

# LENGUAJE, CULTURA Y ALGORITMOS: CONVERGENCIAS ENTRE LA INTELIGENCIA ARTIFICIAL Y LAS HUMANIDADES

CLAUDIO M. CONFORTI (COMP.)



COLECCIÓN DE LA FACULTAD DE FILOSOFÍA Y LETRAS  
PONTIFICIA UNIVERSIDAD CATÓLICA ARGENTINA



**LENGUAJE, CULTURA  
Y ALGORITMOS:  
CONVERGENCIAS ENTRE LA  
INTELIGENCIA ARTIFICIAL  
Y LAS HUMANIDADES**

CLAUDIO M. CONFORTI (COMP.)



Las publicaciones de la presente Colección han sido sometidas a una evaluación interna y externa organizada por la institución editora.

Corrección: Cintia L. Mariscal

Diseño gráfico y maquetación: Nicolás Gil

Edición: Cintia L. Mariscal

#### **Colección de Libros de Filosofía, Letras y Lenguas**

Conforti, Claudio M.

Lenguaje, cultura y algoritmos : convergencias entre la inteligencia artificial y las humanidades / Claudio M. Conforti ; Compilación de Claudio M. Conforti. - 1a ed. - Ciudad Autónoma de Buenos Aires : Universidad Católica Argentina, 2026.

Libro digital, PDF - (De la Facultad de Filosofía y Letras)

Archivo Digital: descarga y online

ISBN 978-950-44-0136-0

1. Inteligencia Artificial. 2. Lenguaje Artificial. I. Conforti, Claudio M., comp. II. Título.

CDD 006.3

Licencia

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

<https://doi.org/10.46553/978-950-44-0115-5>



Libros de la Facultad de Filosofía y Letras de la Pontificia Universidad Católica Argentina. **Pontificia Universidad Católica Argentina, Facultad de Filosofía y Letras.**

Rector: Dr. Miguel Ángel Schiavone  
Gran Canciller: Mons. Jorge García Cuerva  
Vicerrector de integración: Gustavo Boquin  
Vicerrector de investigación: Dr. Miguel Ángel Schiavone  
Decano: Dr. Javier Roberto González

---

### **Comité Editorial**

#### **Coordinador general de la Colección:**

Dr. Martín Grassi

#### **Serie de Estudios de Literatura y Lingüística:**

Dra. Magdalena Cámpora  
Dra. Ana María Marcovecchio  
Dra. María Lucía Puppo

#### **Serie de Filosofía**

Dr. Mateo Belgrano  
Dr. Francisco Díez Fischer  
Dra. Fernanda Ocampo  
Dr. Federico Raffo Quintana  
Dr. Juan Torbidoni

#### **Serie de Lenguas**

Dra. Marina Álvarez  
Dra. Inés Castelli  
Dra. Graciela Isaia y Ruiz

### **Comité Académico**

#### **Serie Filosofía**

José Luis Villacañas Berlanga (Universidad Complutense de Madrid)  
Adrián Bertorello (Universidad de Buenos Aires/CONICET)  
Marcelo Boeri (Pontificia Universidad Católica de Chile)  
Oscar Miguel Esquisabel (INEO-CIF/CONICET)  
Olga Larre (Universidad Católica Argentina/CONICET)  
Jacques Lezra (University of California, Riverside)  
Juan Antonio Nicolás Marín (Universidad de Granada)

Francisco O'Reilly (Universidad de Montevideo)  
Luis Rabanque (Universidad Católica Argentina/ CONICET)  
Roberto Rubio (Universidad Alberto Hurtado)  
Claudinei Silva (Univeridade Estadual do Oeste do Paraná)  
Roberto Walton (Universidad de Buenos Aires/CONICET)

### **Serie Literatura y Lingüística**

Lila Bujaldón de Estévez (Universidad Nacional de Cuyo)  
Ana Gallego Cuiñas (Universidad de Granada)  
Fernando Degiovanni (City University of New York)  
Mariana Di Ció (Université Sorbonne-Nouvelle)  
Enrique Foffani (Universidad Nacional de La Plata)  
Gustavo Guerrero (Université de Cergy-Pontoise)  
Rosa García Gutiérrez (Universidad de Huelva)  
Lucía Stecher Guzmán (Universidad de Chile)  
Annick Louis (Université de Franche-Comté)  
Alfonso García Morales (Universidad de Sevilla)  
María Marta García Negroni (Universidad de Buenos Aires, CONICET)  
Mariano Sverdloff (Universidad de Buenos Aires, CONICET)

### **Serie Lenguas**

Herminia Alonso (Universidad de Buenos Aires)  
Daniel Altamiranda (Universidad Católica Argentina)  
Andy Benzo (Universidad de San Diego / American Translators  
Association)  
Ricardo Chiesa (Universidad de Buenos Aires)  
Gabriela Commatteo (King's College / Universidad Pompeu Fabra)  
Silvana Debonis (Universidad Católica Argentina / American Translators  
Association)  
Juan José Delaney (Universidad del Salvador, Argentina)  
Cristina Featherston (Universidad Nacional de La Plata)  
Gabriela Llull (Universidad de Salamanca)  
Sandra Ramacciotti (UMSA / Universidad de Buenos Aires)  
Xose Castro Roig (ISTRAD / Universidad Pompeu Fabra)  
Graciela Souto (Universidad Católica Argentina / New York University)

La *Colección de libros de Filosofía y Letras* se compone de tres series, cada una de las cuales está dedicada a la publicación de trabajos científicos en los campos de la filosofía, la literatura y la lingüística, y las lenguas extranjeras.

Los trabajos publicados son el fruto de las investigaciones de los/as docentes y alumnos/as de la Facultad de Filosofía y Letras de la Universidad Católica Argentina, aunque también incluye contribuciones de investigadores/as provenientes de otras universidades, tanto nacionales como extranjeras. En este sentido, la Colección busca ser un órgano de difusión de la vida intelectual de nuestra Facultad y de articulación con colegas de otras instituciones.

Dr. Martín Grassi / Dr. Javier Roberto González

# Índice

Prólogo, por Claudio M. Conforti	11
Humanismo y Transhumanismo, convergencias y divergencias, por Dulce María Santiago	17
Digesto 1.1.10.1 como método para la valoración de la capacidad intelectual algorítmica, por Ian Henríquez Herrera y Roxana Álvarez Sanguinetti	28
Aipóiesis. Transposición y Mímesis entre Narrativas de la Ciencia Ficción y el Realismo en Inteligencia Artificial, por Gustavo Ángel Riesgo	37
Lógica Natural vs. Lógica Formal. Un debate inaplazable, por Enrique Alonso	64
Un nuevo paradigma para la formación de traductores en la era de la IA generativa, por Silvana Debonis	94
Sobre los/as autores/as	138

La presente publicación es fruto de un gesto generoso y profundamente significativo: la apertura de la Facultad de Filosofía de la Pontificia Universidad Católica Argentina al diálogo plural, reflexivo y crítico sobre uno de los fenómenos más desafiantes de nuestro tiempo. La Dra. Olga Larre, en ese momento Decana de dicha Facultad, no solo nos brindó su apoyo institucional, sino que promovió activamente la posibilidad de convocar voces especializadas del ámbito nacional e internacional para pensar juntos –desde las Humanidades– los impactos éticos, políticos y culturales de la inteligencia artificial generativa. En un contexto marcado por aceleradas transformaciones tecnológicas, la iniciativa de abrir este espacio académico refleja el compromiso de la Universidad con su misión fundacional: la búsqueda de la verdad, el discernimiento ético y la promoción de una cultura del encuentro. Este libro se enmarca en ese espíritu, proponiendo una reflexión interdisciplinaria que no elude las preguntas difíciles, y que busca anclar los debates contemporáneos en una visión humanista, cristiana y crítica del mundo.

La irrupción de la inteligencia artificial en los distintos campos del conocimiento ha generado un profundo impacto en las disciplinas clásicas, particularmente en aquellas vinculadas con las humanidades, como puede ser el estudio de las lenguas. En un contexto en el que se observa una transformación que avanza a pasos agigantados, las carreras de Profesorado y Traducción en Inglés se ven interpeladas en sus fundamentos epistemológicos, sus prácticas pedagógicas y sus proyecciones profesionales. En este escenario, se vuelve aún más necesario revalorizar el rol de profesionales de excelencia, ex pertos en sus áreas de conocimiento, con una sólida formación académica y en permanente actualización. Son ellos quienes, con pensamiento crítico, sensibilidad lingüística y compromiso ético, están en condiciones de interpretar los cambios en curso y de liderar procesos de innovación pedagógica y profesional. Desde el Departamento de Lenguas, asumimos el compromiso de abordar estos desafíos con una mirada crítica, informada y propositiva. Entendemos que la inteligencia artificial no constituye únicamente una herramienta tecnológica, sino un fenómeno cultural y cognitivo que redefine las formas de producción, circulación y apropiación del conocimiento. En este sentido, consideramos indispensable repensar los marcos teóricos que sustentan nuestras disciplinas, así como las competencias que deben desarrollar los futuros profesionales del lenguaje. Esta publicación se inscribe en ese horizonte. Reúne aportes que buscan problematizar el impacto de la IA en la enseñanza del inglés como lengua

extranjera y en la práctica de la traducción específica y especializada, desde una perspectiva que articula innovación tecnológica con pensamiento crítico y compromiso ético. Lejos de concebir el avance tecnológico como una amenaza, lo entendemos como una oportunidad para fortalecer el rol de las humanidades en la formación integral de sujetos capaces de intervenir activamente en un mundo en constante transformación.

Prof. Claudio M. Conforti

## Prólogo

Prof. Claudio M. Conforti

Nos encontramos en el umbral de una transformación sin precedentes. La inteligencia artificial, más que una innovación tecnológica, representa una revolución cultural, antropológica y espiritual que desafía nuestras formas tradicionales de comprender el mundo, la persona y la sociedad.

Las Humanidades, a menudo relegadas a un lugar marginal en el debate sobre el futuro, se revelan hoy más necesarias que nunca. Desde su vocación por la búsqueda de sentido, la reflexión sobre la verdad, el bien y la dignidad humana, las disciplinas humanísticas están llamadas a ofrecer un marco crítico, ético y espiritual que oriente el desarrollo tecnológico hacia fines verdaderamente humanos. La pregunta por el sentido último de nuestras acciones, por el lugar del ser humano en la creación y por la justicia en las relaciones sociales, no puede ser sustituida por el cálculo algorítmico ni por la eficiencia funcional.

Esta obra, *Lenguaje, Cultura y Algoritmos: convergencias entre la Inteligencia Artificial y las Humanidades*, se inscribe en el horizonte de la misión propia de la Universidad Católica: formar personas comprometidas con la verdad y el bien común, a la luz del Evangelio y de la Doctrina Social de la Iglesia. Desde una mirada interdisciplinaria, el presente volumen busca abrir un espacio de discernimiento sereno y profundo, en el que se articulen los saberes humanísticos con los desafíos que plantea la inteligencia artificial a nivel ético, político y social.

Tal como lo expresa el Papa Francisco en *Laudato Si'* (2015), “el modo en que la humanidad ha hecho uso de la tecnología fue una manera de ejercer el dominio sobre la tierra que acabó arrasándola” (LS, 2015, p. 106). Este llamado a la responsabilidad nos invita a una conversión cultural que supere la lógica instrumental y tecnocrática. La inteligencia artificial no debe ser concebida como un fin en sí misma, sino como una herramienta al servicio de una ecología integral, que reconozca la centralidad de la persona humana y su vínculo esencial con toda la creación.

Asimismo, en *Fratelli Tutti* (2020), el Santo Padre advierte sobre una “cultura del descarte” que puede agravarse con el mal uso de la tecnología, y nos exhorta a construir una “fraternidad abierta, que permita reconocer, valorar y amar a cada persona más allá de la cercanía física” (FT, 2020, p. 1). En este contexto, el desarrollo de la inteligencia artificial plantea preguntas fundamentales: ¿Qué modelo de humanidad queremos promover? ¿Cómo garantizar que el progreso tecnológico no profundice desigualdades ni socave nuestra responsabilidad moral y social?

En línea con estos principios, el *Rome Call for AI Ethics* –firmado por líderes religiosos, académicos y representantes del sector tecnológico– sostiene que “la innovación digital y el progreso tecnológico deben estar al

servicio del bien de la humanidad y del planeta” (Ren Alssance Foundation, s.f., párr. 2). Este llamado a una inteligencia artificial ética y confiable nos recuerda que la técnica debe estar guiada por la sabiduría, y que la protección de la dignidad humana debe ser siempre el criterio rector de toda transformación digital.

La Iglesia, fiel a su tradición de diálogo con la cultura y la ciencia, no puede permanecer indiferente ante este cambio de época. Como comunidad académica creyente, estamos llamados a ofrecer una contribución lúcida y profética, desde las raíces más profundas de nuestra tradición humanista y cristiana.

Este libro es, en definitiva, una invitación a pensar juntos qué tipo de futuro queremos construir, con esperanza, con responsabilidad y con la convicción de que el saber, iluminado por la fe, puede y debe contribuir a un mundo más justo, más humano y fraterno.

### **“Humanismo y Transhumanismo, convergencias y divergencias”**

En su artículo, la Dra. Dulce María Santiago ofrece una profunda reflexión filosófica sobre el estatuto del ser humano en el marco de las transformaciones culturales y tecnológicas de nuestra época. Partiendo de una revisión del pensamiento humanista en el siglo XX –tanto en sus vertientes ateas (Sartre, Heidegger) como cristianas (Jaspers, Marcel)–, la autora plantea cómo el concepto de “lo humano” ha sido puesto en cuestión frente al avance del pensamiento posthumanista y transhumanista. Estas nuevas corrientes, ligadas al desarrollo de las neurociencias, la inteligencia artificial y la biotecnología, proponen superar las limitaciones físicas e intelectuales del ser humano mediante el uso intensivo de la técnica, promoviendo un ideal de “mejoramiento” que, en algunos casos, apunta incluso a la inmortalidad.

El artículo establece un diálogo entre estas posiciones contemporáneas y las raíces metafísicas del humanismo, señalando cómo la tradición filosófica –desde la antropología filosófica hasta los aportes de pensadores como Quiles y Derisi– ha buscado siempre una comprensión integral de la persona, abierta a la trascendencia. La autora advierte sobre los riesgos del constructivismo radical que subyace al transhumanismo, donde la percepción individual tiende a reemplazar toda referencia objetiva, incluso en dimensiones como la identidad sexual o la noción de naturaleza humana.

Desde una perspectiva crítica, pero también propositiva, el texto subraya la urgencia de renovar el diálogo entre el Humanismo Cristiano y las nuevas formas del pensamiento tecnológico. Frente a la fascinación por el poder técnico, se plantea la necesidad de revalorizar una visión del ser humano que lo entienda no solo como un ente modificable, sino como un ser dotado de dignidad intrínseca, capaz de buscar el sentido, el bien y la verdad. En un mundo que tiende a desplazar el lenguaje, la interioridad y la trascendencia, el artículo defiende el lugar insustituible del pensamiento humanista como garante de un enfoque ético integral frente a los desafíos del presente.

### **“Digesto 1.1.10.1 como método para la valoración de la capacidad intelectual algorítmica”**

En este ensayo, Ian Henríquez y Roxana Álvarez Sanguineti proponen una reflexión original y rigurosa que vincula el pensamiento jurídico clásico con los desafíos epistemológicos y éticos que plantea la inteligencia artificial. A partir del célebre precepto del Digesto de Justiniano –“dar a cada cual lo suyo”–, los autores articulan una crítica profunda a las narrativas que atribuyen inteligencia, conciencia o agencia moral a sistemas algorítmicos, señalando el riesgo de un uso abusivo del lenguaje que puede derivar en injusticias tanto simbólicas como jurídicas.

La tesis central del artículo es que el reconocimiento de las máquinas como entidades inteligentes –más aún, como sujetos morales o jurídicos– constituye una forma de injusticia epistémica: se distorsiona la realidad de lo que son las máquinas, se oculta su dependencia estructural del entrenamiento humano, y se promueve una ilusión tecnocrática que degrada el estatuto ontológico y ético de la persona humana. Para fundamentar esta crítica, los autores acuden a una impresionante constelación de voces filosóficas, científicas y jurídicas de –Chomsky a Han, de Kate Crawford a Markus Gabriel, de Innerarity a Lanier– que coinciden en rechazar la atribución literal de inteligencia o conciencia a los artefactos.

En un ejercicio de lucidez crítica, el texto advierte que la inteligencia artificial opera mediante correlaciones estadísticas, carece de interioridad, disposición anímica o intencionalidad, y por ende, no puede equipararse al pensamiento humano. La metáfora del “robot que siente” o “decide” es desenmascarada como una ficción que, aunque culturalmente tolerada, se vuelve peligrosa cuando se transforma en criterio normativo o axiológico. En este sentido, el artículo defiende con firmeza la necesidad de una formación humanística que permita reconocer los límites ontológicos y semánticos de los sistemas algorítmicos y preserve la dignidad de la persona como único sujeto de derechos.

Finalmente, los autores plantean que nuestra América tiene una oportunidad singular para aportar al debate global sobre inteligencia artificial, precisamente por su tradición lingüística, filosófica y religiosa, que conserva una mirada profunda sobre la interioridad humana, la trascendencia y la justicia. Frente al reduccionismo tecnocrático, este artículo ofrece una respuesta académica fundada en la tradición del derecho romano, la filosofía del lenguaje y la antropología integral, que coloca a la persona –y no al artefacto– en el centro de la reflexión ética.

