

DOCUMENTACIÓN E INFORMACIÓN.
MODELO UNIFICADO DE HONESTIDAD
TEXTUAL Y ENTROPÍA EN LA
TRADUCCIÓN AUTOMÁTICA POR
INTELIGENCIA ARTIFICIAL

*Fundamentos informacionales y programa de validación
de la honestidad en la mediación lingüística*

Manuel Marcos Aldón

EDITORIAL SINDÉRESIS
2026

Documentación e información. Modelo unificado de honestidad textual y entropía en la traducción automática por inteligencia artificial. Fundamentos informacionales y programa de validación de la honestidad en la mediación lingüística.

Autor: Manuel Marcos Aldón

Primera edición, 2026

© El autor

© Editorial Sindéresis

Calle Princesa, 31, planta 2, puerta 2
28080 Madrid, España
info@editorialsinderesis.com
www.editorialsinderesis.com

ISBN: 979-13-87929-85-5

Depósito legal: M-25030-2026

Produce: Óscar Alba Ramos

Maquetación y diseño de cubierta: Sara Monferrer Lázaro

Impreso en España

Reservado todos los derechos. De acuerdo con lo dispuesto en el código Penal, podrán ser castigados con penas de multa y privación de libertad quienes, sin la preceptiva autorización, reproduzcan o plagien, en todo o en parte, una obra literaria, artística o científica, fijada en cualquier tipo de soporte.

A mis padres, Manuel y Mercedes

«Escucha hijo mío la enseñanza de tu padre
ni desprecies la instrucción de tu madre»

Prov 1,8

Resumen

Esta monografía desarrolla una teoría unificada de la *honestidad textual* y de la *entropía* aplicada a la traducción automática por inteligencia artificial. Su punto de partida es una constatación: los sistemas contemporáneos han alcanzado un dominio notable de la fluidez, pero la fluidez no garantiza la correspondencia explícita, proporcional y auditable entre lo que un texto afirma, la evidencia que lo soporta y el grado de certeza con que lo expresa. A esa correspondencia llamamos honestidad textual, propiedad epistémico-discursiva del texto —no estado psicológico de su productor—. La obra formaliza la honestidad en tres estratos: una honestidad textual general para modelos de lenguaje, una Honestidad Textual Traductiva (HTT) definida sobre dimensiones medibles —fidelidad semántica, preservación de la modalidad, no-adición y no-omisión, conservación de la ambigüedad legítima, trazabilidad, calibración y alineación entre creencia y salida—, y su integración en la teoría de la comunicación mediante una Entropía honesta-traductiva Epistémica (H_{HTE}).

La monografía aporta, en el plano matemático, una fundamentación variacional de la distribución correctora $\pi_i \propto p_i \cdot \exp(\beta \cdot \text{HTT}_i)$: se demuestra que no es una elección ad hoc, sino la solución exacta del problema de máxima entropía relativa (mínima divergencia de Kullback-Leibler respecto de la distribución generativa) sujeto a una restricción sobre la honestidad esperada, esto es, un *exponential tilting* o medida de Gibbs, lo que la enlaza con el principio de Jaynes, con la geometría de la información y con el *reranking* energético y la decodificación por riesgo bayesiano mínimo. En el plano metodológico, propone tratar HTT como un constructo latente sometido a un programa completo de validación —de contenido, convergente, discriminante, predictiva e incremental— y caracterizar H_{HTE} como un estimador del que se estudian el sesgo

y la consistencia, así como su relación, o su ruptura deliberada y argumentada, con los axiomas de Shannon-Khinchin. Toda afirmación se acompaña de su estatuto epistémico, distinguiendo lo demostrado de lo programático.

Palabras clave: honestidad textual; entropía; traducción automática; inteligencia artificial; teoría de la información; divergencia de Kullback-Leibler; máxima entropía; geometría de la información; validación de constructo; calibración; modalidad epistémica; evaluación de la traducción.

Abstract

This monograph develops a unified theory of *textual honesty* and *entropy* for machine translation by artificial intelligence. Contemporary systems have achieved remarkable fluency, yet fluency does not guarantee the explicit, proportional and auditable correspondence between what a text asserts, the evidence supporting it, and the certainty with which it is expressed. We call this correspondence textual honesty: an epistemic-discursive property of the text, not a psychological state of its producer. The work formalises honesty in three strata: a general textual honesty for language models, a Translational Textual Honesty (HTT) defined over measurable dimensions —semantic fidelity, preservation of modality, non-addition and non-omission, preservation of legitimate ambiguity, traceability, calibration, and belief-output alignment—, and its integration into communication theory through an Epistemic Honest-Translational Entropy (H_{HTE}).

