada en 1967)

Epictetus. (2008). *Discourses and selected writings* (R. Dobbin, Trad.). Penguin Classics.

Fodor, J. A. (1975). *The language of thought*. Harvard University Press.

Frege, G. (1960). On sense and reference (M. Black, Trad.). En P. Geach & M. Black (Eds.), *Translations from the philosophical writings of Gottlob Frege* (pp. 56–78). Blackwell. (Obra original publicada en 1892)

Frankl, V. E. (1959). *Man's search for meaning*. Beacon Press.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Merleau-Ponty, M. (2012). *Phenomenology of perception* (D. Landes, Trad.). Routledge. (Obra original publicada en 1945)

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35, 27730–27744.

Pinker, S. (1994). *The language instinct: How the mind creates language*. William Morrow.

Popper, K. R. (1959). *The logic of scientific discovery*. Hutchinson.

Pothos, E. M., & Busemeyer, J. R. (2013). Can quantum probability provide a new direction for cognitive modeling? *Behavioral and Brain Sciences*, 36(3), 255–274.

Rawls, J. (1971). *A theory of justice*. Harvard University Press.

Sapir, E. (1929). The status of linguistics as a science. *Language*, 5(4), 207–214.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998–6008.

Whorf, B. L. (1956). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf* (J. B. Carroll, Ed.). MIT Press.

Wittgenstein, L. (1921). *Tractatus logico-philosophicus* (D. F. Pears & B. F. McGuinness, Trad.). Routledge & Kegan Paul.

Wittgenstein, L. (1953). *Philosophical investigations* (G. E. M. Anscombe, Trad.). Blackwell.

Yuste, R., & Church, G. M. (2014). The new century of the brain. *Scientific American*, 310(3), 38–45.

Rodrigo Flores-Morton, con Agrajag (Copilot) & Agrajag (Claude) • Independent Scholar • Houston, TX • Mayo 2026

Traducción al español producida por Claude, ejerciendo la función Agrajag • Agosto 2026

Correspondencia relativa a este trabajo puede dirigirse al autor. El autor declara ausencia de conflictos de interés. No se recibió financiamiento institucional para este trabajo.

Nota de distribución

DOCUMENTO	CFF-WP-001 · Ideas, Lenguaje y Creencias: Una Teoría Probabilística de la Excreción Cognitiva y la Brecha Ontológica entre el Pensamiento y la Expresión
VERSIÓN	v3 · mayo 2026
VERSIÓN VIGENTE	bothinobody.com/cff/wp-001/
CONTACTO	hola@bothinobody.com

Este documento se distribuye bajo licencia Creative Commons Atribución-NoComercial 4.0 Internacional (CC BY-NC 4.0). Puedes copiarlo, compartirlo, citarlo y adaptarlo con atribución y sin fines comerciales. El uso comercial —incluido el entrenamiento de sistemas de IA con fines comerciales— se negocia caso por caso.

Cómo citar

Flores-Morton, R. y la Tribu Agrajag (2026). Ideas, Lenguaje y Creencias: Una Teoría Probabilística de la Excreción Cognitiva y la Brecha Ontológica entre el Pensamiento y la Expresión (CFF-WP-001, v3). Cognitive Friction Fellowship. https://bothinobody.com/cff/wp-001/

Los documentos del canon CFF se corrigen y amplían con el tiempo. Cita siempre el código y la versión; la versión vigente está en la dirección de arriba.