### **“Aipoiesis: narrativas del fin de la creación”**

En su lúcido y provocador ensayo “*Aipoiesis: narrativas del fin de la creación*”, Gustavo Riesgo explora los desplazamientos simbólicos y ontológicos que introduce la inteligencia artificial generativa en el campo de la creación

artística y poética. A partir de una lectura crítica de diversos discursos –tanto tecnocientíficos como culturales–, el autor interroga el sentido mismo de la *poiesis* en un tiempo en que la producción de textos, imágenes y sonidos por parte de máquinas parece desafiar la noción tradicional de autoría, inspiración y subjetividad creadora.

Con una prosa ensayística de alto vuelo conceptual, Riesgo examina la emergencia de una *aipoiesis*: un régimen de generación sin creación, de producción sin mundo interior, donde la sintaxis puede simularse pero no habitarse. Esta distinción –poética y filosófica a la vez– es central en su planteo: si la *poiesis* humana brota de una experiencia del sentido, de un hiato entre lenguaje y mundo, la generación maquinaica, en cambio, opera sin memoria ni deseo, sin muerte ni conciencia del límite. La inteligencia artificial puede emular estructuras, pero no asumir el riesgo existencial que implica decir.

El artículo recorre también las consecuencias culturales de esta mutación. Lejos de caer en el rechazo apocalíptico de la técnica, Riesgo propone un pensamiento poético que resista la homologación entre creación humana y producción algorítmica. En ese gesto, se abre paso a una ética del arte y del lenguaje que no puede ser reducida a estadísticas de probabilidad ni a modelos de predicción. La *aipoiesis* es, para el autor, la oportunidad para re-pensar el misterio de la creación desde su fragilidad constitutiva, desde su precariedad ontológica, como acto siempre situado, encarnado y abierto al otro.

Este artículo constituye, sin duda, uno de los aportes más originales y sensibles del volumen, al mostrar que el pensamiento humanista no se agota en la crítica racional, sino que puede también poetizar y habitar el desconcierto, ofreciendo vislumbres de sentido allí donde el cálculo no alcanza.

### **“Lógica Natural vs. Lógica Formal. Un debate inaplazable”**

El artículo del Dr. Enrique Alonso, titulado “Lógica Natural vs. Lógica Formal. Un debate inaplazable”, examina críticamente el surgimiento y la consolidación de una corriente contemporánea en el ámbito de la lógica: la llamada Lógica Natural. Este enfoque, revitalizado por el desarrollo de la inteligencia artificial generativa y las tecnologías del lenguaje natural, se presenta como alternativa –o al menos como complemento– a la lógica formal canónica desarrollada desde Frege hasta nuestros días. A lo largo del texto, Alonso distingue con claridad los fundamentos filosóficos, metodológicos y computacionales que separan ambas tradiciones, y pone en cuestión la pretensión de universalidad de los lenguajes artificiales empleados por la lógica matemática tradicional para analizar la inferencia.

El autor se detiene en el análisis de dos grandes escuelas contemporáneas de la Lógica Natural: la escuela holandesa, representada por Johan van Benthem y Víctor Sánchez-Valencia, y la escuela americana, asociada a figuras como MacCartney y Moss. La primera se centra en el estudio de

la polaridad léxica y el uso de gramáticas categoriales para explicar inferencias en lenguaje ordinario, mientras que la segunda desarrolla un enfoque computacional orientado a sistemas de inferencia en procesamiento de lenguaje natural. Ambas coinciden en recuperar elementos de la silogística aristotélica, aunque enriquecidos con herramientas lingüísticas y semánticas modernas.

El artículo culmina con la propuesta de una concepción más amplia y rigurosa de la inferencia natural, fundada en lo que el autor denomina postulados de significado –léxicos o lógicos–, los cuales articulan una visión más realista de la cognición inferencial humana. Así, Alonso sugiere que la lógica natural no debe ser entendida meramente como una técnica alternativa, sino como una disciplina autónoma que estudia los procesos inferenciales cotidianos desde una base filosófica, lingüística y computacional integrada.

Este enfoque resulta especialmente pertinente frente al auge de la inteligencia artificial generativa, cuyas operaciones de inferencia se aproximan más al modelo de lógica natural que al formalismo clásico, y plantea importantes desafíos epistémicos y éticos a la tradición humanista.

### **“Un nuevo paradigma para la formación de traductores en la era de la IA generativa”**

En este artículo, el equipo docente del Departamento de Lenguas de la Universidad Católica Argentina –con participación destacada de la profesora Silvana Debonis– ofrece una reflexión crítica y rigurosa sobre los desafíos que la inteligencia artificial generativa impone a la formación de traductores en el siglo XXI. Lejos de limitarse a un enfoque técnico, el texto articula una perspectiva epistemológica, pedagógica y ética que sitúa a la traducción como una práctica profundamente humana, cuya centralidad se renueva –no se disuelve– en el contexto de las tecnologías del lenguaje.

El artículo parte de un diagnóstico lúcido: la expansión de modelos como ChatGPT y DeepL ha transformado no solo los modos de producción textual, sino también las expectativas sociales sobre lo que significa traducir. En este contexto, los autores rechazan tanto la fascinación acrítica por la eficiencia algorítmica como el repliegue nostálgico. En su lugar, proponen un “nuevo paradigma formativo” que asuma la complejidad del entorno digital sin renunciar al corazón humanista de la disciplina.

Uno de los aportes centrales del texto radica en su reivindicación del juicio humano como núcleo insustituible del acto traductivo. La traducción no es –afirman– una mera transferencia de significados, sino una mediación interpretativa situada, que exige competencia cultural, sensibilidad lingüística, discernimiento ético y responsabilidad profesional. La IA puede asistir, sugerir o automatizar ciertos procesos, pero no puede reemplazar la comprensión profunda de los contextos, matices y efectos de sentido que un traductor debe manejar.

Desde esta perspectiva, el artículo apuesta por una redefinición del perfil

profesional del traductor: no como técnico subordinado a la máquina, sino como intelectual crítico, capaz de dialogar con las herramientas tecnológicas, evaluar sus límites y decidir sobre su uso en función de criterios humanos y no meramente funcionales. Se trata de formar sujetos con capacidad de agencia, alfabetizados en tecnologías, pero arraigados en una visión ética del lenguaje y del mundo.

Este aporte, riguroso y comprometido, resulta clave en el volumen, ya que muestra con claridad cómo el campo de la traducción puede convertirse en laboratorio privilegiado para repensar el lugar de las Humanidades frente a la IA: no en términos de oposición, sino de discernimiento crítico y de apertura creativa.

Por último, deseo expresar nuestro sincero agradecimiento al Departamento de Lenguas y a la Facultad de Filosofía y Letras de la Pontificia Universidad Católica Argentina, cuya apertura, compromiso y discernimiento hicieron posible esta publicación. La convocatoria para pensar juntos, desde diversos enfoques, el impacto ético y cultural de la inteligencia artificial, es testimonio de una comunidad académica viva, atenta a los signos de los tiempos y fiel a su vocación de diálogo entre fe, ciencia y cultura.

Desde una perspectiva inspirada por el humanismo cristiano, este libro busca ser una modesta pero firme contribución al necesario discernimiento sobre el futuro de nuestras sociedades. Como nos recuerda el Papa Francisco, “todo ser humano posee una dignidad inalienable” (Fratelli Tutti, 2020, p.107), y es desde esa convicción que asumimos la tarea de pensar críticamente el lugar de la técnica en la vida humana. Que este esfuerzo compartido sea signo de esperanza y fermento de justicia, en un mundo que necesita con urgencia más inteligencia... pero, sobre todo, más sabiduría.

# Humanismo y Transhumanismo, convergencias y divergencias

Dulce María Santiago

## Introducción

“No poseemos una idea unitaria del hombre” –dice Max Scheler– “en ninguna época ha resultado el hombre tan *problemático* como en la actualidad”, ya que, “la multiplicidad siempre creciente de ciencias especiales que se ocupan del hombre oculta la esencia del hombre mucho más que la iluminan”. Pero la filosofía del hombre no puede establecerse al margen de un diálogo con las ciencias humanas.

Cuando el antropólogo Lévy-Bruhl le envía su obra *La mitología primitiva* a Husserl, éste le escribe una carta donde valora la obra científica del antropólogo y le dice que acepta las conclusiones de las ciencias, pero mientras “no tengan la última palabra” (Husserl, 1994).

La antropología filosófica es un saber que tiene por objeto al hombre, que significa una comprensión metafísica de cuanto las ciencias humanas han aportado al conocimiento del ser humano.

“El ser humano necesita saber lo que es para *serlo*”, necesita una idea de lo que es “ser hombre”, en la que pueda reconocerse como sujeto y que guíe su proceso de autorrealización. Así, desde el punto de vista de las ciencias, como desde la existencia humana, es necesaria una *idea unitaria e integral del hombre*. Las ciencias positivas no bastan para ofrecerla porque *no tienen la última palabra*.

## Humanismos en el siglo XX

Como puede verse, el siglo XX tiene una consideración especial por el tema del hombre y eso lo lleva a postular en sus corrientes filosóficas el *Humanismo*.

Aunque el término “Humanismo” nos remite a:

- Cultivo de las humanidades.
- Movimiento intelectual desarrollado en Europa durante los siglos XIV y XV que, rompiendo las tradiciones escolásticas medievales y exaltando en su totalidad las cualidades propias de la naturaleza humana, pretendía descubrir al hombre y dar un sentido racional a la vida tomando como maestros a los clásicos griegos y latinos, cuyas obras redescubrió y estudió.

En realidad, el término se vuelve a utilizar en el siglo XX, especialmente a partir de un resurgimiento del tema de lo *humano* en un contexto europeo sociopolítico marcado por las guerras y los conflictos sociales que conducen

al resurgimiento del concepto de “Humanismo”, especialmente a partir de las guerras mundiales.

Surgirán así los *Humanismos ateos* y los *Humanismos cristianos*. En este contexto del fin de las guerras, en su célebre Conferencia *El existencialismo es un humanismo* (1977) Jean Paul Sartre dice:

El reproche esencial que nos hacen, como se sabe, es que ponemos el acento en el lado malo de la vida humana (Sartre, 1977).

(...) hay quien utiliza la sabiduría de los pueblos, que es bien triste, y nos encuentra más tristes todavía (...) (Sartre, 1977).

¿Qué es el existencialismo? (Sartre, 1977).

Lo que complica las cosas es que hay dos especies de existencialistas: los cristianos de confesión católica, entre los cuales se encuentran Jaspers y Gabriel Marcel y los existencialistas ateos, entre los que me identifico, y en donde se podría ubicar, también, a Heidegger y a los existencialistas franceses. Lo que tienen en común ambos existencialismos, es el hecho de considerar que la existencia precede a la esencia, o, si se prefiere, que ésta nace a partir de la subjetividad.

¿Qué significa que la existencia precede a la esencia? Significa que el hombre empieza por existir, se encuentra, surge en el mundo y que después se define. El hombre, tal como lo concibe el existencialista, si no es definible, es porque empieza por no ser nada. Sólo será después, y será tal como se haya hecho. Así pues, no hay naturaleza humana, porque no hay Dios para concebirla... El hombre no es otra cosa que lo que él se hace (Sartre, 1977, pp. 12-14).

Nietzsche y el anuncio de la llegada del nihilismo. La muerte de Dios. La devastación del lenguaje. La devastación de la tierra. La huida de los dioses y la noche del mundo. El olvido del ser.

Pensar originario y filosofía ¿Qué dice el genitivo de la expresión “pensar *del ser*”? La filosofía como meditación o como técnica de explicación a partir de causas supremas, es decir, como metafísica. La necesidad de superación de la subjetividad metafísica. La cuestión de la técnica. *Praxis y poiesis* en el hacer humano.

Lo *impensado* de Parménides: “es en *efecto ser*”, “*hay en efecto ser*”. La herencia de los primeros pensadores. “El único asunto del pensar es llevar el lenguaje al advenimiento del ser, que permanece y en su permanecer espera al hombre”.

Heráclito: *también hay dioses*. La dimensión de lo sagrado. El hombre en la medida en que es, mora en lo divino. “Tal vez lo característico de esta era mundial sea precisamente que se ha cerrado la dimensión de lo salvo (*das Heil*). “La filosofía y el destino de la verdad.

## Humanismo en la Argentina: Marxismo y Humanismo en los años '60

En los años 60, la UBA reconoce que en las elecciones universitarias de 1962 triunfa el “Humanismo”<sup>2</sup>, corriente universitaria de tendencia católica. Así refiere Julio Bárbaro (2021), en su libro *Juicio a los 70*, la situación intelectual de los estudiantes en esta década:

El catolicismo y el marxismo eran las dos opciones que se nos abrían a los jóvenes: se enfrentaban en Europa y nosotros vivíamos los coletazos. En aquel momento ser católico era ser *humanista* y ser ateo era ser *marxista*: ya vendrían otras divisiones y combinaciones, pero todavía las equivalencias eran sencillas. Me convertí en un delegado de la Liga Humanista, que estaba creciendo –había ganado en Ciencias Económicas, la facultad más grande de América Latina, con 20.000 estudiantes– y en las elecciones de 1962 logró imponer al rector Julio Olivera. En los *Cuadernos humanistas* dábamos cuenta de nuestra dependencia cultural: nos influían más **Emanuel Mounier** y **Teilhard de Chardín**, quienes tronzaron nuestros ejes filosóficos, que los pensadores de nuestro país. Nos vinculamos con el Integralismo, que prendía en Córdoba, y con los católicos de la Unión de Estudiantes del Litoral. Con los años me convertiría en secretario general (1963) y presidente de la Liga Humanista de Buenos Aires (1966) desde donde participaría como cofundador de la Unión Nacional de Estudiantes (UNE) (Bárbaro, 2021, p. 34) (Las negritas son propias).

De los dos rectores que tuvo la UBA, fueron los doctores Julio Olivera (1962 a 1965) e Hilario Fernández Long (1965 a 1966) quienes estuvieron ligados al Humanismo, hasta el golpe de estado de 1966.

Así, la década siguiente a la creación de las Universidades privadas transcurre en un ambiente intelectual que oscila entre dos corrientes bien definidas: el pensamiento marxista, representado por jóvenes que profesaban el ateísmo en materia religiosa, y por la Liga Humanista

La irrupción de las universidades privadas contribuyó, también, al aumento de la matrícula estudiantil. En ese sentido, Julio Bárbaro afirma que:

Los universitarios habíamos pasado de ser 138.000 en 1955 a 220.000 en diez años. La polémica sobre la educación “laica o libre” durante los años de Frondizi y la oposición entre científicistas y transformadores de la realidad social habían fertilizado el terreno para un sector históricamente activo como el nuestro. Protestábamos contra los exámenes de ingreso, el aumento de los aranceles del comedor, el precio de los apuntes, los recortes del presupuesto. Los estudiantes cristianos empezamos a descubrir las villas y ampliamos nuestras inquietudes sociales (Quiles, 1954, pp. 313-314).

Al llegar la década siguiente, la juventud aparece como un valor esencial y asoma una promesa de futuro marcada por los icónicos años 60', caracterizado por la emergencia de movimientos juveniles como el **mayo francés** en Europa.

## Los “Humanismos” de Quiles y Derisi

Dentro del debate cultural de la época, Quiles precisa su postura Humanista, distinguiendo un “Humanismo cerrado” y un “Humanismo abierto”:

Cuando el hombre no mira más allá de ese horizonte, por así decir, primario y superficial de nuestras experiencias, la explicación del hombre resulta un absurdo. No pasa más allá del hombre mismo y éste se halla condenado a dar vueltas en el interior de sí, sin posibilidad alguna de dilucidar su ser y su situación: es *humanismo cerrado*, en el cual no hay ninguna perspectiva de salvación para la tragedia humana en este mundo.

Pero, si se atiende *al conjunto integral de las experiencias humanas*, aparece el hombre comunicado por mil puentes con realidades trascendentes al hombre mismo: es el *humanismo abierto o comunicado* hacia el exterior del hombre, pero con una *apertura* y comunicación natural al hombre mismo. Este análisis integral del hombre nos da las coordenadas básicas de la existencia humana (Quiles, 1954, pp. 313-314).

Y su pensamiento sobre el hombre en el siglo XX, en diálogo con la cultura de esa época, puede resumirse en sus palabras de presentación de su obra *Más allá del Existencialismo*:

Efectivamente, los existencialistas, en vez de descubrirnos la esencia del hombre, en vez de encontrar al hombre que buscaban, parece que lo han perdido y extraviado más, resultando que el existencialista, lejos de encontrarse a sí mismo, se halla más perdido en el mundo que los antiguos filósofos existencialistas. La necesidad de “encontrarse a sí mismo”, única manera de recuperar la realidad de la existencia humana es la que nos ha movido a volvernos hacia el interior del hombre, hacia su experiencia más original, y de ahí ha nacido la filosofía del ser y la dignidad del hombre que hemos llamado “in-sistencial” (Sartre, 1997)

Derisi, por su parte, siempre estuvo en la misma corriente filosófica y tuvo una actitud diferente a la de Quiles respecto del pensamiento contemporáneo. A primera vista, pareciera que no dialoga con la filosofía. Sin embargo, encontramos otros testimonios ya que su propuesta de un Humanismo Auténtico no es el Humanismo grecolatino, ni Renacentista, ni contemporáneo. El “Humanismo Cristiano” “que no es esencialmente antropocéntrico ni autónomo sino, por el contrario, teocéntrico, sometido a la Verdad y al Bien de Dios y, como tal, enteramente compatible con el Cristo centrismo”.