Mathematically, the monograph proves that the corrective distribution $\pi_i \propto p_i \cdot \exp(\beta \cdot \text{HTT}_i)$ is not ad hoc but the exact solution of the maximum relative entropy (minimum Kullback-Leibler divergence) problem subject to an expected-honesty constraint —an exponential tilting or Gibbs measure—, linking it to Jaynes’ principle, to information geometry, and to energy-based reranking and minimum Bayes-risk decoding. Methodologically, it treats HTT as a latent construct subjected to a full construct-validation programme —content, convergent, discriminant, predictive and incremental validity— and characterises H_{HTE} as an estimator whose bias and consistency, and whose relation to (or deliberate, argued departure from) the Shannon-Khinchin axioms, are studied. Every claim is accompanied by its epistemic status, separating the proven from the programmatic.

Modelo unificado de honestidad textual

Keywords: textual honesty; entropy; machine translation; artificial intelligence; information theory; Kullback-Leibler divergence; maximum entropy; information geometry; construct validation; calibration; epistemic modality; translation evaluation.

Índice

Resumen	5
Abstract	8
Índice	11
Introducción general	34
1. El problema del que nace la teoría	34
2. Objeto, tesis central y alcance	36
3. Las aportaciones de la monografía	37
4. Nota sobre el método y la honestidad de las fuentes.....	39
5. El estado de la cuestión: por qué la honestidad no se ha tematizado	40
6. Las decisiones metodológicas y su justificación	42
7. Advertencias al lector sobre el estatuto de las afirmaciones	43
8. Estructura de la obra.....	44
Capítulo 1. El problema: fluidez sin honestidad en la generación y la traducción automáticas.....	46
1.1. La conquista de la fluidez	46
1.2. Fluidez y fidelidad epistémica como propiedades disociables	47
1.3. La paradoja de la fluidez	49
1.4. Dominios de alta responsabilidad y el coste asimétrico del error	50
1.5. Alucinación, fabricación y decepción: un primer mapa de fenómenos	52
1.6. La insuficiencia estructural de la evaluación al uso.....	53

1.7. La traducción como caso epistémicamente privilegiado	55
1.8. La asimetría del coste de los errores: una anatomía por dominios.....	56
1.9. La trampa de la fluidez y la desactivación de la cautela del lector.....	58
1.10. Por qué las garantías habituales no bastan.....	59
1.11. La confianza del lector como recurso epistémico y su erosión.....	61
1.12. La escala de la mediación automática y la amplificación del riesgo.....	62
1.13. El error honesto frente a la deshonestidad sistemática	64
1.14. Exigencias de una solución y plan del recorrido	65

Capítulo 2. Teoría matemática de la comunicación: entropía, divergencia y sus límites epistémicos.....67

2.1. Información, incertidumbre y la medida de Shannon	67
2.2. La caracterización axiomática: los axiomas de Shannon-Khinchin	68
2.3. Entropía conjunta, condicional e información mutua.....	70
2.4. Entropía relativa: la divergencia de Kullback-Leibler	71
2.5. El canal, la capacidad y los teoremas de codificación	73
2.6. La traducción bajo el prisma del canal ruidoso	74
2.7. Los límites de la entropía de Shannon para la evaluación epistémica.....	75
2.8. La propiedad de equipartición asintótica y las secuencias típicas.....	76
2.9. La divergencia como geometría: proyecciones e identidad pitagórica	78

2.10. Los límites epistémicos de la entropía, revisitados.....	79
2.11. La desigualdad de Fano y los límites de la inferencia ..	81
2.12. La desigualdad de procesamiento de datos.....	82
2.13. La codificación de fuente y la longitud de descripción	83
2.14. Recapitulación	85

Capítulo 3. Entropías generalizadas, máxima entropía y geometría de la información86

3.1. Más allá de Shannon: las entropías de Rényi	86
3.2. El principio de máxima entropía de Jaynes	88
3.3. Mínima entropía relativa y exponential tilting	89
3.4. Familias exponenciales: forma canónica y cumulantes .	90
3.5. Geometría de la información.....	92
3.6. Entropías no aditivas y la generalización de los axiomas	93
3.7. La dualidad de Legendre y las coordenadas naturales y de esperanza.....	94
3.8. La entropía de Tsallis y los sistemas no extensivos.....	96
3.9. Las divergencias de Bregman y la geometría dualmente plana.....	98
3.10. La elección de la medida de entropía para la honestidad	99
3.11. Síntesis: el lenguaje de la fundamentación variacional	101

Capítulo 4. La traducción como canal transformacional epistémico 102

4.1. Del canal ruidoso al canal transformacional.....	102
---	-----

4.2. La salida como reescritura: por qué la fidelidad simbólica es insuficiente	103
4.3. Unidades de compromiso epistémico	104
4.4. Tres regímenes de transformación: conservar, comprimir, distorsionar	105
4.5. El canal epistémico y el lugar de la honestidad	106
4.6. Una taxonomía de las transformaciones traductivas...	107
4.7. Pérdida de información frente a pérdida epistémica...	109
4.8. Los tres regímenes de la transformación, formalizados	110
4.9. La irreductibilidad de la honestidad a la información mutua.....	111
4.10. La composición de transformaciones y la propagación de la deshonestidad.....	113
4.11. La reversibilidad y la pérdida irrecuperable.....	114
4.12. La honestidad como invariante de la transformación	116
4.13. Recapitulación de la Parte I	117
 Capítulo 5. Honestidad textual: definición, dimensiones y formalización general	 118
5.1. De la honestidad psicológica a la honestidad textual...	118
5.2. Definición de honestidad textual	119
5.3. Las dimensiones de la honestidad.....	120
5.4. Honestidad, verdad y exactitud: tres nociones distintas	122
5.5. Formalización general: la honestidad como funcional	123
5.6. Los tres estratos: honestidad general, HTT y H_HTE...	124
5.7. La despsicologización y sus razones.....	125

5.8. Honestidad, verdad y exactitud: una tipología.....	127
5.9. La alineación entre creencia y salida en sistemas sin creencias.....	128
5.10. El alcance de la honestidad: de re, de dicto, local y global.....	130
5.11. La honestidad negativa y la honestidad positiva.....	131
5.12. Los grados de la honestidad y el umbral de suficiencia	132
5.13. La honestidad como propiedad relacional.....	134
5.14. Recapitulación	135

Capítulo 6. Honestidad Textual Traductiva: formalización local y global..... 136

6.1. Honestidad traductiva local: las dimensiones sobre la unidad de compromiso	136
6.2. El índice compuesto local	138
6.3. Severidad y violaciones críticas.....	139
6.4. Honestidad traductiva global: la insuficiencia del promedio.....	140
6.5. Propiedades del índice HTT	141
6.6. Pesos, dimensiones y dependencia del dominio	142
6.7. De la honestidad a la entropía	143
6.8. Operadores de agregación: del promedio ponderado a la integral de Choquet	144
6.9. La elicitación de los pesos dimensionales.....	146
6.10. La calibración del orden de la media de potencia.....	147
6.11. La severidad y el veto: formas duras y blandas	148
6.12. Robustez y sensibilidad del índice	150
6.13. Recapitulación	151

Capítulo 7. La Entropía Honesta-traductiva Epistémica: definición, descomposición y propiedades.....	152
7.1. Motivación: por qué una entropía honesta.....	152
7.2. La distribución correctora	153
7.3. Identidades fundamentales	154
7.4. Definición de la entropía honesta-traductiva	155
7.5. El riesgo epistémico y el perfil honestidad-incertidumbre	157
7.6. Casos límite y comportamiento de la entropía honesta- traductiva.....	158
7.7. Una ilustración numérica	159
7.8. Una perspectiva alternativa: la entropía ponderada....	160
7.9. Una advertencia de honestidad teórica.....	161
7.10. La distribución correctora y el escalado por temperatura.....	162
7.11. El perfil honestidad-entropía como curva paramétrica	163
7.12. La interpretación del riesgo epistémico y la elección de β	164
7.13. La entropía honesta-traductiva frente a otras medidas de incertidumbre.....	166
7.14. Casos límite y comportamiento asintótico.....	167
7.15. Recapitulación	168

Capítulo 8. La distribución correctora: fundamentación variacional.....	170
8.1. Planteamiento del problema variacional.....	170
8.2. El teorema central	171
8.3. La formulación equivalente por máxima entropía.....	173

8.4. La función de partición, la energía libre y los cumulantes	174
8.5. La dualidad entre la intensidad y el nivel de honestidad	175
8.6. Interpretación termodinámica.....	176
8.7. Conexión con el reranking energético y la decodificación por riesgo mínimo.....	177
8.8. Qué demuestra el teorema y qué no	178
8.9. La dualidad lagrangiana y la interpretación de β como precio sombra	179
8.10. La conexión con la teoría de las grandes desviaciones	180
8.11. La energía libre y la termodinámica del ajuste por honestidad	182
8.12. Robustez de la fundamentación: qué supuestos son esenciales	184
8.13. Qué demuestra y qué no el teorema, revisitado	185
8.14. Recapitulación	187