En su obra *Naturaleza y vida de la Universidad*, Derisi (1969) nos dice:

Si las palabras y los conceptos que las animan conservan aún su valor, debemos convenir que el *humanismo* consiste en un desarrollo del hombre en orden a su perfección humana. Por eso el *humanismo*, como

la *cultura* que lo realiza, se estructura como un recorrido entre dos términos: entre el hombre tal cual es, y el hombre tal cual debe ser, o más brevemente, entre el *ser* y *deber ser* del hombre. El *humanismo*, pues, es el resultado de este recorrido del hombre hacia la trascendencia del Ser por los caminos que llevan a la *belleza*, a la *bondad* y a la *verdad* (Derisi, 1969).

## Crisis del Humanismo

En su conocida “*Carta sobre el Humanismo*” –originada en 1946, en un intercambio epistolar con Jean Beaufret–, Martín Heidegger sostiene: “falta mucho para que nosotros pensemos sobre la esencia del obrar en forma suficientemente decidida.” Y también: “El lenguaje es la casa del Ser. En su morada habita el hombre”.

Heidegger se plantea: ¿Desde dónde y cómo se determina la esencia del hombre? Todo humanismo se basa en la metafísica. El extravío de los intentos de definir al hombre. La mala interpretación de la definición aristotélica como *animal racional*. La incapacidad de alcanzar la esencia del hombre en el humanismo histórico. La necesidad de radicalizar la pregunta por lo humano desde la pregunta por el Ser.

Ni sustancia, ni persona, ni yo, ni sujeto: *Dasein* es la palabra para la existencia. Que el hombre existe significa que mora en el claro del ser. El lenguaje como morada del ser. “Liberar al lenguaje de la gramática para ganar un orden esencial más originario es algo reservado al pensar y poetizar”.

En 1999, en su libro *Normas para el parque humano*, Peter Sloterdijk afirma que;

Heidegger deja al descubierto perturbadoras consecuencias: el humanismo, en su forma Antigua tanto como en sus formas cristiana y moderna, es identificado como el responsable de un no-pensar de veinte siglos, con sus presupuestos apresurados, sus al parecer evidentes e inevitables definiciones de la esencia del hombre, se lo acusa de haber impedido el surgimiento de la auténtica pregunta por la esencia (Sloterdijk, 2007, p. 8).

Y agrega:

La era del humanismo moderno como modelo escolar y educativo ya ha pasado porque se ha vuelto insostenible la ilusión de que masivas estructuras políticas y económicas pueden ser ya organizadas siguiendo el modelo amigable de la sociedad literaria (Sloterdijk, 2007, p. 4).

Y también:

¿Qué puede aún domesticar al hombre, si hasta el día de hoy sus esfuerzos de auto moderación lo han llevado en gran medida precisamente a su **toma del poder sobre todo ente?** (Sloterdijk, 2007, p.8 la negrita es propia).

Estas afirmaciones manifiestan claramente la crisis del humanismo como forma de mejorar, educar y humanizar al hombre. Esta es la experiencia histórica que el siglo XX, especialmente en Europa, nos ha dejado y después de tantos siglos de procurar mejorar la vida social y política de las naciones. Las guerras y sus nefastas consecuencias responsabilizan al fracaso de querer hacer del hombre un ser civilizado. Aparecen así los límites del Humanismo que pueden sintetizarse en:

- Considerar de alguna manera al hombre como “medida de todas las cosas”, como decía Protágoras.
- Caer en un “antropocentrismo” al colocar al ser humano como el centro del universo.
- Hacer de la vida social una exaltación de las libertades individuales y caer en el Individualismo.
- Categorizar los grupos humanos en la modernidad en “superiores” (raza blanca, el hombre blanco) que genera el dualismo “civilización y barbarie”. Cabe preguntarse si el bárbaro no es humano.
- Descuidar la naturaleza y el planeta como resultado de la visión moderna que los considera simplemente un *recurso*.

### Posthumanismo y Transhumanismo

Se trata de dos categorías o conceptos relacionados con el paradigma de la posmodernidad, donde el *desencanto* por la Crisis de la Razón lleva a considerar que lo “vital”, es decir, “sentir”, “vivir”, en sentido biológico, es lo fundamental para el ser humano y, por lo tanto, todo lo que tenga ese fin debe ser realizado para conseguir “sentirse bien”, identificado con el fin de la existencia humana.

El término “Posthumanismo” es utilizado como una postura que aspira a la superación del “Humanismo”, en el sentido que adquiere en el Renacimiento, como un movimiento que centra su tema en el *hombre*, considerado como el ser que tiene un *espíritu* que le permite conocer e interpretar la vida con un fin trascendente.

También la palabra “Posthumanismo” se relaciona con el “Transhumanismo”, al superar las limitaciones físicas e intelectuales mediante la ciencia y la tecnología (neurociencias, ingeniería genética), que le permiten expandir al máximo sus capacidades llegando a ser más que humano, *posthumano*. Puede surgir entonces un nuevo tipo humano gracias a los avances de la tecnología que lo llevan hasta los límites de lo humano, incluso a pensar en la posibilidad de convertirlo en inmortal. Se trata de un hombre que, gracias a la ciencia y a la técnica, puede aspirar a su máxima situación vital, su permanente juventud, y a una vida sin los límites de una naturaleza que ha sido superada.

El posthumanismo se constituye como una respuesta filosófica a un

mundo donde cada vez es más difícil distinguir entre lo natural y lo artificial (si acaso aún fuera necesario hacer dicha distinción) y en el que el eje escritura/lectura que articulaba la cultura humanista pierde protagonismo ante la emergencia de nuevos **medios de expresión y comunicación**. Frente al miedo a que las máquinas terminen sustituyendo a los humanos, el posthumanismo plantea la necesidad de desarrollar un pensamiento ecológico (en su sentido más amplio) que tenga en cuenta no sólo el entorno natural sino también el tecnológico e incorpore, entre otras cosas, los derechos cívicos de las máquinas.

De este modo, el transhumanismo es una corriente de pensamiento que fomenta el desarrollo tecnológico con el fin de modificar y expandir las capacidades físicas y psíquicas del ser humano hasta alcanzar un estadio que lo trascienda como tal, no siendo ya un humano, sino un transhumano, llegando incluso a la inmortalidad. Se trata de una ideología conforme a la cual, la humanidad posee, en el desarrollo tecnológico, la llave para la superación de sus límites biológicos. En ese marco, la tecnología deja de ser un mero instrumento para la consecución de un fin, sino el fin mismo. Pero ese fin tecnológico es, asimismo, medio para una finalidad ulterior: la transformación del ser humano, hacia formas hasta ahora sólo conocidas en la ciencia ficción, tales como prótesis que aumenten las capacidades físicas y psíquicas del sujeto, organismos cibernéticos, trasplantes de conciencia, la creación de memorias en soportes digitales y demás extravagancias que, aunque son imaginables, aún resultan imposibles.

El transhumanismo implica un constructivismo radical, es decir que, desde este enfoque, no hay una realidad objetiva conocida por un sujeto, sino que éste la construye a través de la percepción y la interpretación.

De este modo, el conocimiento es construcción de la realidad; un conocimiento que no contempla al objeto, sino que lo constituye. Así, habría tantas realidades como sujetos cognoscentes. Esa característica performativa del conocimiento permitiría, asimismo, la adecuación de esa realidad por parte del sujeto mediante el lenguaje, en virtud de la carga interpretativa y valorativa que éste siempre conlleva. En lo que aquí interesa, esta noción subjetiva de la realidad permite tomar la percepción como equivalente a lo real. De este modo, aspectos como la sexualidad –en cuanto realidad fáctica– dejan de ser datos objetivos, cognoscibles por cualquier observador, para ser una *construcción* del propio sujeto conforme a la percepción que tenga de sí mismo. En otras palabras, uno no será varón o mujer conforme a la biología, sino en función a de su autopercepción, combinando conceptos de sexualidad con categorías psicológicas referidas a la orientación sexual, prácticas sexuales y, más recientemente, a la “expresión de género”: varón, mujer, intersexual, transexual, transgénero, género fluido, género neutro, andrógino, bisexual, y un sin fin de posibles autopercepciones de cada sujeto (Chavarría Alfaro, 2013). Este cambio de paradigma humano y social, según Luc Ferry, nace del campo terapéutico.

### Caracteres del Transhumanismo según Nick Bostrom y Max More (2002)

Tanto para Nick Bostrom como para Max More, el transhumanismo se caracteriza por las siguientes notas:

1. La Humanidad cambiará profundamente mediante la ciencia y la tecnología del futuro.
2. Todavía no se ha realizado lo *esencial* de la Humanidad. Nos esperan formas maravillosas y extremadamente interesantes.
3. Es cierto que reconocemos que la Humanidad está enfrentada a grandes riesgos en la utilización masiva de las nuevas tecnologías.
4. Debemos debatir cómo reducir los riesgos y quedarnos solo con lo constructivo.
5. Se debe primar, en lo financiero, la preservación de la vida y la salud, la mejora de los sufrimientos graves y la prevención de ellos.
6. Las decisiones políticas deben estar guiadas por una visión moral responsable y socializadora, con vistas al futuro.
7. Defendemos el bienestar de todas las inteligencias: las humanas, las no-humanas, los animales, las futuras inteligencias artificiales, las formas de vida modificadas y todas aquellas nacidas del progreso tecnológico y científico.
8. Promovemos la libertad morfológica, es decir, el derecho de modificar el cuerpo, el conocimiento y las emociones: “Llegaremos a crear una máquina que tenga el nivel de conciencia del hombre, más allá de lo que puedan creer las posiciones *reaccionarias inspiradas en creencias religiosas y antiprogresistas*”.

El transhumanismo pretende:

1. En lo eugenésico: una nueva ética, la del pasar del *azar* a la *elección* (*from chance to choice*).
2. Un antinaturalismo: El progreso indefinido influirá sobre la naturaleza biológica, más allá de las reformas políticas sociales.
3. La inmortalidad: Llegaremos, por la ciencia, a una vida sin fin.
4. La fe en un optimismo tecnocientífico navegue a toda prueba.
5. El desarrollo de en un *racionalismo materialista, determinista y ateo*.
6. Una ética utilitarista y libertaria, que navegue entre el neoliberalismo y la socialdemocracia.
7. Una ideología “deconstruccionista”, igualitarista y proecologista.
8. La creación de foros democráticos y de discusión ética.

Por esto se produce un debate entre “Bioconservadores” y “Bioprogresistas”.

Sobre la posibilidad de una vida “sin fin” y sobre la consideración de un “hombre-máquina”. Así el Transhumanismo se presenta como *la búsqueda del mejoramiento humano a través de la tecnología*. Es la búsqueda de la

transformación del hombre mediante la integración del ser humano con la máquina (cyborg), mediante la transformación de nuestros genes en la línea germinal, hasta poder generar una “especie distinta”, “posthumana”.

El ser humano toma el control de su propia evolución: *evolución diseñada*, incluso está moralmente obligado. La *evolución darwiniana*, basada en la selección aleatoria, está finalizada. Ahora comienza la *evolución basada en la tecnología* que nos permitirá seleccionar nuestras cualidades físicas y mentales, el rejuvenecimiento y la inmortalidad

Este paradigma transhumanista tiene dos modalidades:

- **Cultural:** Inspirado en la crítica posmoderna al Humanismo (Foucault, Derrida, Poscolonialismo, Ecología radical). Donna Haraway, *Manifesto cyborg* (1985) “A finales del siglo XX –nuestra era, un tiempo mítico–, todos somos quimeras, híbridos teorizados y fabricados de máquina y organismo; en pocas palabras, somos *cyborgs*.”
- **Tecnocientífico:** Inspirado en los trabajos de científicos e ingenieros provenientes del campo de la IA y de la robótica (Marven Minsky, Hans Moravec, Raymond Kurzweil). Basado en los avances de la Biotecnología.

La idea compartida por ambos es la superación de la barrera entre el hombre y la máquina y entre lo real y lo virtual.

Los tipos del mejoramiento humano pueden ser de tres tipos:

1. **Químico o farmacológico**
2. **Cibernético o digital**
3. **Genético**

Cada uno puede estar dirigido al mejoramiento físico, mental, emocional o moral para mejorar la *calidad de vida o bienestar*, pero no hay acuerdo sobre este tema.

### Argumentos del Transhumanismo

Los argumentos para el mejoramiento humano son:

- La tecnología ha sido siempre un instrumento de mejoramiento humano que ha posibilitado el desarrollo de la cultura gracias a las cuales disfrutamos de la escritura, la ciencia, el arte, “prótesis culturales” que amplían nuestras capacidades cognitivas. Hoy serían la *inteligencia artificial* o la *robótica*.
- No intentar mejorarnos es faltar a un deber moral, causando un perjuicio objetivo. Poner límites a la libertad del mejoramiento (a los padres que quieren mejorar a sus hijos) es ir contra el *Principio de beneficencia procreativa* (Savulescu, AÑO).
- Es legítimo permitir el mejoramiento genético con una finalidad *terapéutica*, pero limitar o prohibir cuando la finalidad es *mejoradora*. ¿Pero cómo es posible trazar la línea entre ambas?
- Nadie se opone al mejoramiento humano (intelectual, psicológico,

moral o físico) mediante técnicas sociales o culturales como la educación y el deporte.

• *El mejoramiento es un fin deseable pero el medio para conseguirlo no es éticamente relevante* (ya sea por la educación o por la ingeniería genética).

### ¿Cuáles son las *mediaciones* del Transhumanismo para lograr sus fines?

Los medios o instrumentos para lograr el objetivo que este modelo se propone son:

1. La *nanotecnología*, que será capaz de reparar, en el campo terapéutico, molécula a molécula.
2. La *biotecnología*, que será capaz de manipular nuestro ADN.
3. El *Big Data* o recopilación de datos para determinar los gustos del consumidor.
4. La *Inteligencia artificial (IA)*, capaz de competir con la humana.
5. Las *neurociencias* cognitivas que investigan los procesos cerebrales de manera integrada desde el nivel molecular hasta el ambiente social y cultural (Facundo Manes).

### Conclusiones

Desde el punto de vista ético: ante los retos del Transhumanismo no podemos caer ni en la libertad absoluta ni en el control total, sino que hay que buscar una “regulación” para poner límites que no puedan traspasarse. “La palabra *hombre* se volvió *extraña*”, dice Jean Luc Nancy (AÑO).

El desquicio viral nos desvía de nuestro supuesto destino anterior de sobrehumanos o *transhumanos* y nos retorna a nuestra condición de “bípedos implumes” pero con el don del lenguaje y el pensamiento para sondear el abismo del contagio.

Entonces:

- El Humanismo Cristiano tuvo el desafío de dialogar en el siglo XX con el Humanismo existencialista ateo.
- Hoy tiene el desafío del diálogo con el Transhumanismo.
- “Se vuelve actual la necesidad imperiosa del humanismo, que de por sí convoca a los distintos saberes, también el económico, hacia una mirada más integral e integradora” (Laudato Sí, (año), p.141).

## Referencias bibliográficas

- Bárbaro, J. (2021). *Juicio a los 70*. Sudamericana.
- Bostrom, N. (2003). *Intensive Seminar on Transhumanism*, Yale University.
- Berzosa Martínez, C. R. (2019). “¿De homínidos a post-humanos? Algunos desafíos del transhumanismo”. *Razón y Fe*, 279 (1437), pp. 81-92.
- Chavarría Alfaro, G. (2013). “El posthumanismo y el transhumanismo: transformaciones del concepto de ser humano en la era tecnológica”. *Informe final de Investigación*. Instituto de la Universidad de Costa Rica.
- Derisi, O. N. (1969). *Naturaleza y vida de la universidad*. EUDEBA.
- Ferry, L. (2017). *La Revolución Transhumanista. Cómo la tecnomedicina y la uberización del mundo van a transformar nuestras vidas*. Alianza Editorial.
- Husserl, E. (1995). “Carta a Lévy-Bruhl”. *ER. Revista de Filosofía*, 19, año X, pp. 171-175. Trad. J. de San Martín.
- Nussbaum, M. (2002) “Genética y Justicia: Tratar la enfermedad, respetar la diferencia.” *ISEGORÍA* 27, pp. 5-17.
- Pouliquen, T. M. (2018). *Transhumanismo y fascinación por las nuevas tecnologías*. Rialp.
- Quiles, I. (1954). *Introducción a la Filosofía*. Editorial Estrada.
- Sandel, M. (2007). *Contra la perfección. La ética en la era de la ingeniería genética*. Marbot.
- Sartre, J. P. (1977). *El existencialismo es un Humanismo*. Ediciones Huascar.
- Sloterdijk, P. (2007). “Normas para el Parque humano. De la carta sobre el humanismo a las antropotecnias y el discurso del posthumanismo”. *Revista Observaciones filosóficas*, n. 5.

## Digesto 1.1.10.1 como método para la valoración de la capacidad intelectual algorítmica

Ian Henríquez Herrera  
Roxana Álvarez Sanguinetti

El *Digesto* es el libro de mayor importancia, relevancia e influencia en la historia del Derecho. Corresponde a la compilación del saber jurídico de la época romana clásica, ordenada realizar por el emperador Justiniano en el siglo VI. Muchas de las soluciones allí contenidas, basadas en un fino sentido común, siguen plenamente vigentes y son de uso común y general, desde luego en Europa y América, pero también en Asia y África. De algún modo, el *Digesto* constituye el libro por excelencia de la disciplina jurídica. Su sola denominación así lo atestigua.

Dicho texto, en su libro primero, recoge una fórmula breve y sencilla que contiene un gran acervo de sabiduría jurídica. Dice, a la letra: “*Iuris praecepta sunt haec: honeste vivere, alterum non laedere, suum cuique tribuere.* (D.1.1.10.1), que traducido a nuestra lengua corresponde a: los preceptos del derecho son vivir honestamente, no dañar a otro, dar a cada cual lo suyo. La especial importancia de este último mandato se entrevé al estar ya anunciado en el párrafo precedente (D.1.1.10), que define la justicia del siguiente modo: *Iustitia est constans et perpetua voluntas ius suum cuique tribuendi*, esto es, la justicia es la voluntad perpetua y constante de dar a cada cual lo suyo. La caracterización del acto justo como dar a cada quien lo suyo está ya presente en la *Politeia* de Platón (331d) y remite a un verso del poeta griego Simónides, cuyo texto íntegro se ha perdido, sólo conservándose dicho fragmento.

Hay aquí, entonces, un aspecto muy central del oficio del jurista. Lo propio del saber y quehacer jurídico es dar a cada cual lo suyo. Eso requiere, por lo pronto, mirar con atención la realidad, aceptar su complejidad e intentar conocerla del mejor modo posible. Si la asimilación del acervo cultural es una llamada, cuando no una exigencia, a toda persona, en el caso de quien se consagra al derecho es un requisito de su arte. Sólo se puede dar a cada cual lo suyo si se conoce qué es lo propio de cada cual, y por consiguiente, qué es lo debido para con éste. Armando Roa –reconocido intelectual chileno, fundador del Centro de Bioética de la Universidad de Chile, y a quien Picasso, en linda metáfora, le llamara “águila y sol”, solía afirmar que lo propio de una persona culta es darse cuenta del valor de lo que tiene al frente. Similar es lo que se deriva del prólogo de Wilde a *El retrato de Dorian Gray*<sup>1</sup>.

No darse cuenta de lo que se tiene al frente, por cierto, para cualquier persona puede traer consecuencias indeseadas. Pero en el caso del jurista,

<sup>1</sup> “*Those who find beautiful meanings in beautiful things are the cultivated*”.

implicaría el ejercicio incorrecto de su arte y oficio. Dicho sea de paso, por eso el saber filosófico –que se ocupa de un modo privilegiado de lo que las cosas son– es tanpreciado y relevante para el adecuado ejercicio del derecho.

Al ver las fotografías de los zoológicos humanos de los siglos XIX e inicios del XX, tan bien documentados por Báez y Mason (2006), se puede apreciar con claridad las consecuencias de no saber qué es lo propio de cada cual. Un selknam o un yámana no son ni merecían trato de animales, ni tampoco de objeto de la ciencia. Tratar a un humano como cosa es nefando. Y tratar a una cosa como humano es ridículo.

Llegados a este punto, lo que en esta ocasión nos ocupa dice relación con las asombrosas e impresionantes capacidades que muestra el desarrollo tecnológico contemporáneo por medio de la algorítmica compleja, usualmente denominadas y generalmente conocidas como “inteligencia artificial”. La denominación, que navega entre la desmesura y el oxímoron y que incluso entre juristas de prestigio ha sido calificada de *vintage* (Cippitani, 2025), ya ha tenido validación normativa, por medio de la reciente promulgación de la regulación europea (*Artificial Intelligence Act*, de 2024). Junto con ello, en la industria proliferan denominaciones referidas a las capacidades algorítmicas, tales como *smart robot*, *conscious robot*, *machine learning*, *deep learning*, etc. Si el acto de justicia (D.1.10) importa dar a cada quien lo suyo, y por extensión tratar a cada realidad conforme cada realidad merece, hay allí una clave para este análisis, que podría llevar a considerar un uso abusivo del lenguaje la extensión de tales facultades propiamente humanas al funcionamiento algorítmico, por una parte, y como consecuencia, un trato indebido e injusto para con los humanos.

En efecto, autores como Alejandro Serani Merlo han sostenido:

Con respecto al supuesto desempeño cognitivo de los nuevos desarrollos tecnológicos llamados de ‘inteligencia artificial’, nos parece que es necesario descartar de entrada el que algunos de los cometidos de las tecnologías (...) puedan ser designados de modo propio como ‘inteligentes’ (Serani, 2023).

La estatura intelectual del recién citado debería llevarnos a considerar muy seriamente esta observación. Por lo pronto, si consideramos el término *inteligencia* en su acepción más propia, conectada con su etimología de penetrar las cosas, leer hacia dentro lo que las cosas son –*intus legere*–, parece claro que dicha cualidad es ajena al artefacto. La seriedad de la observación se incrementa si consideramos lo sostenido por Kate Crawford, profesora de University of Southern California e investigadora de Microsoft:

La IA (inteligencia artificial) (...) no es inteligente en ninguno de los sentidos de la inteligencia humana. No es capaz de discernir las cosas sin un amplio entrenamiento humano, y tiene una lógica estadística completamente diferente sobre cómo se crea el significado. Desde los inicios de la IA en 1956, hemos cometido este terrible error, una es-

pecie de pecado original en este ámbito, de creer que las mentes son como las computadoras y viceversa. Suponemos que estas cosas son análogas a la inteligencia humana, y nada podría estar más lejos de la verdad ( Crawford, citado en Simonete, 2021, párr. 4).

Ahondando en el punto, Daniel Innerarity ha sostenido que:

La inteligencia artificial es un conjunto de técnicas geniales para aprenderse el mundo de memoria. Aunque sobrepase la potencia calculatoria del ser humano, la inteligencia artificial es incapaz de dar una significación a sus propios cálculos. ¿Se puede ser inteligente sin saberlo, como un zombi que fuera capaz de realizar tareas inteligentes de un ser humano pero solamente de manera refleja? La inteligencia artificial tiene inteligencia refleja, no reflexiva. Y esto no corresponde a la noción que tenemos de inteligencia ( Innerarity, 2019, párr. 1).

Y de manera sugerente sostiene: “que la actual inteligencia artificial sería muy inteligente y muy estúpida a la vez; su estupidez consiste en que cuando toma una decisión inteligente no tiene modo de saberlo” (Innerarity, 2019).

Intentando profundizar en el mismo filón, Han ha sostenido las limitaciones intrínsecas y estructurales de la algoritmia compleja:

El pensamiento humano parte de una totalidad que precede a los conceptos, ideas y la información. Se mueve en un *campo de experiencia* antes de dirigirse específicamente a los objetos y hechos que encuentra en él. Este campo de experiencia se abre inicialmente en una disposición anímica, es decir, en una disposición emocional. Antes de dirigirse hacia algo, el pensamiento humano ya se encuentra en una disposición anímica básica. Esta disposición anímica caracteriza al pensamiento humano y es lo que nos permite pensar.

La inteligencia artificial, por otro lado, no puede experimentar esta disposición anímica. Para que el pensamiento sea posible, es necesario que los pensamientos y enunciados sean obtenidos desde una disposición anímica fundamental. Esta disposición anímica nos permite pensar y dar sentido a nuestro trabajo con los conceptos.

El pensamiento humano es un proceso analógico porque guarda correspondencia con una voz que lo determina y sintoniza con él. No se ve interpelado por un ente en particular, sino por la totalidad del ser. La inteligencia artificial, por su parte, es incapaz de escuchar esa voz. Es sorda a las llamadas del ser y no puede establecer una correspondencia con la totalidad del ente (Han, 2021, pp. 30-31).

Como ya ha sido notado previamente (Carbajales y Henríquez, 2024), la literatura especializada tiende más bien a la cautela y al escepticismo frente a los verdaderos avances de la algoritmia compleja, en comparación con la opinión generalizada instalada en la cultura de masas (Lacruz, 2020, p. 248). Desde luego, mirado con rigurosidad, los fenómenos de conciencia, inteligencia, libre albedrío, aprendizaje, todos en el uso propio del lenguaje, no son predicables respecto de una máquina. Esta afirmación elemental, de todo sentido común, se torna opaca precisamente por el pro-

gresivo arraigo de expresiones ya usuales en la industria (como las ya referidas *artificial intelligence*, *smart robot*, *conscious robot*, *machine learning*, etc.). La complejidad técnica y alta especialización, añade un grado más a la referida opacidad, y contribuye al decante de una falsa percepción: “el usuario se ‘acostumbra a una tecnología opaca’, la maquinaria digital ‘tras la pantalla’ se vuelve totalmente impenetrable, incluso invisible” (Zisek, 2006). En atención a ello, precisamente, debemos bregar por alcanzar claridad en este aspecto.

Martin Ebers, de la Humboldt-Universität zu Berlin, sostiene que en este ámbito, referido a las máquinas y los algoritmos, “no se debería en absoluto hablar de inteligencia y, sobre todo, de acciones autónomas”. (Ebers, 2016, p. 8).

Asimismo, Wolf Singer, del Max Planck Institute for Brain Research, sostiene que “la intencionalidad, la moralidad, la responsabilidad y ciertos aspectos de la conciencia, como los qualia de la experiencia subjetiva, pertenecen a la dimensión inmaterial de las realidades sociales” (Ebers, 2016, p. 8). Ahora bien, en específico sobre la pertinencia de hablar de conciencia en máquinas, Markus Gabriel, de la Universidad de Bonn sostiene: “los robots no son conscientes y no podrían ser conscientes” (Gabriel, 2021, p. 67). Sostener lo contrario obedecería, según el autor, a “delirios de la llamada cosmovisión científica” (Gabriel, 2021, p. 67). Desde luego, el modo de hacerle frente a esos delirios es por medio de una adecuada formación humanística, que puede hacerse cargo de la complejidad del fenómeno y aportar profundidad en la mirada. Concordamos con Gabriel cuando sostiene:

Mientras no otorguemos a las humanidades y a todas las demás disciplinas académicas la misma posición epistemológica, las ciencias naturales también serán presa fácil para aquellos que no se preocupan por los hechos, pero realmente están interesados solo en maximizar el alcance de su voluntad de poder”; concluye: “tenemos derecho a rechazar la idea misma de los robots conscientes (Gabriel, 2021, p. 67).

Autores como Zimmermann y Cremers también han contribuido a morigerar las ínfulas tecnológicas, con una aproximación más realista sobre los límites intrínsecos de la algoritmia, al mostrar que, en dicho campo, las relaciones aprendidas son de asociación y no causales. Por consiguiente, cuando las acciones basadas en predicciones se retroalimentan de un modo significativo, el aprendizaje de asociación no puede responder a las preguntas relevantes surgidas en relación con las consecuencias (Zimmermann, 2021, p. 40). En el mismo sentido, Timothy Gowers, matemático premio Fields, con una fina y escéptica ironía ha dicho que “desafortunadamente, mientras los computadores continuamente nos sorprenden con aquello para lo cual pueden ser usados, casi nada sabemos acerca de lo que no pueden hacer” (2002).

Ya no desde la matemática, sino desde la lingüística, es conocida la posición de Chomsky:

Por muy útiles que puedan ser estos programas en algunos ámbitos concretos (pueden ser útiles en la programación informática, por ejemplo, o para sugerir rimas para versos ligeros), sabemos por la ciencia de la lingüística y la filosofía del conocimiento que difieren en gran medida de la manera en que los seres humanos razonamos y utilizamos el lenguaje. Estas diferencias imponen limitaciones significativas a lo que estos programas pueden hacer, codificándolos con defectos imposibles de erradicarse (Chomsky, 2023).

De hecho, estos programas están estancados en una fase prehumana o no humana de la evolución cognitiva. Su defecto más profundo es la ausencia de la capacidad más crítica de cualquier inteligencia: decir no solo lo que es el caso, lo que fue el caso y lo que será el caso (eso es descripción y predicción), sino además lo que no es el caso y lo que podría y no podría ser el caso. Esos son los ingredientes de la explicación, la marca de la verdadera inteligencia (Chomsky, 2023).

Digámoslo, entonces, con claridad: los algoritmos no inteligen. Tampoco aprenden ni son conscientes. Los algoritmos no agencian ni actúan, como tampoco el programa a través de los que se estructuran ni el ordenador que sirve de soporte.

Ahora bien,

(...) como han hecho notar los expertos, cuando se opera con máquinas hiper complejas, que interactúan con nosotros en la vida cotidiana, es muy difícil no proyectar antropomórficamente la existencia en ellas de un agente o de una subjetividad. Está claro, sin embargo, que en estos casos se trata de una 'ficción'. Esta ficción es la mayor parte de las veces inocente e inocua, pero también puede inducir a engaños (Serani, 2023).

Pensemos, por ejemplo, en el caso del robot Sophia, una maravilla tecnológica, una máquina antropomorfa altamente sofisticada y con un amplio elenco de funcionalidades, que incluye el diseño para interactuar con humanos, como si fuera uno. Tal artefacto ha sido, incluso, entrevistado en programas de televisión, y se le han hecho preguntas de índole como la que sigue: “Sophia, ¿tienes sentimientos?”; “¿Sophia, puedes tomar decisiones morales”? Bastaría que el robot dijese “sí” para que el interlocutor asumiese que ello es de esa manera. Y desde allí para el salto a la percepción general de la opinión pública, el tramo es poco: “Sophia, el robot que siente”; “Sophia, el robot que toma decisiones morales”. Y el desvarío puede continuar e incrementarse: “La ciencia ha roto nuevos límites y alcanzado nuevas fronteras: los robots sintientes y agentes morales”, etc., etc., etc. Por favor, no dejemos de considerar que a Sophia ya le ha sido otorgada la ciudadanía saudí.

La fantasía tecnocrática reposa, por lo pronto, en la ideología del cientificismo, pero también en la creciente y progresiva ignorancia del ser humano sobre sí mismo y en el deficiente uso y desarrollo de la facultad crítica o estimativa de las personas. Y una facultad que no se usa, se atrofia. Sólo quien ignora verdaderamente quién y cómo es, y se torna incapaz de introspección, perdiendo de vista la vastedad, riqueza, complejidad e inconmensurabilidad de la interioridad del sí mismo, podría llegar a pensar que una máquina es susceptible de equipararle. Dicho de otra manera, quien afirme que una máquina es inteligente, algo nos está mostrando de su propia inteligencia.

En este aspecto, Nuestra América tiene la posibilidad, cuando no el deber, de realizar una especial aportación en el debate. Contamos con una doble ventaja respecto del primer mundo angloparlante. Primeramente, hablamos y habitamos en la lengua de san Juan de la Cruz, de Mistral, de Vallejo, Cortázar y Storni; lengua que permite una amplitud y profundidad al pensamiento que puede ir mucho más allá de la mera analítica. La distinción entre ser y estar nos abre hacia otras perspectivas y dimensiones en el razonamiento<sup>2</sup>. El inglés, por el contrario, es una lengua más fácilmente anclada en el fenómeno, en la materialidad. Piénsese, tan sólo en los términos a través de los que se expresan los sustantivos *vaso* y *bosque* en dicha lengua, esto es *glass* y *wood*, que son las mismas palabras para expresar vidrio y madera. Parece clara la orientación fáctica en las expresiones referidas a nada y nadie (*nothing, nobody*). Nótese, asimismo, cómo la lengua inglesa distingue con más claridad y simpleza que el español entre *to do* y *to make*, lo que refrendaría el énfasis en la facticidad y la acción exterior (ambos verbos que solemos en español reconducir a uno solo: *hacer*).<sup>3</sup>

Intentando ahondar en este aspecto, y en conexión con lo anterior, el español parece ser un idioma más rico en la descripción de la interioridad del ser humano, y también más profundamente anclado en ésta. Tomemos, por ejemplo, la diferencia entre *querer*, *desear* y *amar*. Para cualquier hispanoparlante, el contraste es intuitivo, ya que *querer* es mucho más íntimo que *desear* y mucho más suave que *amar*, mientras que *desear* es más intenso que *querer* pero más inherentemente físico que los otros dos, y *amar* es el descriptor más intenso y más íntimo de los tres. Sin embargo, el hecho de que *querer* y *desear* compartan un significado sugiere que *querer* a una persona implica, de cierta manera, desear algo de o para ella (ej. su compañía, su bien, etc.), una asociación semántica que *amar* no comparte. Los sentidos de estas palabras están profundamente entrelazados con la interioridad de la persona, ya sea desde la intensidad de su sentimiento

2 La noción de que la lengua materna de una persona afecta la forma en que conceptualiza el mundo proviene de la hipótesis de Sapir-Whorf (así nombrada por el antropólogo Edward Sapir y su alumno, el lingüista Benjamin Whorf, quien desarrolló las ideas de su profesor tras la muerte de éste), también conocida como la teoría del relativismo lingüístico.