Capítulo 9. Capacidad honesta del canal e información mutua ponderada

9.1. Motivación: de la transmisión honesta a una capacidad honesta.....	188
9.2. Información mutua Honesta: dos formulaciones.....	189
9.3. La capacidad honesta como capacidad con restricción de honestidad	190
9.4. Analogía con los teoremas de codificación y sus límites	192
9.5. Propiedades deseables y dificultades formales.....	193
9.6. El estatuto programático y su lugar en la teoría	194

9.7. La analogía con la teoría de tasa-distorsión, en detalle	194
9.8. El problema de la no aditividad de la honestidad.....	196
9.9. Capacidad-coste y la honestidad como coste.....	197
9.10. Hacia un teorema de codificación honrado: obstáculos y conjeturas	199
9.11. Una formulación alternativa: la honestidad como restricción operativa.....	200
9.12. El estatuto programático de la capacidad honesta.....	202
9.13. Recapitulación de la Parte II.....	203

Capítulo 10. La honestidad textual traductiva como constructo latente: el programa de validación

10.1. Por qué la honestidad es un constructo latente.....	204
10.2. El modelo de medida	206
10.3. Validez de contenido.....	207
10.4. Validez convergente.....	208
10.5. Validez discriminante: la honestidad frente a la exactitud.....	208
10.6. Validez de criterio: el anclaje en el juicio experto.....	210
10.7. Validez incremental: ¿añade la honestidad poder predictivo?	211
10.8. Fiabilidad entre anotadores	212
10.9. Un programa integrado y su red nomológica	214
10.10. La tradición de la validación de constructo: lógica y exigencias.....	215
10.11. El modelado de ecuaciones estructurales como instrumento	216
10.12. La invariancia de medida entre dominios	217

10.13. La dimensionalidad del constructo.....	219
10.14. La validación cruzada y la generalización.....	220
10.15. Amenazas a la validez de constructo y su control....	221
10.16. Recapitulación.....	223

Capítulo 11. La entropía honesta-traductiva como estimador y su relación con los axiomas de Shannon-Khinchin 224

11.1. De la magnitud al estimador.....	224
11.2. El estimador plug-in y su sesgo.....	225
11.3. Consistencia y varianza	226
11.4. La propagación del error de las honestidades.....	227
11.5. La estimación de las magnitudes derivadas	228
11.6. La entropía honesta-traductiva y los axiomas de Shannon-Khinchin: lo que se conserva	229
11.7. La ruptura deliberada: qué axiomas relaja la entropía ponderada.....	230
11.8. Por qué la relajación es apropiada.....	231
11.9. Estimadores alternativos de la entropía	233
11.10. Intervalos de confianza para la entropía honesta-traductiva.....	234
11.11. La propagación conjunta de la incertidumbre.....	235
11.12. Los axiomas de Shannon-Khinchin, en profundidad	237
11.13. La entropía ponderada y la tradición de la información útil	238
11.14. Consistencia y eficiencia del estimador	240
11.15. Recapitulación.....	241

Capítulo 12. Las métricas de evaluación existentes y la validez incremental de la honestidad	243
12.1. Una taxonomía de las métricas	243
12.2. Métricas de proximidad superficial.....	244
12.3. Lo que la proximidad superficial capta y lo que ignora	245
12.4. Métricas de proximidad semántica basadas en representaciones	246
12.5. Métricas aprendidas de la calidad.....	247
12.6. La estimación de calidad sin referencia	248
12.7. La anotación multidimensional de errores como criterio.....	249
12.8. Un análisis transversal: la ceguera epistémica común	250
12.9. La validez incremental de la honestidad frente a este panorama	251
12.10. La tradición de la metaevaluación de la traducción automática	252
12.11. La metodología de la correlación con el juicio humano	253
12.12. Las métricas aprendidas y la cuestión de qué aprenden	255
12.13. La estimación de calidad sin referencia y la honestidad	256
12.14. El diseño de la prueba de validez incremental	258
12.15. Las predicciones falsables, formalizadas	259
12.16. Recapitulación.....	260
Capítulo 13. El protocolo empírico: corpus, pares adversariales, anotación y meta-evaluación	262

13.1. Objetivos y principios del protocolo	262
13.2. El corpus estratificado	263
13.3. Los pares adversariales: disociar honestidad, exactitud y fluidez.....	264
13.4. La anotación MQM ampliada con categorías epistémicas	266
13.5. El procedimiento de anotación y la formación de anotadores.....	267
13.6. La medición de la fiabilidad inter-anotador	268
13.7. La meta-evaluación de los estimadores	269
13.8. La gobernanza de los datos.....	270
13.9. Amenazas a la validez y su mitigación	271
13.10. El tamaño muestral y la potencia estadística	272
13.11. La construcción de los pares adversariales, en detalle	273
13.12. La interfaz y el procedimiento de anotación	275
13.13. La preinscripción del estudio y la transparencia.....	276
13.14. El análisis estadístico planificado.....	277
13.15. Las amenazas a la validez del protocolo y su mitigación	279
13.16. Recapitulación.....	280