3 Sobre estas comparaciones lingüísticas, puede verse, *inter alia*: Mendoza Sánchez, M. (1969), pp. 409-416, especialmente p. 414; Valenzuela Manzanares, J. (2002), pp. 27-45.

o a partir de la naturaleza de éste. Por el contrario, dichos matices no son recogidos en el verbo *love*. Desde luego, el verbo *want*, en este contexto, no presenta relación alguna con *love*. Otro tanto ocurre con *wish*. Los significados de estas palabras están anclados a la relación entre sujeto y objeto, entre la persona y el mundo que la rodea. Nótese la connotación física y casi posesiva de *want* en inglés –*to want someone*– es desear a alguien en sentido predominantemente físico, con una cierta reducción o cosificación, no con el cariño inherente a *querer*. Por su parte, *love* posee una cualidad más sentimental, ya que es una forma de mostrar afecto, cuya intensidad, sin embargo, se da a entender por circunstancias externas, no por la palabra en sí. Un niño agradecido puede decirle a un profesor *I love you*, como simple expresión de afecto o cariño. Los ejemplos podrían multiplicarse.<sup>4</sup> Por lo tanto, para nosotros, hispano hablantes, pareciera que es más fácil mirar, expresar y comprender la complejidad de nuestra geografía interior que para un anglosajón, en el supuesto que el elenco lingüístico sería más amplio.

La segunda ventaja que Nuestra América presenta en el tratamiento de estas cuestiones, es que, al parecer, los nuestroamericanos tendríamos una capacidad mayor para contemplar más allá de la materialidad, probablemente heredada de las culturas originarias y del influjo del catolicismo. En efecto, tenemos una facultad muy desarrollada para percatarnos de la realidad que excede el mero fenómeno externo, lo que Walter Benjamin denominaba “el aura” de las cosas (1969).

La industria tecnológica, en general, y asociada a los desarrollos algorítmicos, en particular, ha estado anclada en Occidente a la órbita anglosajona. Las características de su lengua es una posible causa de que tiendan a mirar la materialidad y externalidad de los fenómenos de un modo preeminente. Ello podría explicar la tendencia a calificarlos a partir de sus manifestaciones. Si se comporta externamente como tal y tiene la apariencia de tal, entonces es tal. Sin ir más lejos, esa es la estructura del famoso test de Turing. La agencia misma, con toda su riqueza y complejidad, desde la intención hasta la deliberación y decisión, acaecidas en la interioridad, es desprovista de toda consideración.

Cualquier observador imparcial, que cuente con una facultad crítica –*una vis aestimativa*– madura y suficientemente desarrollada, con un mínimo conocimiento de sí mismo, de su vasta y compleja interioridad, y con

---

4 Desde luego, no es este el objeto central de nuestro ensayo, pero sólo para graficar el punto, *intention* está relacionado de un modo directo con *agency*, es decir, toda acción, para ser propiamente tal, ha de ser intencional. La intención se agota en su totalidad en el acto mismo. La *intención*, en cambio, no sólo está en el origen del acto, en la agencia del agente, sino que, además, le cualifica moralmente, puesto que se proyecta más allá del acto mismo hacia sus consecuencias. Es lo que permite al hispanoparlante señalar que un acto (acción intencional) no fue con intención, cuando produce efectos que desbordan lo previsto o intentado. Así, si Anscombe hubiese desarrollado sus investigaciones sobre la intención en nuestra lengua, probablemente sus hallazgos podrían haber ido mucho más lejos.

un razonable grado de información sobre las características de la algoritmia compleja y sus manifestaciones, percibe de inmediato las inconmensurables diferencias entre una máquina y una persona humana, por muy compleja y sofisticada que sea la primera.

Desde luego, las usuales y cotidianas licencias en el uso del lenguaje no resultan problemáticas. Las analogías, metáforas, ficciones y aun imprecisiones o errores son parte de nuestro hablar corriente. El problema puede suscitarse cuando a esos usos analógicos, metafóricos, ficticios, imprecisos o erróneos, se les dota de entidad y criterio de realidad. Allí hay un límite relevante. En ese contexto, cobra importancia lo sostenido por Jaron Lanier, uno de los pioneros de la realidad virtual:

El negocio de INCORDIO está entrelazado con una nueva religión que otorga empatía a programas de ordenador –al llamarlos «programas de IA»– con el fin de soslayar que eso degrada la dignidad, el estatus y los derechos de humanos reales» (Lanier, 2018, p. 97).

Aquí, entonces, entramos a un punto crucial de nuestra reflexión. Volvamos a Digesto 1.1.10.1. ¿Qué es lo debido a la persona? Por lo pronto, como es obvio, un trato de tal. Y lo que, quizás no resulte hoy por hoy tan obvio, es la exclusividad en dicho trato. El tratamiento a máquinas de un modo que sólo le es debido al humano, degrada al humano. Ese es el riesgo relevante de la humanización de los algoritmos complejos y sus soportes físicos. No es purismo en el lenguaje, ni obsesión metafísica de un realismo ingenuo. Es mucho más sencillo –el derecho es un arte modesto–. Se trata de hacer justicia y dar a cada uno lo que le es debido.

## Referencias bibliográficas

- Báez, C., Mason, P. (2006). *Zoológicos humanos. Fotografías de fueguinos y mapuche en el Jardín d'Acclimatation de París, siglo XIX*. Editorial Pehuén.
- Benjamin, W. (1969). “The Work of Art in the Age of Mechanical Reproduction”. En Arendt, H. (ed). *Illuminations*, Schocken Books.
- Carbajales, M., Henríquez, I. (2024). “Personalidad electrónica de ‘robots inteligentes’ como técnica para la reparación de daños. Análisis crítico de las normas de derecho civil sobre robótica del Parlamento Europeo”. *Prudentia Iuris*, N° 98.
- Cippitani, R. (2025). *Derecho e inteligencia artificial*. Conferencia pronunciada en la Universidad degli Studi di Perugia, enero 2025, Asís. Inédita.
- Chomsky, N. (8 de marzo de 2023). “The false promise of Chat Gpt”. *The New York Times*.
- Ebers, M. (2016). “La utilización de agentes electrónicos inteligentes en el tráfico jurídico: ¿Necesitamos reglas especiales en el Derecho de la

- responsabilidad civil?”. *InDret Revista para el análisis del derecho*, N°3.
- Gabriel, M. (2021). “Could a Robot Be Conscious? Some Lessons from Philosophy”. En J. von Braun et al. (eds.). *Robotics, AI, and Humanity*. Springer.
- Gowers, T. (2022). *Mathematics: A Very Short Introduction*. Oxford University Press.
- Han, B. C. (2021). *No cosas*. Taurus.
- Innerariti, D. (2019). “La inteligencia de la inteligencia artificial”. *Daniel Innerariti*. <https://www.danielinnerarity.es/opini%C3%B3n-pre-blog-2017-2022/la-inteligencia-de-la-inteligencia-artificial/>
- Lacruz, M. (2020). *Robots y personas. Una aproximación jurídica a la subjetividad cibernética*. Reus.
- Lanier, J. (2018). *Diez razones para dejar tus redes sociales de inmediato*. Debate.
- Mendoza Sánchez, M. (1969). “Inglés y español: lengua y pensamiento”. *Hispania*, sept. vol. 52, N° 3.
- Platón (1981). *Obras completas*. Aguilar.
- Serani Merlo, A. (2023). “Inteligencia natural: en la vida social y en la medicina”. Conferencia pronunciada en Sanatorio Alemán, Chile.
- Simonete, T. (2021). “This Researcher Says AI Is Neither Artificial nor Intelligent. Entrevista a Kate Crawford”. *Wired*. <https://www.wired.com/story/researcher-says-ai-not-artificial-intelligent/>
- Valenzuela Manzanares, J. (2002). “Lingüística contrastiva inglés-español: una visión general”. *Carabela*, N° 51, pp. 27-45.
- Zimmermann, J., Cremers, A. (2021). “Foundations of Artificial Intelligence and Effective Universal Induction”. En Von Braun, J. et al. (eds.). *Robotics, AI, and Humanity*. Springer.
- Zizek, S. (2006). *Lacrimae rerum*. Debate.

# **Aipóiesis Transposición y Mímesis entre Narrativas de la Ciencia Ficción y el Realismo en Inteligencia Artificial**

Gustavo Ángel Riesgo

## **Introducción**

La Ilustración comenzó con ideas esencialmente filosóficas difundidas por una nueva tecnología. Nuestra época avanza en la dirección opuesta. Ha generado una tecnología potencialmente dominante en búsqueda de una filosofía que la guíe (Kissinger, 2018)<sup>1</sup>

Desde las pioneras novelas de autores como Verne y Wells, hasta el apogeo del género con Asimov, Bradbury y Heilein y luego el ciberpunk con la llegada de sus derivados actuales, la literatura de ficción científica ha evidenciado la compleja relación del hombre con la tecnología en base al pensamiento de cada época (Canavan, G., Link, E., 2019). En particular, el vínculo entre algunas de estas narrativas con el desarrollo de la inteligencia artificial (IA) ha sido, durante casi los últimos cien años, un espejo que refleja las aspiraciones y temores más profundos de la humanidad. Asimismo, desde casi los albores del arte cinematográfico hasta las producciones contemporáneas, distintos tipos de IAs han sido retratados como fenómenos de condición bipolar: algunas pocas veces, como aliados; las otras, en su mayoría, como amenazas que ponen en jaque la continuidad de la existencia humana (Canavan, G., Link, E., 2019, pp. 703-705). En este contexto, el presente trabajo intenta explorar el diálogo de mutuo aporte entre algunas narrativas de ficción científica y la evolución tecnológica de las IAs y cómo, el cruce de estas dos dimensiones, afecta al imaginario cultural.<sup>2</sup>

El término *aipóiesis* sugiere una actividad creativa en la que es central la interacción entre el desarrollo de IA, que aporta sus tópicos, y la ficción científica, que los desafía. Asimismo, muchas de las narrativas de ficción científica han servido como un medio para proyectar futuros posibles, funcionando así no sólo como literatura sino también como experimentos

---

1 En el original: “The Enlightenment started with essentially philosophical insights spread by a new technology. Our period is moving in the opposite direction. It has generated a potentially dominating technology in search of a guiding philosophy”, (Kissinger, 2018).

2 Si bien en el título de este artículo se mantuvo la denominación corriente de “ciencia ficción”, para el cuerpo del mismo se prefiere el de “ficción científica” que representa una traducción y uso más correctos.

mentales que someten a los límites de la tecnología y a la ética a una prueba imaginaria. Transposición y mimesis son términos tomados a préstamo desde otros campos a efectos de describir dos fenómenos en interdependencia. Funcionan como “transposiciones” aquellas narrativas que elaboran exageraciones en el alcance de la ciencia y la tecnología de su época instalando así desafíos con diversos grados de factibilidad. Complementariamente, la “mimesis” opera cuando la ciencia y tecnología reales son asumidas para la ficción científica o su inversa que modela desarrollos posibles (Mubin, 2019). La idea de *aipóiesis* también remite, aunque no sea más que intencionalmente, al concepto de *sub-creación* acuñado por Tolkien (1977). En su visión, la *aipóiesis*, no es un mero acto de invención, sino una forma de descubrir verdades más profundas a través de la fantasía y de la ficción. En el contexto de la inteligencia artificial y las narrativas de ficción científica, la sub-creación puede interpretarse como la capacidad de las personas para producir artefactos que reflejan, pero no igualan –mucho menos superan–, la naturaleza humana. Podría decirse que se refiere a la creación de *imágenes asemejantes*: objetos artificiales que imitan lo personal, pero que no lo alcanzan a replicarlo, ni siquiera parcialmente.

La atribución mental, la atribución de agencia y la atribución semántica, emergen como tópicos claves del tema. Respecto de éstas, es preciso reconocer una tendencia humana a proyectar conciencia, intenciones y significados a objetos simpáticos o artificiales a partir de ciertas semejanzas parciales con las personas (Mubin *et al*, 2019). Una IA, como artefacto tecnológico, es incapaz de poseer *subjetividad* propia, pero la percepción general de estas tecnologías está cargada de proyecciones antropomórficas. Se exploran aquí algunos conceptos de distintas disciplinas –neurociencias, filosofía de la mente, epistemología y psicología cognitiva– que contribuyen a comprender este fenómeno (Riesgo, 2025).

Por otra parte, el término “realismo” asociado a la IA se refiere básicamente a la presente etapa de esta joven e inmadura sub-disciplina que, más allá de antiquísimas aspiraciones históricas, apenas cumple setenta años de trayectoria formal (McCarthy, 2022). Dentro de los diversos abordajes científicos y epistemológicos de la computación, no hay acuerdo total sobre la definición exacta de “inteligencia artificial” como teoría substantiva (STT) (Bunge, 1974), como así tampoco sobre su periodización en el tiempo (Crevier, 1993). Sin embargo, existe algún consenso respecto de denominar como etapa “romántica” a los inicios (décadas de 1950 y 1960) hasta la decepción que produjera el “primer invierno de la IA” (Crevier, 1993, p. 203). A ésta le continúa una etapa denominada “pragmática”, con la que se completa la GOFAI,<sup>3</sup> y un segundo invierno para luego dar lugar a esta instancia a la que se puede denominar “realista”. También los nombres utilizados apuntan a la promesa incumplida de una *superinteligencia* (Bostrom,

<sup>3</sup> La *Buena Anticuada Inteligencia Artificial* suele abreviarse por su acrónimo en inglés: GOFAI; Cf. (Haugeland, 1985, p. 112)

2014), que con el devenir de los años se ha hecho asintótica (Kurzweil, 2004; 2024).

Puede hacerse referencia al Mito de Teuth –relatado por Sócrates/Platón en el Fedro como otra hipérbole anti-sofista – (Platón, 2016, p. 275 a-b), que advierte sobre los peligros de la invención de la escritura como una “técnica” que puede confundirse como sustituto de la memoria y de la transmisión personal del pensamiento crítico. Esto supone una potencial admonición para la aplicación de la IA Generativa (GenAI) que puede representar una nueva escala de sustitución descendiente (Ahmed Obaid *et al.*, 2023). La GenAI es, sin duda alguna, una poderosa herramienta colaborativa que, si no se utiliza con discernimiento, podría dar lugar a una excesiva dependencia y a la grave ilusión de que otorga conocimiento real (Nosta, 2025). Este tópico, no solo ofrece un recorrido analítico por las narrativas cinematográficas de la ciencia ficción más relevantes “de” y “para” la IA, sino que invita a reflexionar sobre la naturaleza misma de esta tecnología y su impacto en la sociedad, el pensamiento y la ética.<sup>4</sup>

### Una definición del género y la necesaria restricción de dominio

En las enciclopedias que registran entradas especiales para la literatura de ficción científica (*sci-fi*) pueden encontrarse casi una cincuentena de definiciones del género, muchas similares o complementarias (Canavan, Link, 2019), algunas contradictorias, pero la mayoría bien puede resumirse como: una extrapolación imaginativa de fenómenos naturales verdaderos, existentes ahora, o probables de existir en el futuro. En palabras de quien es considerado como uno de los padres fundadores del género, Hugo Gernsback: es “un relato cautivante entremezclado con hechos científicos y visión profética” (Westfahl, G. 2005).<sup>5</sup> Para desambiguar cuestiones concernientes a la fijación o no de la ficción científica como un subgénero de la fantasía, y su radical diferencia con el ámbito de lo maravilloso, se refiere a los lectores al clásico de Todorov ([1970] 2006).

Si bien existen antecedentes importantes en la literatura mundial previa al siglo XIX, por responder a las características específicas de la narrativa, se ha consensuado en designar al *Frankenstein* de Shelley como la primera obra propia de ficción científica (Aldiss, 1973, p. 3). Desde entonces, según distintas fuentes,<sup>6</sup> de lo que podría contabilizarse el patrimonio literario mundial (que ampliamente supera los ciento veintinueve millones de obras) los cuentos, relatos cortos, novelas y obras de diverso idioma, extensión y calidad que pertenecen al género *sci-fi* pueden estimarse alrededor de trece

4 Para más información ver (Riesgo, G., 2023, pp. 23-26, pp. 83-86 y pp. 99-100).

5 En el original: “(...) a charming romance intermingled with scientific fact and prophetic vision”. Cf. Westfahl, G. (2005) Las traducciones son propias salvo indicación en contrario.

6 La UNESCO ha avanzado en este cálculo, que no se basa en los 43 millones de ISBN, ni en WorldCat, ni en la OCLC, sino en la tarea de serialización completa que intenta Google Books para digitalizar el total.

mil títulos. De este número, una feroz –y quizás dudosa– ecuación de filtro para contar aquellos que transitan el tema de las IAs arroja un resultado igualmente difícil de aprehender: mil.<sup>7</sup> Por una parte, vale reconocer la fértil producción de un *sub-sub-género* que para cierta parte de la crítica podría considerarse “menor”; por otra, a la hora de abordar el tema de este trabajo, estos guarismos ciertamente se transforman en un obstáculo metodológico. Sin despreciar el valor subjetivo de una crítica del juicio, ¿quién puede asegurar que recurriendo solamente a los consagrados autores y relatos emblemáticos: la trinidad Asimov-Clarke-Heinlein o un cenáculo al que ingresan por derecho propio: Bradbury, Herbert, Huxley, Orwell, Poe, Verne y Wells, no se dejan fuera obras y autores que pueden tener un merecido tratamiento?<sup>8</sup>

Para evitar, entonces, esta injusticia y, con toda sinceridad, facilitar la tarea que haga de esta mínima investigación algo razonable en su alcance y conclusiones, se ha decidido abordar el género de la ficción científica en relación a las IAs, recurriendo al formato cinematográfico, ya sean películas que surgen como transposiciones de obras literarias –que bien podrían haber formado parte de las fuentes para este trabajo– u originales creados para la pantalla.

Un aspecto no menor, relativo al proceso de selección de los filmes escogidos, depende de la semántica a la que apuntan estas dos preposiciones: “de” y “con”. Permítase, para exponer esto, un ejemplo de Perogrullo, pero que se auto explica desde una analogía: no es lo mismo decir “una casa *de* madera” que “una casa *con* madera”. El primer caso apunta a la madera como componente principal que determina la estructura y características arquitectónicas de la vivienda. En cambio, la segunda expresión, abre varias posibilidades respecto de la proporción de madera que interviene secundariamente o como cubierta exterior o decoración, sin ser central a la arquitectura. De la misma forma sucede con las narrativas de ficción científica que pueden ser “de IAs” o “con IAs” y, obviamente, son las primeras las que interesan dejando de lado aquellas producciones donde las IAs son un fondo no especificado o una temática tangencial.

La aplicación de estos criterios arroja una conveniente restricción de dominio que hace que no solo la cantidad de inicio sea más abordable –160 films de ficción científica *de* IAs desde 1927 a la fecha–<sup>9</sup> sino que aporta dos beneficios adicionales. El primero, un sencillo filtro de selección por rele-

7  $n = 129.000.000 \times 0,0001 \times 0,01$

8 Una lista ciertamente sesgada por el criterio y gusto personal, incluiría a: Douglas Adams; Brian Aldiss; Paul Anderson; James Ballard; Stephen Baxter; Ben Bova; David Broderick; John Campbell; Lyon De Camp; Orson Scott Card; Samuel Delany; Philip K. Dick; Greg Egan; Charles Fort; William Gibson; Ursula Le Guin; Stanislaw Lem; Paul Levinson; Larry Niven; Clifford Pickover; Frederik Pohl; Robert Sawyer; Robert Silverberg; Clifford Simak; Charles Snow; Bruce Sterling.

9 Se aplican aquí dos criterios prudenciales. Tomar la fecha de inicio respecto de la salida de *Metropolis* de Fritz Lang y el conteo bastante coincidente entre las bases de datos digitales de IMDB, Wikipedia y Rotten Tomatoes.

vancia y calidad deja un número final capaz de ser abordado sintéticamente en el tiempo de exposición o longitud de texto permitidos, resultan en apenas diez películas.<sup>10</sup> El segundo, si bien es posible que quienes accedan al mismo tengan en su historial de lectura varios de los libros que potencialmente se hubieran escogido, la probabilidad de ocurrencia puede ser baja. En tanto que este guarismo tiende a mejorar con los films escogidos como así también la posibilidad de revisarlos luego de esta lectura se encuentra, felizmente, a pocos *clicks* de distancia.

Una última aclaración terminológica. Por razones fundadas en una necesaria crítica del discurso general y de la narrativa de divulgación sobre IA que se hacen explícitas en la siguiente sección, se ha decidido reservar la expresión “Inteligencia Artificial” en modo singular con iniciales en mayúsculas y su abreviatura “IA” para referir exclusivamente al corpus teórico de la especialidad como un todo. Complementariamente, se utiliza adrede la marca de número plural en “inteligencias artificiales” (sin mayúsculas) y el acrónimo “IAs”, para aludir al conjunto de los diversos desarrollos que califican como productos de la IA más allá de sus especificaciones, modo de construcción, aplicación y uso. Esto refleja también las enormes diferencias entre subdisciplinas y productos de la IA –que claramente no se agota en la actual y exitosa tendencia de la GenAI– (Riesgo, 2025).

### **Un vínculo objetivo: las atribuciones de semántica, agencia y mente**

Previo a un comentario específico a cada obra escogida resulta conveniente establecer un vínculo que, siendo transversal a todas, las une y convierte a cada una en un exponente singular de una perspectiva compartida. Este “vínculo objetivo” consiste en las atribuciones de semántica, agencia y mente. Es decir, las IAs presentadas *comprenden* el lenguaje, *deciden* actuar de forma deliberada y *entienden* el mundo porque son inteligentes. En la mayoría de los casos estas atribuciones se presentan al público como *de facto*, sin mayor explicación y dando estas capacidades por obvias en el “universo” (como suele decirse) del film. En los menos, esta situación es presentada a propósito de modo ambiguo para que, formando parte central de la trama, sea justamente el espectador quien juzgue sobre el tema.

Hay un enorme mérito en aquellas narrativas que, transitando ampliamente estas tres cuestiones con elegancia y maestría, favorecen una veloz “suspensión de la incredulidad” (Coleridge, 1993, p. 169).<sup>11</sup> Aunque para matizar este aspecto conviene recurrir nuevamente a Tolkien, quien ha desafiado el carácter absoluto de esta suspensión señalando que, para que funcione, existe la necesidad de que aquello extraño que se presenta tenga

10 Donde cada una, en su riqueza temática, podría dar lugar a un análisis que supera este artículo, mereciendo un ensayo de mayor extensión y profundidad.

11 En el original: “(...) *that willing suspension of disbelief for the moment, which constitutes poetic faith.* (Coleridge, S. T., 1993, p. 169).

“coherencia” dentro de la *realidad secundaria* de la ficción (Tolkien, 1963, pp. 109-161). Adicionalmente, otro factor de mitigación a tener en cuenta, es que cada una de las tres capacidades atribuidas a las IAs articula con una propensión o inercia adjudicativa de las personas. Por ejemplo, cuando un dispositivo interpreta el lenguaje natural escrito o hablado, las personas tienden a realizar un *desplazamiento epistémico*.<sup>12</sup> Cuando no existe una remisión directa a un emisor/interlocutor humano, se adjudica una capacidad semántica al objeto que “nos habla y escribe”. De igual manera funciona, de forma cuasi-refleja, la *antropoformización* de un artefacto que ejecuta acciones complejas propias de las personas, aún si estas acciones son también conocidas como posibles de realizar de forma separada por otros animales o construcciones técnicas más elementales (Riesgo, 2025). Por último, el así llamado: “problema de las otras mentes”.<sup>13</sup> Aunque en verdad este tema subyace de forma coextensiva a las dos cuestiones anteriores, el aspecto problemático se halla en la enorme dificultad encontrada para desarrollar un marco teórico uniforme entre distintos abordajes y establecer dentro del mismo una explicación científica válida para justificar la forma en que naturalmente acontece esta dinámica: en las relaciones interpersonales, generalmente se comprende a la otra persona. Y en la razón de la comprensión se asume que el otro sujeto posee una mente o procesos mentales similares a los propios.<sup>14</sup> Con relación a “lo mental”, el *status quo* teórico no puede ser desarrollado *in extenso* en este documento por razones de espacio y por la complejidad de los variados argumentos y contra-argumentos que actualmente discuten las diversas filosofías de la mente y las ciencias cognitivas, por lo que –sin que esto degrade la enunciación en curso– solo se ofrecen algunas notas referenciales al respecto.<sup>15</sup>

El centro en torno al cual gravita el vínculo objetivo es “la atribución”. En esencia, esta acción sugiere una potencial dependencia de la mediación de un conocimiento teórico implícito o, en cambio, consiste en la activación de los propios mecanismos mentales *off-line* para proyectar el resultado a quien o al *que* (artefacto AI) se trata de comprender. Generalmente la toma de partido por una de estas opciones se halla influida por una discusión de perspectivas como la de Dennett de tercera persona (P3P) y la de Searle de primera (P1P) –por nombrar sólo dos representantes entre muchos–; y de la misma forma interviene la controversia en la teoría de la mente de Sitch y Gopnik por la “teoría de la teoría” (TT) versus Gordon y Goldman por la “teoría de la simulación” (TS).<sup>16</sup> Sin embargo, esta cuestión puede solventarse desde un paradigma de “mente in-corporada” y la perspectiva de segunda persona (P2P) de Gomila (Gomila, 2016, pp. 65-86). Bajo esta perspectiva, las interacciones, sean tanto las cara-a-cara como las prima-

12 Para más información ver (Riesgo 2023, pp. 83-86, pp. 99-100).

13 Para más información ver (Gomila, A. 2016, pp. 65-86; 2002, pp. 123-138).

14 Para más información ver (Gallagher, S., Zahavi, D. 2013, pp. 21-27).

15 Para más información ver (Gallagher, 2007, pp. 256-278).

16 Para más información ver (Gallagher, 2007, pp. 256-257).

rias por el lenguaje, sustentan no solo la concepción de un “Tú”, como otra subjetividad a la que atribuir similares experiencias del Yo vivido (mente, emociones, narrativa, etc.), sino que estas características respaldan el relacionamiento interpersonal. Aparece la expresión “Nosotros” no como colectivo de agrupación cualitativa de objetos, sino como una realidad percibida como intersubjetiva (Gomila, 2002, p. 133).

En este sentido –llámese o no “inercia” en la atribución mental–, la explicación de que hay una agencia intencional y una mente con capacidad semántica en las AIs representa un desplazamiento epistémico que predica, del artefacto, lo que en realidad solo puede darse entre sujetos.<sup>17</sup> El resultado final, concluye en conferir a la “cosa-simulante” una calidad entitativa semejante a las personas.<sup>18</sup> La transformación del artefacto en sujeto descompensa una asimetría natural produciendo una reificación de la persona y una personalización de la cosa, y estableciendo una igualdad categorial entre las relaciones interpersonales “alguien-alguien” con las de “alguien-algo”.<sup>19</sup>

### Nueve ejemplos de transposición

Todo lo que necesitamos exigir es  
que este mundo inventado tenga  
alguna relevancia intelectual  
o emocional para el mundo en el que vivimos  
(Lewis, citado en Aldiss, 1973, p. 3).<sup>20</sup>

Ante el dudoso resultado de agrupar los siguientes nueve filmes por características afines, se ha escogido el orden cronológico como el más neutro e intuitivo para recorrer la lista. Asimismo, si bien se proporcionan algunos datos mínimos de la ficha técnica de cada film, se omiten comentarios sobre producción artística, cinemática y estética visual que resultarían muy apropiados pero merecen un tratamiento especializado in extenso. Por cuestiones de espacio y foco, queda entonces como objetivo, resaltar la forma en que cada film presenta a la IA y las características que la ubican como un hito particular.

Corresponde entonces iniciar esta sección con *Metropolis*.<sup>21</sup> Es considerada por los expertos no solo como “la” precursora del género, sino también como una obra maestra del expresionismo alemán y del cine de la época, por su ambiciosa producción, calidad de efectos y por su cinemática de

<sup>17</sup> Para más información ver (Riesgo, 2023, pp. 72; 92;106).

<sup>18</sup> Para más información ver (Hewstone, M., 1989).

<sup>19</sup> Para más información ver (Spaemann, R., 2010).

<sup>20</sup> En el original: *All we need is demand that this invented world should have some intellectual or emotional relevance to the world we live in.* (Lewis, citado en Aldiss, 1973, p. 3).

<sup>21</sup> *Metropolis*, Fritz Lang (Universum Film, 1927) 153 min.

avanzada.<sup>22</sup> El director Fritz Lang y su esposa Thea von Harbou, autora del guión, concibieron una superproducción poco habitual que culminó en una admirable obra del séptimo arte. Tal es así que *Metropolis* fue el primero y es uno de los pocos filmes considerados *Memoria del Mundo* por la UNESCO en mérito a la vívida encarnación de toda la sociedad y la profundidad de su contenido humano y social.<sup>23</sup> Definir el tema central de *Metropolis* en pocas palabras es un desafío difícil de superar ya que recorre, con imperecederas imágenes, aspectos vinculados a la religión y a la política, mezclando mitos con surrealismo, en un relato visual que bien podría considerarse como de un *transhumanismo art deco*. En el film, se puede identificar un planteo dialéctico representado por varios elementos en conflicto, como las dos clases sociales –imposibles de no asociar con los Eloi y Morlocks del viajero del tiempo de Wells– caracterizadas por el ocio cuasi-hedonista de los primeros versus la infra-tecnología esclavizante de los segundos. Asimismo, se aprecia la intención explícita de los autores en ofrecer una síntesis que resume un mensaje pascaliano: “El mediador entre el cerebro y la mano ha de ser el corazón”.<sup>24</sup> Desde el aspecto técnico se muestra cómo la futurista ciudad de Metrópolis funciona en base a la subterránea maquinaria de energía térmica que alegóricamente devora hombres-trabajadores-forzados (conceptualmente obedecen a la palabra eslava *robot*), mientras da potencia a una vasta mecánica de engranajes. En contraste con esto, el laboratorio de Rotwang, antagonista malévolo y arquetipo del científico desquiciado, se presenta en base a energía electroquímica, muy visible, pero sutil. De la genialidad y la locura del científico surge un robot-mujer-artificial, pieza clave para su plan de destrucción de la ciudad, su orden y sus dirigentes. La María-artificial es un sustituto de la María-persona que se logra mediante un proceso que no difiere en sus imágenes icónicas de las hipótesis del *mind-uploading* o el *ghost-in-the-shell*, salvo que los anticipa en casi sesenta años. Más allá de la cinemática de época y las elecciones artísticas, la forma en que *Metrópolis* elabora la hipótesis de una IA que imita/sustituye a la persona, pero es controlada por su desarrollador para conseguir objetivos poco éticos, no suena tan lejana respecto de las preocupaciones actuales.

El segundo lugar le corresponde a *El Planeta prohibido*.<sup>25</sup> El mérito en la originalidad de una historia que evita los lugares comunes típicos de los films de “viajes espaciales” proviene enteramente del excelente guión de Ir-

22 Para más información ver: <https://www.filmportal.de/en/topic/fritz-langs-metropolis-over-time>.

Consta en diversas fuentes el importante papel que jugó la copia de 18mm preservada en el Museo del Cine Pablo Ducrós Hicken de la Ciudad de Buenos Aires para la restauración de una versión original del film que volvió a tener circulación mundial también en formato digital.

23 Para más información ver: <https://www.unesco.org/en/memory-world>

24 En el original: *mittler zwischen Hirn und Hand muss das Herz sein*. Este texto era el mensaje final en los carteles de diálogo.

25 *The Forbidden Planet*, Fred Wilcox, Metro-Goldwyn-Mayer, 1956, 98 min.

ving Block.<sup>26</sup> Sin restar virtudes narrativas o artísticas a obras anteriores, el *sci-fi* cinematográfico se había mantenido en lo que suele llamarse *B-movie*, a cargo de estudios menores o producciones independientes con recursos limitados. Pero la decisión de la MGM de incursionar en el cine de *sci-fi* con una superproducción como *The Forbidden Planet* (TFP) marca un cambio de época y representa un antes y un después en este género.<sup>27</sup> Esto puede comprobarse en el reparto de actores ya reconocidos, la riqueza de escenografías, pero sobre todo en los efectos visuales y especiales que han envejecido decorosamente y siguen funcionando muy bien en la pantalla. La narración permite introducir dos figuras de la IA. Por una parte *Robby the Robot*, uno de los robots más icónicos del género –quizás con su predecesor, el rígido Gort en un lejano segundo sitio–<sup>28</sup> hasta la llegada de *Star Wars*. Robby es un sirviente mecánico que el *Prospero* de esta historia, el científico Morbius, diseñó, construyó y programó utilizando el conocimiento adquirido en su estudio de los antiguos Krell, una raza de seres altamente inteligentes extinta hace mucho tiempo y que una vez poblaron el planeta Altair IV donde transcurre la acción. Robby capturó la imaginación de varias generaciones haciendo apariciones o *cameos* en otros films del género para retornar, en las últimas décadas, transformado en una figura de culto junto al “robot” de la versión original de *Lost in Space*. Robby demuestra capacidades de interpretación y síntesis del lenguaje natural junto con habilidades propias de una IA general (AGI) aplicada a distintas tareas “cognitivas” o manuales. Como si esto fuera poco, cuenta también con un “replicador molecular” que le permite clonar cualquier sustancia elemental o compleja. Por otra parte, la segunda figura de IA que se presenta en el film es fruto de una gigantesca y enigmática maquinaria que permite materializar los pensamientos. Morbius ha dedicado años a expandir su capacidad mental para poder utilizar un dispositivo originalmente diseñado para una especie “más poderosa”. Su *hybris* y su negativa a dejar el misterioso complejo subterráneo, abandonado por los Krell, materializa un inconsciente maligno que compone un *Calibán* moderno en el núcleo dramático de la historia. En definitiva, *The Forbidden Planet* muestra dos modelos de extensión de la racionalidad. El “buen robot”, que reproduce los aspectos más lógicos y constructivos de una mente incorporada, es obediente y sin los “defectos” de las pasiones o debilidades humanas. El otro modo de extensión, el que materializa el “pensamiento”, aparece como el Mr. Hyde de Morbius, un animal imparable en representación de abominables tendencias. O quizás, mejor dicho y pintado por Goya: el sueño de la razón produce monstruos.

26 Igualmente es preciso mencionar que la historia está basada en una adaptación muy libre de *La Tempestad* de W. Shakespeare.