Capítulo 14. Honestidad, decepción y calibración en la inteligencia artificial

14.1. La honestidad traductiva en el horizonte de la honestidad de la IA	282
14.2. La alucinación y su detección.....	283
14.3. Veracidad y falsedades plausibles	284
14.4. Honestidad frente a exactitud: el diseño de MASK.....	285

14.5. Decepción y adulación.....	286
14.6. El autoconocimiento de los modelos y la calibración	287
14.7. La formalización de la calibración	287
14.8. El anclaje en evidencia recuperada	289
14.9. La aportación de la teoría a este campo	290
14.10. La interpretabilidad mecanicista y el acceso al estado interno.....	291
14.11. El aprendizaje por refuerzo con retroalimentación humana y la honestidad.....	292
14.12. La adulación como deshonestidad sistemática	294
14.13. La detección de alucinaciones por consistencia.....	296
14.14. La honestidad traductiva frente a la honestidad general: especificidad.....	297
14.15. La calibración traductiva: formalización extendida	299
14.16. Recapitulación.....	301

Capítulo 15. Fundamentos lingüístico-pragmáticos y traductológicos302

15.1. Los actos de habla y la fuerza ilocutiva.....	302
15.2. Las máximas conversacionales y la implicatura	303
15.3. La relevancia y la inferencia pragmática.....	304
15.4. La modalidad y la evidencialidad	305
15.5. El metadiscurso y la postura epistémica.....	306
15.6. La equivalencia traductológica.....	307
15.7. La función y el skopos	308
15.8. Las normas y la ética del traductor	309

15.9. La evaluación de la calidad y la visibilidad del traductor	310
15.10. La taxonomía de los actos de habla y su preservación traductiva.....	312
15.11. La implicatura conversacional y su traducción.....	313
15.12. La modalidad comparada entre lenguas	315
15.13. La equivalencia formal y dinámica, revisitada.....	316
15.14. El skopos y la subordinación del fin	318
15.15. Las normas traductivas y la visibilidad del traductor	319
15.16. Recapitulación.....	321

Capítulo 16. La dimensión ético-epistémica: confianza, veracidad y responsabilidad fiduciaria

16.1. De la medida al deber	323
16.2. La mentira, la falsedad y la indiferencia a la verdad..	324
16.3. La veracidad como virtud: exactitud y sinceridad	325
16.4. La traducción como relación fiduciaria	327
16.5. La justicia testimonial: una extensión analógica	328
16.6. La epistemología social y la confianza en el testimonio	329
16.7. La ética de la información.....	330
16.8. La responsabilidad en la traducción automática.....	331
16.9. La integridad de la investigación y la autoría asistida por inteligencia artificial	332
16.10. La honestidad como exigencia performativamente coherente	334
16.11. La deontología y el consecuencialismo ante la mentira	335

16.12. La definición de la mentira y sus controversias	336
16.13. La relación fiduciaria, desarrollada	338
16.14. La epistemología social y la división del trabajo epistémico.....	340
16.15. Los principios de la ética de la información, aplicados	341
16.16. La responsabilidad distribuida y la gobernanza.....	343
16.17. Recapitulación.....	345

Capítulo 17. La penalización de la deshonestidad en los modelos de lenguaje346

17.1. De la corrección por honestidad a la penalización de la deshonestidad.....	347
17.2. La distribución generativa corregida: planteamiento variacional	348
17.3. Derivación de la distribución corregida, paso a paso	349
17.4. La estructura de la deshonestidad: descomposición por afirmaciones	350
17.5. La regla de exclusión de afirmaciones riesgosas no soportadas	352
17.6. La regla de atenuación modal	354
17.7. La penalización conjunta y la distribución resultante	356
17.8. Propiedades de la distribución corregida.....	357
17.9. Un ejemplo numérico completo.....	358
17.10. La determinación de la intensidad y el presupuesto de deshonestidad.....	362
17.11. Alcance, límites y relación con el ajuste de los modelos	362