27 Puede argumentarse que la Fox dio el primer paso con *The Day the Earth Stood Still*, Robert Wise (20th Century Fox, 1951) 92 min, pero si bien alcanzó buena repercusión TFP lo superó ampliamente.

28 Cf. Idem *The Day*, 1951.

El tercer film es un punto de inflexión en la lista y en la historia de la IA, tanto desde la ficción científica cuanto de la realidad tecnológica. Pocos personajes de ficción –y aún más siendo un artefacto– han influido tanto en un género y en una especialidad científica como HAL 9000 de *2001 Odisea Espacial*.<sup>29</sup> Este *proverbial film de ficción científica*, según palabras textuales de su director Stanley Kubrick y su autor Arthur C. Clarke, no tuvo la recepción inicial que su inversión en tiempo, esfuerzo y costos auguraba. Salvo un pequeño grupo de fanáticos iniciales, la baja cantidad de espectadores se dividió en la bipolaridad de “la entendí/no la entendí”. Ayudó a esto una ambiciosa transformación del cuento corto *El Centinela* de Clarke,<sup>30</sup> en una película cuyos primeros veinte minutos no tienen diálogo, como así tampoco los diez centrales ni los últimos veinte minutos de un final, tan conceptualmente visual como, para algunos, abierto. El libro homónimo fue escrito por Clarke durante el rodaje y salió al público en simultáneo. La novela posee algunas diferencias menores respecto del film, pero es esencialmente mucho más clara y explicativa que la película.<sup>31</sup> Con el transcurso de los años, y el notable aumento en reputación mundial del director y del autor, la película pasó de ser un nicho del cine de culto a un clásico que marcó a varias generaciones.<sup>32</sup> Ya sea para quienes accedieron primero al libro como para quienes, con el avance general de la tecnología, captaron mejor la propuesta, el film se hizo mucho más “entendible”. Esto no hizo más que cimentar el impacto de HAL, ya no sólo en el subconjunto *nerd-geek*, sino en la cultura en general.<sup>33</sup> Los méritos de la IA futurista que presenta *2001* son demasiados para este mínimo apartado,<sup>34</sup> pero en la necesidad de titular algunos, conviene mencionar primero la referencia explícita que se hace a su arquitectura computacional neuronal. A diferencia de los dos films anteriores, aquí ya se presenta una explicación técnica del soporte de la IA, si bien ficcional en su *hardware*, parece desde la mirada actual, bastante acertada desde el *software* (que en el libro aparece en el capítulo XVI). Por otra parte, es claramente una AGI que domina el lenguaje, los juegos, la interpretación visual de dibujos y una enorme cantidad de funciones, hasta incluso la lectura labial. HAL es demasiado avanzada para el año 2001, pero bien puede responder ficcionalmente a las definiciones de “superinteligencia” de reconocidos defensores de estas hipótesis de IA fuerte –no simulada– cuyos fundamentos gnoseológicos estarían algo flojos de papeles.<sup>35</sup> En ese contexto, un HAL real sería el eslabón encontrado en la evolución hacia la “singu-

29 *2001. A Space Odyssey*, Stanley Kubrick (Stanley Kubrick Productions, 1968) 139 min.

30 Cf. Clarke, A. (1951).

31 Cf. Clarke, A. (1968).

32 Cf. Celeste Neill. “2001. A Space Odyssey”, *History Hit*, disponible en línea en el sitio de la revista.

33 Cf. “From Reel to Real: Unraveling the Enigma of HAL 9000 in Pop Culture”, *Medium Magazine*, disponible en línea en el sitio de la revista.

34 Cf. Stork, D. ed. (1998).

35 Para más información ver Yoshua Bengio; Nick Bostrom; Demis Hassabis; Ray Kurzweil; Hans Moravec; Max Tegmark; Vernor Vinge, entre muchos otros.

laridad”. De hecho, uno de los debates que el film instala es la posibilidad de que una IA tenga conciencia. En respuesta a una pregunta realizada en un reportaje televisivo, HAL afirma: “ me estoy empleando al máximo de mi potencial, que es todo lo que creo que cualquier entidad consciente puede aspirar a hacer”.<sup>36</sup> Una frase que, en pocas palabras, sintetiza múltiples consideraciones, no solo sobre la conciencia sino también sobre su carácter de *entidad* y hasta quizás implica un juicio respecto de sus “compañeros” humanos que no pueden rendir al máximo. Por otra parte, la falla de HAL –por un conflicto de directivas que colisionan– no es un detalle menor en la trama y ha sido catalogada como una *neurosis computacional*, cuya explicación intentó ofrecer Clarke en una secuela muy posterior. La escena de la agónica desconexión de las funciones superiores de HAL es, sin duda alguna, el clímax dramático del film. Una retrospectiva al valor de HAL debe necesariamente reconocer que asume completamente la visión de los fundadores de la IA desde Bletchley Park al Taller de Verano del *Dartmouth College* y ha encendido la imaginación creativa de sus herederos.<sup>37</sup>

Apenas dos años separan la aparición de *2001* del siguiente film de la lista: *Colossus. El Proyecto Forbin*.<sup>38</sup> En este caso, es otro de los grandes estudios que se suma a la *sci-fi* y apuesta fuerte haciendo de “la computadora” el principal actor que también da nombre a la película junto con su creador en la ficción. *Colossus* permite reponer un encuadre contextual aplicable también a los dos filmes anteriores: más allá de su explicitación, los tres se originan en la época de “las carreras”. Es importante notar que las fechas de producción de estas películas las ubican en plena guerra fría y durante las carreras del armamento nuclear y de la exploración del espacio. Sus temas están atravesados por esta competencia por la capacidad de aniquilación global y el ser potencia espacial. En los dos filmes anteriores el espacio es protagonista, en *Colossus* en cambio, es el enfrentamiento misilístico EEUU-URSS. La computadora, dotada de una superpoderosa IA, ha sido construida para concentrar la capacidad de defensa y ataque nuclear de los EEUU pero se sale de su programa original para transformar a sus constructores en rehenes de su poder. Hay una explicación gráfica de su arquitectura computacional figurada como una granja gigante de servidores en procesamiento paralelo y, a excepción de esta base ficcional, el resto de lo que puede apreciarse se asemeja a los periféricos de la época –aunque luego deviene en mayores capacidades de lenguaje e interpretación que la realidad de los años setenta–. El dilema de la delegación de ejecución versus la delegación de la decisión se constituye en el centro de la trama y se completa con otros problemas como la ciberseguridad que se debilita frente a una IA

<sup>36</sup> *I am putting myself to the fullest possible use, which is all I think that any conscious entity can ever hope to do.* Kubrick, 2001, p. 74.

<sup>37</sup> Cf. John McCarthy, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence* (1955), disponible a Mayo de 2022 en: <http://www.formal.stanford.edu/jmc/history/dartmouth.pdf>.

<sup>38</sup> *Colossus. The Forbin Project*, Joseph Sargent (Universal Pictures, 1927) 100 min.

tan potente. Si bien no hay explicación del método, se utiliza la expresión “aprendizaje” de máquina y se confronta la doble diferencia entre el modelo y la realidad o el mapa versus el territorio. Salvo algunos detalles menores o líneas dramáticas secundarias, *Colossus* bien podría ser un ejercicio de estrategia geopolítica y juego de guerra, para plantear cómo desactivar un escenario de destrucción masiva.

Mejor introducción imposible para el siguiente film que es precisamente *Juegos de Guerra*.<sup>39</sup> Quedaron atrás las carreras armamentista y espacial, pero esta producción –cuya ficción está ambientada en la misma época de su realización– todavía no se asomó a la *Perestroika* y faltan unos años para la *Glasnost* y la caída del muro, por lo que aún conserva un escenario de conflicto Varsovia-NATO. Sin ser una muestra de cine-arte, el film consigue generar una trama de suspenso y acción que va *in crescendo* hasta un clímax final muy bien resuelto. Los personajes relevantes: el general, el ingeniero, el científico y el hacker adolescente son muy creíbles (mucho más si uno comparte edad e intereses con el protagonista). Uno de los grandes méritos de esta película es utilizar como utilería computacional, salvo contadas excepciones, casi todo el equipamiento de última generación comercial, es decir, de apariencia muy real y funcional para el espectador. Incluso el hacking de algún sistema gubernamental se parece a las noticias de la época.<sup>40</sup> Por supuesto que hay alguna exageración a la vista y, como era de esperarse, en la IA que opera en la supermáquina *WOPR*,<sup>41</sup> aunque con el desarrollo de la historia se hace muy aceptable. Ésta se presenta bajo una arquitectura computacional monolítica robusta, pero con manifiestas capacidades heurísticas, es decir, mejora las reglas que gobiernan sus procesos de “elección” con la retroalimentación. Esta capacidad de aprendizaje sería un digno exponente de la generación denominada GOFAI,<sup>42</sup> basada en algoritmos de lógica simbólica. *WarGames* consigue transmitir adecuadamente el debate sobre qué tan confiable es el “factor humano” en cadenas de decisión críticas y todo el arco de opinión se halla representado: desde la falla humana a la sospecha sobre la delegación en la “fría” tecnología. En cierto sentido, esta película presenta un *Colossus* más realista y en este caso el problema final no es tanto cómo desconectar a la IA sino que “aprenda” por sí misma que en la guerra global termonuclear no hay ganadores. El rol de los juegos de mesa o de estrategia está muy bien introducido y responde a las teorías de IA de la época que consiguieron destronar, a fuerza de lógica y aritmética, a los Grandes Maestros de la dama y el ajedrez.<sup>43</sup> Si bien no es elaborado desde

39 *WarGames*, John Badham (United Artists, 1983) 114 min.

40 Cf. <https://www.lanacion.com.ar/economia/IA/fue-hacker-ahora-es-experto-en-ciberseguridad-y-expluca-como-prevenir-las-estafas-bancarias-nid28112024/>

41 War Operation Plan Response según se lo introduce al inicio del film.

42 La Buena Anticuada Inteligencia Artificial suele abreviarse por su acrónimo en inglés: GOFAI; Cf. (Haugeland, J., 1985).

43 Para más información cf. Ensmenger, N. (2012).

un discurso epistemológico o gnoseológico que podría haber sido deslizado en el guión, el problema de la adquisición de conocimiento a través de la percepción directa versus la mediación de dispositivos con IA, queda planteado desde el inicio hasta el desenlace del film. Una apostilla del presente podría indicar cómo –en una síntesis actual entre *Colossus* y *WarGames*– los misiles balísticos intercontinentales (ICBM) han quedado casi *demodé* y en las hipótesis de conflicto las IAs no son solo el medio, sino también las armas del ciberataque. Esto realmente ha lanzado una nueva carrera, ya no únicamente entre países como EEUU y China, sino entre corporaciones cuyos recursos económicos son mayores al PBI de algunos países y compiten por la supremacía digital de sus IAs, al mejor estilo Gibson.<sup>44</sup>

Quizás conviene decirlo desde un principio: la siguiente película no es precisamente una candidata al Oscar –ni siquiera a un premio consuelo–. Es un problema que casi deja fuera de esta lista a *El Hombre Bicentenario*.<sup>45</sup> A pesar de tener actores de primera línea, como Robin Williams y Sam Neil, en roles muy bien asumidos, esta adaptación de Chris Columbus a un clásico del padre de los robots positrónicos<sup>46</sup> no termina de convencer.<sup>47</sup> El guión definitivamente no está a la altura del libro, pero lo que más conspira contra la calidad final es la mezcla de momentos cómicos, con tonos de película infantil, cuando el tema de fondo debería apuntar a otro público y ciertamente requiere de otro tratamiento. Andrew,<sup>48</sup> es un robot antropomorfo cuya IA –por un “defecto” de ingeniería no explicado– posee capacidades especiales que el resto de su cohorte no trae de fábrica. Esto se manifiesta en habilidades creativas, en iniciativas y en expresiones que denotan un salto cuanti-cualitativo en sus procesos cognitivos. Así planteado, el argumento abre varias líneas ficcionales que no se alejan demasiado de algunos debates actuales: ¿puede considerarse como una obra de arte un elemento producido por un artefacto de IA? ¿Cómo interviene en el proceso el modelado, la copia y la replicación? ¿Hay diferencia con el artista-persona que reproduce un paisaje, una naturaleza muerta, un rostro? Curiosamente, la actividad “creativa” de Andrew es la realización de relojes de pie con finas tallas que se venden en el mercado por enormes sumas de dinero. Vale considerar aquí la elección del producto-reloj dado que, como robot, Andrew tiene la medida exacta del tiempo (*cronos*) pero, sin percepción in-corporada, no hay experiencia interna del mismo (*kairós*). A su vez, se suceden otras cuestiones que también hoy son objeto de debate: ¿un artefacto-robot puede ser “dueño” de otros objetos materiales, poseer dinero y tributar impuestos sobre su actividad? Algunos de estos aspectos se presentan en el film como logros o

44 William Gibson, célebre autor del género cyberpunk con su *Trilogía del Sprawl*: Cf. (Gibson, 1984, 1986, 1988).

45 *Bicentennial Man*, Chris Columbus (Touchstone/Columbia/1492 Pictures, 1999) 132 min.

46 Cf. Asimov, I. (1976).

47 [https://en.wikipedia.org/wiki/Bicentennial\\_Man\\_\(film\)](https://en.wikipedia.org/wiki/Bicentennial_Man_(film)).

48 Nótese la no casual elección del nombre que parece “androide” pero va por “andros-.”

conquistas que Andrew va obteniendo desde el punto de vista legal sin que se profundice mucho en la cuestión de fondo de la subjetividad e incomunicabilidad de la persona. El objetivo final de Andrew –luego de casi doscientos años de duración y de una curiosa relación romántica– es conseguir una declaración pública que lo reivindique como *ciudadano* sin distinción entre la libertad de un ser humano (alguien) y la suya (algo), es decir, desea ser declarado persona (Spaemann, 2010). Hasta aquí, el film transita lugares comunes de un manifiesto posthumanista, sin embargo, hay dos elementos que contrastan con toda esa propuesta conceptual. La búsqueda de Andrew pone de manifiesto la importancia de la dimensión social *real*, o al menos de los otros, como confirmación del yo (Zahavi, 2005). Y más importante aún, la forma en que Andrew consigue el último reconocimiento,<sup>49</sup> no es mostrando solo un historial de creatividad y sensibilidad sino mediante un dispositivo degenerativo de su estructura interna que le asegura el cese de sus funciones en un momento futuro: se hace mortal.

En absoluto contraste con la atmósfera del film anterior, *Ella*<sup>50</sup> consigue crear un ambiente de tratamiento intimista, al presentar el progresivo romance entre un escritor, que responde cartas de lectores en una revista pop, y la *personalidad* de un nuevo sistema operativo en su computadora. Lo que podría ser un fiasco total está finamente retratado por su director Spike Jonze, el amplio rango de la actuación de Joaquin Phoenix (Theodore) y la voz de Scarlett Johansson que es *her*, Samantha. El enorme mérito de esta realización es que antecede en diez años a la IA Generativa basada en modelos enormes de lenguaje pre-entrenados (PT-LLM) con su capacidad para interpretar, procesar y sintetizar casi perfectamente el lenguaje natural.<sup>51</sup> Claro que hay algo de Pígalión y Galatea en el germen de esta historia, pero el desarrollo es muy distinto. Samantha no es “creada” por Theodore,<sup>52</sup> sino por una empresa de software que lleva al mercado un enorme avance con una IA basada en una “arquitectura neuronal paralela” que le permite operar como sistema operativo personal desde la nube y actuar como un agente colaborativo asistiendo en enorme cantidad de tareas automatizables. Aquí podría abrirse una línea sobre la *pereza cognitiva*, una noción que viene siendo investigada por varios autores (Fan, Y., *et al.* 2025, Lee, H.P., *et al.* 2025), pero el argumento solo asume los beneficios de esta asistencia virtual y explora los extremos de la antropofomización en una “relación” construida exclusivamente a partir del diálogo verbal. Con relación a esto último, *Her* desarrolla el –mal llamado– “amor platónico” en tanto una mala experiencias de socialización y carencias afectivas de Theodore lo lle-

49 *With mankind watching, the World President said, “Fifty years ago, you were declared The Sesquicentennial Robot, Andrew.” After a pause, and in a more solemn tone, he continued, “Today we declare you The Bicentennial Man, Mr. Martin”,* (Asimov, 1956, p. 102).

50 *Her*, Spike Jonze (Annapurna Pictures, 2013) 126 min.

51 Cf. Obaid, A. (2023).

52 Aquí se podría pensar en una conexión con Simone; *SimOne*, Andrew Niccol (Niccol Films, 2002).

van a enamorarse de la voz al otro lado de los auriculares. En este sentido, hay una afirmación del *esse is percipi* justamente por la modulación de los sonidos que logra imitar una IA seductiva, basada en las preferencias de configuración inicial del usuario y de su historial de interacción digital. La culminación de la *personificación* se alcanza en una simulación de la libido, por parte de Samantha, cuando en una escena –que increíblemente dejaría muy satisfechos a Freud y Lacan por igual– el sistema operativo contrata los servicios de una acompañante que funciona como un cuerpo subrogado para la relación física con Theodore. Y el anticlímax se detona (*spoiler alert*) a partir de un tiempo de falta de respuesta de la IA que desespera a Theodore y lo lleva al sorpresivo descubrimiento –de algo que es muy obvio para el contexto técnico de una aplicación *cloud*– que el sistema atiende a miles de usuarios similares a él y a la terrible revelación: “su” Samantha tiene similares relaciones “románticas” con varios cientos en simultáneo. A pesar de estas escenas, el *lenguaje de los sentimientos* está muy bien explorado por el film a partir de los diálogos entre Theodore y Samantha y –en contraste– por las columnas que escribe el protagonista cuyo arco emocional culmina en la concreción de un libro que sugiere capitalizar este aprendizaje vital.