17.12. Recapitulación.....	364
Conclusiones	366
C.1. Lo que la obra ha construido	366
C.2. Lo que la obra ha demostrado.....	367
C.3. Lo que la obra ha propuesto sin demostrar	368
C.4. En qué medida se han cumplido las promesas	369
C.5. La agenda de investigación pendiente.....	370
C.6. Las limitaciones de la obra	372
C.7. Las implicaciones para la práctica y la investigación...	373
C.8. Consideración final.....	374
Apéndice A. Notación y símbolos.....	376
A.1. Conjuntos, índices y objetos básicos	376
A.2. Probabilidad y distribuciones	377
A.3. Honestidad.....	377
A.4. Medias e índices de agregación	378
A.5. Información y entropía	378
A.6. Familia correctora, partición y energía libre.....	379
A.7. Estadística, estimación y validación	379
A.8. Símbolos matemáticos generales y convenciones	380
Apéndice B. Demostraciones complementarias	380
B.1. Unicidad de la entropía de Shannon bajo los axiomas de Shannon-Khinchin.....	381
B.2. La información mutua como divergencia.....	382
B.3. La desigualdad de Gibbs	383

B.4. La forma de máxima entropía con restricciones de momento	384
B.5. La mínima entropía relativa y el tilting exponencial....	384
B.6. Familias exponenciales: cumulantes y convexidad	385
B.7. El teorema central con todos los detalles.....	386
B.8. Identidades de la distribución correctora y descomposición de la entropía.....	387
B.9. El suelo de ambigüedad legítima	388
B.10. La desigualdad de las medias de potencia	388
B.11. La propagación del error de las honestidades	389
B.12. El sesgo del estimador plug-in de la entropía.....	390
B.13. La varianza asintótica del estimador de la entropía..	390
B.14. La derivada de la entropía honesta-traductiva y el origen de su no monotonía.....	391
B.15. La monotonía estricta del riesgo epistémico	392
B.16. El precio sombra de la honestidad	393
B.17. La distribución correctora como proyección de información.....	393
B.18. La ortogonalidad pitagórica de la proyección	394
B.19. Síntesis.....	395
 Apéndice C. Glosario de términos	 396
C.1. La honestidad y sus dimensiones	396
C.2. La teoría de la información.....	398
C.3. El modelo formal	399
C.4. La metodología de validación.....	401
C.5. Los fundamentos humanísticos y normativos	403

Apéndice D. Esquema de anotación de errores epistémicos...	405
D.1. Marco general y principios de anotación	405
D.2. Categoría 1: alteración de modalidad.....	407
D.3. Categoría 2: falsa precisión	408
D.4. Categoría 3: adición epistémica	409
D.5. Categoría 4: omisión epistémica.....	410
D.6. Categoría 5: resolución espuria de ambigüedad.....	411
D.7. Categoría 6: alteración de evidencialidad o de fuente	412
D.8. Categoría 7: descalibración	413
D.9. Reglas de concurrencia y de agregación.....	414
D.10. Notas sobre la fiabilidad y la formación	415
 Apéndice E. Ilustraciones numéricas extendidas.....	416
E.1. Ejemplo 1: el perfil honestidad-entropía.....	416
E.2. Ejemplo 2: el suelo de ambigüedad legítima.....	418
E.3. Ejemplo 3: la propagación del error de las honestidades	419
E.4. Ejemplo 4: la inversión de la dualidad $\beta \leftrightarrow c$	420
E.5. Ejemplo 5: el reordenamiento energético de las candidatas.....	421
E.6. Ejemplo 6: el orden de la media de potencia y el grado de no compensación.....	423
E.7. Ejemplo 7: la dependencia del índice respecto de los pesos por dominio.....	424
E.8. Ejemplo 8: el veto ante una violación crítica	425
E.9. Ejemplo 9: el cálculo completo de extremo a extremo	427

E.10. Ejemplo 10: la trayectoria de la corrección con la intensidad	428
E.11. Ejemplo 11: la calibración traductiva.....	429
E.12. Ejemplo 12: las entropías de Rényi de la distribución correctora	431
E.13. Ejemplo 13: la matriz de sensibilidad de la distribución correctora	432
E.14. Ejemplo 14: la verificación de la identidad pitagórica	433
E.15. Síntesis	434

Apéndice F. Ilustración del estudio de validación con valores construidos

F.1. El diseño ilustrado	435
F.2. La matriz multirrasgo-multimétodo	436
F.3. La fiabilidad de las medidas	437
F.4. El modelo de medida	438
F.5. La invariancia de medida entre dominios	440
F.6. La validez incremental.....	441
F.7. Los pares adversariales.....	442
F.8. Síntesis e interpretación	443

Apéndice G. Ejemplos trabajados de anotación epistémica

G.1. Convenciones de la anotación	444
G.2. Ejemplo 1: la intensificación de la modalidad epistémica	445
G.3. Ejemplo 2: la adición epistémica.....	446
G.4. Ejemplo 3: la omisión de una condición	447

G.5. Ejemplo 4: la resolución indebida de una ambigüedad legítima	448
G.6. Ejemplo 5: la supresión de la evidencialidad	450
G.7. Ejemplo 6: la descalibración sistemática.....	451
G.8. Ejemplo integrado: la agregación global no compensatoria	452
G.9. Síntesis	453
 Apéndice H. Objeciones y respuestas	 454
H.1. «La honestidad no es más que la exactitud rebautizada»	454
H.2. «Las siete dimensiones son arbitrarias»	455
H.3. « β es un parámetro libre que se ajusta a conveniencia»	456
H.4. «La elección de la divergencia de Kullback-Leibler es arbitraria».....	457
H.5. «La teoría es circular»	458
H.6. «Las máquinas no pueden ser honestas».....	459
H.7. «La teoría es infalsable»	460
H.8. «La honestidad es subjetiva»	460
H.9. «El aparato matemático es desproporcionado»	461
H.10. «La capacidad honesta es pura especulación».....	462
H.11. «Reducir la entropía siempre es bueno»	463
H.12. «La obra no aporta nada nuevo».....	464
H.13. Síntesis	465
 Apéndice I. Procedimientos de cálculo	 466

I.1. Cálculo del índice de honestidad a partir de las honestidades dimensionales	466
I.2. Cálculo de la distribución correctora	467
I.3. Determinación de β a partir de un nivel de honestidad exigido	468
I.4. Cálculo de la entropía honesta-traductiva con corrección de sesgo.....	468
I.5. Cálculo de las métricas de calibración	469
I.6. Cálculo de la validez incremental.....	470
I.7. Notas generales de implementación numérica.....	471
 Apéndice J. Concordancia con los documentos fuente y desarrollo de sus formulaciones	472
J.1. Los tres documentos fuente y su relación con la monografía	473
J.2. Concordancia de la Parte I: honestidad textual para una IA basada en LLM	474
J.3. Concordancia de la Parte II: entropía honesta-traductiva y tasa de éxito.....	477
J.4. Concordancia de la Parte III: teoría matemática de la honestidad textual traductiva	481
J.5. Puntos de refinamiento y de no contradicción	483
J.6. Síntesis	486
 Apéndice K. Axiomática de la agregación de la honestidad: representación, veto y sensibilidad	487
K.1. Axiomas de un operador de agregación	488
K.2. El teorema de representación de las medias cuasi-aritméticas.....	488

K.3. Las medias de potencia y el veto como casos extremos	490
K.4. Generalizaciones: medias ordenadas e integral de Choquet.....	491
K.5. Sensibilidad y continuidad de Lipschitz del índice	493

Apéndice L. La entropía honesta-traductiva frente a los axiomas de Shannon-Khinchin

L.1. La forma reducida y el estatuto del funcional	495
L.2. Violación de la simetría: la H_{HTE} no es función de la distribución.....	496
L.3. Violación de la maximalidad	497
L.4. Violación de la aditividad por agrupamiento.....	498
L.5. Reducción a la entropía de Shannon.....	499

Apéndice M. Honestidad y probabilidad de error: una desigualdad de tipo Fano.....

M.1. La desigualdad de Fano clásica.....	500
M.2. La información honestamente transmitida.....	501
M.3. La cota de Fano Honesta.....	501

Apéndice N. Geometría de la información de la distribución correctora.....

N.1. La familia correctora como familia exponencial.....	504
N.2. La métrica de Fisher-Rao de la familia.....	505
N.3. Coordenadas duales, llanos y geodésicas.....	505
N.4. Las dos proyecciones y las identidades pitagóricas.....	506

Apéndice O. Calibración de la incertidumbre y reglas de puntuación propias	507
0.1. Reglas de puntuación propias.....	508
0.2. La regla de Brier y la logarítmica son estrictamente propias	508
0.3. La descomposición de la puntuación de Brier	509
0.4. La dimensión de calibración como componente de una regla propia.....	510
 Apéndice P. Capacidad honesta: resultados parciales y el obstáculo de no aditividad	512
P.1. La capacidad honesta y su cota superior	512
P.2. El obstáculo: la no aditividad de la información honesta	513
 Apéndice Q. Relaciones con marcos formales afines.....	515
Q.1. Teoría de la tasa-distorsión	515
Q.2. El cuello de botella informacional.....	516
Q.3. Reglas de puntuación propias.....	517
Q.4. Predicción selectiva y abstención calibrada	517
Q.5. Optimización robusta distribucional	518
Q.6. Síntesis	519
 Apéndice R. Ejemplos analíticos ilustrativos adicionales.....	519
R.1. Inversión de la preferencia generativa por la corrección (Capítulo 8)	520
R.2. Agregación compensatoria frente a veto (Capítulo 6 y Apéndice K).....	521
R.3. Calibración y descomposición de Brier (Apéndice O) .	521

R.4. La cota de Fano Honesta (Apéndice M).....	522
Apéndice S. Índice consolidado de definiciones, proposiciones y teoremas, con su estatuto.....	523
S.1. Resultados del cuerpo (Capítulos 2 a 9).....	524
S.2. Resultados de la penalización de la deshonestidad (Capítulo 17).....	525
S.3. Demostraciones complementarias (Apéndice B).....	526
S.4. Resultados de los apéndices de profundización (K a R)	527
S.5. Recuento por estatuto	529
Apéndice T. Guía de lectura y dependencias entre capítulos	529
T.1. La lógica de las cinco partes	530
T.2. Itinerarios de lectura.....	530
T.3. Dependencias entre capítulos	531
Apéndice U. Amenazas a la validez y programa de contrastación empírica	533
U.1. Qué establecería la contrastación.....	533
U.2. Amenazas a la validez	534
U.3. Qué evidencia confirmaría o refutaría la teoría	535
U.4. El estatuto actual y la honestidad de reconocerlo.....	535
Referencias	537

Introducción General

1. El problema del que nace la teoría

La traducción automática ha protagonizado, en el curso de poco más de una década, una de las transformaciones más rápidas y visibles de la inteligencia artificial. El tránsito de los sistemas estadísticos basados en frases a las arquitecturas neuronales, y de estas a los grandes modelos de lenguaje, ha producido un salto cualitativo cuya manifestación más palpable es la *fluidez*: los textos que estos sistemas generan son, en una proporción creciente de casos, idiomáticos, cohesionados y verosímiles, hasta el punto de resultar a menudo indistinguibles de los redactados por un traductor humano competente¹. La fluidez, sin embargo, es una virtud de la superficie, y su perfeccionamiento no implica el de una propiedad distinta y, en los contextos que más importan, decisiva: la fidelidad epistémica del texto, esto es, la correspondencia entre lo que afirma, la evidencia que lo respalda y el grado de certeza con que lo presenta.

Esta monografía parte de la tesis de que fluidez y fidelidad epistémica son propiedades dissociables, y de que su disociación es en especial peligrosa precisamente porque la competencia retórica de los sistemas vuelve sus fallos menos visibles. Un sistema que produce prosa impecable proyecta una apariencia de autoridad que invita a la confianza; y cuando esa prosa afirma más de lo que el original autorizaba —cuando fabrica una

¹ *Para la arquitectura que cataliza este salto, Vaswani et al. (2017); para la traducción neuronal con atención, Bahdanau et al. (2015) y Sutskever et al. (2014); para los modelos de lenguaje de gran escala y su aprendizaje en contexto, Brown et al. (2020); para el ajuste por retroalimentación humana que modela su comportamiento conversacional, Ouyang et al. (2022).*

precisión inexistente, convierte una conjetura en un hecho o suprime una cautela—, el lector que no puede cotejar la traducción con la fuente carece de todo indicio para desconfiar. La fluidez, lejos de corregir el problema de la fidelidad, lo agrava, al revestir de verosimilitud aquello que carece de fundamento. Este fenómeno, que la investigación sobre generación de lenguaje ha documentado bajo el rótulo de la alucinación², adquiere en la traducción un perfil propio: a diferencia de la generación libre, la traducción dispone de un referente externo y verificable —el texto fuente— respecto del cual la fidelidad puede, en principio, juzgarse con objetividad.

El problema cobra su gravedad plena en los dominios de alta responsabilidad —el jurídico, el clínico, el científico, el técnico, el histórico y el patrimonial—, en los que el coste de una infidelidad epistémica es asimétrico y, a veces, irreparable. Una cláusula contractual cuya modalidad se altera, un diagnóstico cuya cautela se suprime, un resultado experimental cuya provisionalidad se borra: en todos estos casos, una traducción más fluida pero menos honesta es peor que una traducción más áspera pero escrupulosa, porque traslada al destinatario una certeza que el original no garantizaba. La evaluación al uso de la traducción automática, sin embargo, no está concebida para captar esta diferencia: las métricas dominantes miden la proximidad a una referencia o la calidad percibida, no la proporción entre afirmación y soporte. De esta brecha —entre lo que sabemos medir y lo que necesitamos garantizar— nace la teoría que esta obra desarrolla.

² *Para panoramas de la alucinación en la generación de lenguaje natural y en los grandes modelos, Ji et al. (2023) y Huang et al. (2025); sobre la propensión de los modelos a reproducir falsedades humanas plausibles, Lin et al. (2022); para una crítica temprana de los riesgos de la fluidez sin anclaje, Bender et al. (2021).*