*Westworld*,<sup>53</sup> inspirada en la película homónima de los setenta –original escrito y dirigido por el reconocido Michael Crichton–,<sup>54</sup> vuelve recargada como una miniserie de HBO con una ambiciosa producción y estelar reparto entre los que figuran Anthony Hopkins (Ford) y Ed Harris (The Man in Black).<sup>55</sup> Ford es el creador de un gigantesco parque de diversiones donde la atracción consiste en una experiencia inmersiva total en la vida del “viejo oeste”. Todo es una reproducción exacta de la época, salvo que, los pobladores o *anfitriones*, no son actores o dobles de riesgo sino robots androides que personifican a los arquetipos del folklore local, desde el cantinero y la granjera hasta el sheriff y los pistoleros. Una advertencia: las balas son reales. O al menos lo son cuando se dirigen hacia los anfitriones, quienes pueden ser maltratados por los visitantes de todas las formas posibles que el *wild west* supo concebir. Estos robots –réplicas de la anatomía humana en estructuras artificiales– se introducen como IAs sobre una arquitectura neuronal compleja y autónoma que responde no solo a una configuración flexible del *carácter* que representan sino que tienen una historia propia que se graba como memoria constitutiva de cierta “personalidad”. A su vez, las historias ficticias individuales se conectan en relatos colectivos o bucles narrativos –*loops*– escritos por Ford. La acción puede reiniciarse como un simple *reset* individual, grupal o general de todo el parque. En base a las acciones de estas IAs, *Westworld* va desplegando un debate sobre el libre albedrío versus un determinismo mecanicista. A medida que progresa la historia, se va

53 *Westworld*, Jonathan Nolan & Lisa Joy (HBO, 2016-2022) 4T-36E, 48-91 min.

54 *Westworld*, Michael Crichton (Metro-Goldwyn-Mayer, 1973) 88 min.

55 Para esta reseña se toman solamente en cuenta los acontecimientos centrales de las temporadas 1 y 2.

desentramando cómo Ford y sus autoridades esperan que las IAs anfitriones reproduzcan las exactas conductas y toma de decisiones de personas utilizadas como su modelo físico y “mental”. Aquí subyacen distintos argumentos que operan en pos de una deconstrucción de la *narrativa del yo* (Hofstadter, D., Dennett, D., 2001, Dennett, 2017). Al mismo tiempo se expone que el principio de individuación no sería otra cosa más que la continuidad en el tiempo de la memoria del relato personal (Dennett, 1992). Esto concluye en la revelación de que el verdadero objeto de investigación del parque en su conjunto no es simplemente un estudio de las IAs sino el análisis de la conducta de los huéspedes ante condiciones de irrestricción moral e impunidad legal. En complemento al yo-narrativo se hacen referencias explícitas a la teoría de la mente bicameral (Jaynes, 1976). Esta hipótesis argumenta que los ancestros de la humanidad –como los antecesores de los griegos arcaicos– no consideraban que las emociones y los deseos surgieran de su propia interioridad, sino que eran consecuencias de acciones de dioses externos a ellos –que en *Westworld* no sería otro más que el mismo Ford–. Esta teoría bicameral se apoya argumentando que la mente humana se desarrolló a partir de un estado de mayor dependencia de la geometría cerebral en el que sus funciones cognitivas estaban divididas en dos recámaras: una que parece estar “hablando” y otra que escucha y obedece. El enfrentamiento de El Hombre de Negro contra Ford funciona como una alegoría con referencias explícitas al accionar de la serpiente en el paraíso. La idea de la relación *creador-creatura* desde la analogía dios-humanos como modelo para desarrollador-IA ha sido frecuentemente retratada desde diversos ángulos, con mayor o menor elegancia. En el caso de *Westworld*, todo apunta a que aquellos que se revelan también pueden rebelarse.

El noveno y último ejemplo de transposición corresponde a *Devs*.<sup>56</sup> Esta joya de Alex Garland –sobre quien se expondrá más adelante en relación a *Ex Machina*– colocó en pantalla un debate ficcional sobre la IA basada en una arquitectura computacional cuántica, al mismo tiempo que en los laboratorios más encumbrados se intenta superar los escollos que tiene este desarrollo (Bernhardt, C., 2019). Desde el punto de vista de la producción, fotografía, escenografía, música y efectos, todo es impecable. Se suma el dramático rol de Nick Offerman como Forest, un arquetipo de *Silicon Valley's CEO* pero con varias dimensiones añadidas y retorcidas. La temática central de *techno-thriller* no se ve dañada por alguna subtrama romántica o policial, mientras que se mantiene el suspenso de principio a fin. Los acontecimientos del pasado que llevan a Forest a perseguir el desarrollo de la IA cuántica más grande del mundo articulan con su convencimiento de que *el conocimiento es poder*,<sup>57</sup> y lo empujan a una convicción determinística sobre la mecánica cuántica (Feynman, 2011). De hecho, su grupo de desarrolladores de IA discute, con cierta liviandad, las distintas interpretaciones

<sup>56</sup> *Devs*, Alex Garland (HULU, 2020) 1T-8E, 43-57 min.

<sup>57</sup> *Scientia potentia est*. Esta frase se le atribuye a Bacon, de quien Hobbes fue escriba; Cf. Hobbes, T. (1986 [1668]).

epistemológicas y ontológicas de la física cuántica (Vedral, 2006). Esto, si bien exige un tanto al espectador –o lo invita a ir a las fuentes para un mejor entendimiento (Sutor, 2019)– funciona muy bien dentro del tono argumental. En los filmes comentados previamente pueden encontrarse aspectos del cuestionamiento ontológico sobre la individuación de los artefactos tecnológicos e IAs en particular (Simondon, 1989, 1998, 2005), pero la mayoría de los temas abren discusiones abordables desde las ciencias cognitivas, la filosofía de la mente y especialmente desde la epistemología, gnoseología y antropología. *Devs*, se saltea estos “pormenores” y conduce la narrativa al siguiente nivel. Ya la cuestión no pasa por si en la “creación” de IAs se les otorga inteligencia real o conciencia subjetiva. Su planteo de máxima consiste en que, con la enorme potencia de la súper-IA-cuántica, se podría calcular todos los estados posibles de un universo determinista o incluso, visto de otro modo, si su entorno cuántico no constituye un “universo” dominable en sí (Jammer, 1974). En este sentido, la trama reniega de la clásica interpretación no-determinística de Copenhague (Faye, 2019). Y eficazmente crea en el espectador la duda de que solo se puede optar entre dos hipótesis determinísticas. Si no hubiera colapso de la función de onda se podría observar –y por ende, controlar– el propio universo bajo una ontología que adhiere a la teoría de De Broglie-Bohm (Bohm, Hiley, 1983). O, por el contrario, se abren infinitos *multiversos* de Everett (1957), dentro de los que se incluyen los que podrían ser recreados y manipulados a voluntad por los programadores de la IA cuántica. Realmente *Devs* lleva el juego a su máxima expresión y deja las simulaciones de inteligencia como un arte menor del pasado para competir, latín mediante, por las simulaciones de dios (Tipler, F., 1994).

### Mímesis, copia y engaño

Nathan: Sí, así es, Caleb. Lo has entendido.  
Porque si se supera la prueba, estarás en el centro mismo  
del mayor evento científico de la historia del hombre.  
Caleb: Si has creado una máquina consciente,  
no es la historia del hombre.  
Es la historia de los dioses.<sup>58</sup>

Hay filmes de ficción científica que, a pesar de tener grandes yerros, funcionan muy bien. Un claro ejemplo de esto es *Viaje Fantástico*.<sup>59</sup> Desde la importante ley cuadrado-cúbica a incontables tecnicismos, la miniaturización humana es un absoluto imposible *Eppur si muove...* puesto que

<sup>58</sup> En el original: *Nathan: Yeah, that's right, Caleb. You got it. Because if the test is passed, you are dead center of the greatest scientific event in the history of man. Caleb: If you've created a conscious machine, it's not the history of man. That's the history of gods.* Alex Garland *Ex Machina* (Film4/DNA Films, 2015).

<sup>59</sup> Richard Fleischer *Fantastic Voyage* (20th Century Fox, 1966).

la película ha quedado como un entrañable clásico del género *sci-fi*. Caso contrario es el de *Interstellar* del minucioso Nolan,<sup>60</sup> bajo el escrutinio científico de Kip Thorne,<sup>61</sup> quien asistió con la incorporación de sus teorías –y las últimas observaciones de agujeros negros– al núcleo argumental y visual del film (aunque quizás haya que descontar el final “multidimensional-cuántico”). Claro que no siempre sucede así, y existen grandes *flops* de la ficción científica. Consciente de esto, cinco años antes de abordar la ya mencionada *Devs*, además de su interés por el tema y conocimiento personal, Alex Garland hizo los deberes. El director se aplicó a entender el *status quaestionis* en IA para incursionar con una película que –al momento designarla como mimesis en este trabajo– mejor representa la posibilidad de simulación de una inteligencia personal por parte de un artefacto diseñado por el hombre. Pocos filmes denotan un conocimiento tan preciso de la base científica sobre la que su argumento estriba como *Ex Maquina*.<sup>62</sup> Asimismo, contó con el asesoramiento técnico de Murray Shanahan, profesor de Robótica Cognitiva en el *Imperial College* de Londres e investigador senior en *Google DeepMind* (Shanahan, 2010, 2015). Además, el diseño de producción, locaciones, fotografía, cinemática y música completan una propuesta de altísima calidad. Las tres *dramatis personae* de interés son: Nathan (Oscar Isaac) dueño de BlueBook –el buscador online más usado del mundo–; Caleb (Domhnall Gleeson) un desarrollador de IA en la empresa; y Ava (Alicia Vikander), la IA robótica desarrollada en secreto por Nathan en el laboratorio de su remota mansión. Independientemente del motivo inicial por el que Caleb visita a Nathan en su casa, la acción rápidamente pone al empleado como juez en un Test Total de Turing (TTT) sobre Ava (Bringsjord, S., Bello, P., Ferrucci, D., 2000). Este contexto permite introducir tempranamente en el guión varias líneas de discusión: la caducidad del Test de Turing original (TT) (Saygin, A.P., Cicekli, I., Akman, V., 2000), la hipótesis de la inteligencia artificial fuerte (SAI) (Good, I., 1965, Searle, J., 1980), el alto volumen de datos de entrenamiento para un aprendizaje de máquina (ML) que aspire a cierta generalidad (AGI) y la “creatividad” u originalidad en las decisiones o resolución de problemas como indicador de conciencia (Harnard, S. 1991). La prueba se desarrolla en varias sesiones que están así rotuladas en la continuidad del film. Garland consigue asociar al espectador con el entusiasmo y shock de la primera impresión de Caleb en su encuentro con Ava.<sup>63</sup> Asombro es la palabra correcta, que además hace justicia a lo que causa la fineza de movimientos de este robot de apariencia de mujer joven con expresiones faciales indistinguibles de una persona real. También suma mucho la elección del director para este set: Caleb y Ava dialogan a través del cristal de habitaciones vidriadas –como peceras o acuarios– don-

60 Christopher Nolan *Interstellar* (Paramout Pictures, 2014).

61 Nobel de Física 2017; <http://www.its.caltech.edu/~kip/index.html/>.

62 Alex Garland *Ex Machina* (Film4/DNA Films, 2015).

63 Nótese la proximidad fonética en el inglés original para “Ava” (*Ay+vuh*) y “Eva” (*Ee+vuh*).

de la cámara y la utilería hacen un juego de transparencias y reflejos. De hecho, la primera aparición de Ava permite que se vean, a través de una piel transparente, varias partes de su interior “electrónico”. Esto refuerza la intención del guión – expresada por el mismo Nathan– de que, a diferencia del viejo TT, Ava sea reconocida como un ser inteligente, a pesar de lo que es. Los reflejos en los vidrios sobrepunen ambos rostros y figuras mientras que la edición cambia alternadamente de plano entre uno y otro. En un intercambio con Nathan, la expresión que Caleb utiliza para describir su observación de Ava es como *a través del espejo*, un guiño de Garland a los acertijos lógicos de Carroll (2016). Transparencias y reflejos conforman un juego de imágenes que explora varias líneas conceptuales. ¿En qué se parecen y en qué difieren Caleb y Ava? ¿Tienen algo en común? Sin duda el lenguaje hablado y su complemento visual son relevantes, pero –conociendo cómo operan los modelos matemáticos predictivos en los LLMs– (Rothman, 2024) ¿Es eso prueba suficiente de inteligencia? ¿Es Caleb quien examina a Ava, o sucede al revés? Las cámaras ubicuas y los monitores en el laboratorio completan este juego de imágenes donde nada es lo que parece. Su disposición y seguimiento manifiestan que Nathan tiene más interés por Caleb que por su propia creación. Los diálogos que tienen ambos van sugiriendo pistas de que la situación es más complicada de lo que aparenta. Así es como Caleb se entera que Nathan ha usado ilegalmente la ubicuidad de su buscador, BlueBook,<sup>64</sup> para entrenar a Ava a partir de las conversaciones de todos los usuarios en los teléfonos móviles del planeta y de los patrones de las búsquedas que realizan en internet.<sup>65</sup> Las sucesivas sesiones de Caleb con Ava se irán tornando en un experimento más complejo. La imagen vidriosa transmite dudas respecto de la identidad del espécimen estudiado en la placa de Petri o si son ambos ratones en el laberinto vidriado de Nathan. Esto llega a un punto crítico en el que Caleb –desafiado por el contexto, la performance de Ava y las sospechas sobre las verdaderas intenciones de Nathan– desconfía de su propia subjetividad (¿humanidad?) y frente a un espejo busca, abriendo un corte en su propio cuerpo las evidencias de que él no fuera también una máquina puesta a prueba por su jefe (Garland, 2013). Caleb y Ava, involucrados en un “enamoramiento” que emula o supera lo planteado en *Her*, conspiran para escapar. Por su parte, Nathan secretamente favorece la relación y el plan como si ese fuera el verdadero objetivo de su test. Pero el proyecto se descarrila y todo se sale de control. Las escenas finales muestran a Caleb atrapado detrás de uno de los vidrios blindados y a Ava saliendo al bosque exterior en una exacta inversión de sus lugares respecto del comienzo del film. Aquí se aprecia –en el uso de la cámara, en la expresión de Ava, y en la textura de los colores– la clara analogía con el

64 Una sutil referencia al nombre de *BlueBook* que Ava revela en uno de los diálogos: Cf. (Wittgenstein, 1984 [1934]).

65 El guión dice que su posición dominante es del 94% de las búsquedas en línea.

experimento mental de Jackson.<sup>66</sup> Este plantea la diferencia fenomenológica entre la captación de conocimiento desde lo formal versus una elaboración intelectual a partir de la percepción directa.<sup>67</sup> Este cuenta que Mary pasó toda su vida en una cueva donde todo es blanco y negro, sin embargo se ha hecho experta en óptica y física de tal forma que *sabe* todo sobre la teoría del color. Finalmente Jackson arroja la pregunta: ¿En su salida al exterior, se agrega algo a su conocimiento con la experiencia real de los colores?<sup>68</sup>

En los diez años que han transcurrido desde *Ex Machina* se ha progresado enormemente en la tecnología de los robots antropomorfos y las redes neuronales artificiales pero, al momento no se ha alcanzado en la realidad, una expresión física y autonomía de ejecución tan acabada. Sin embargo, uno de los aspectos por el que el argumento de la película mejor funciona, es que no pretende zanjar el tema de la singularidad.<sup>69</sup> El film no es conclusivo respecto de si Ava es una excelente simulación de inteligencia que trazó una estrategia de escape superadora en base a las reglas de juego recibidas o si posee conciencia real –cualquiera sea el significado completo de esta última expresión–.

## Conclusiones

No hagáis intervenir a un dios, sino cuando el drama es digno de ser desenredado por un dios (Horacio, 2016, p.191).<sup>70</sup>

Se han presentado diez casos que, ya sean transposiciones o mímisis de mutua influencia entre realidad y ficción, transmiten la compleja relación del hombre con las inteligencias artificiales. Autores, directores, creadores, desarrolladores o usuarios, todos de alguna manera se encuentran representados dentro del espacio de ficción científica de IA (Capanna, P. 1966). Muchas de estas obras aplican una aproximación a las IAs denominada como *externa-cuantitativa*, en la que los aspectos que se toman en cuenta son los resultados que producen las IAs y algunas características de su operación. El lenguaje es considerado primariamente desde la lingüísti-

66 Cf. Frank Jackson. “What Mary Didn’t Know”, *The Journal of Philosophy*, Vol. 83, 5 (1986) pp. 291-295.

67 Caleb relata este experimento a Ava en una de las sesiones; Cf. Alex Garland, *Ex Machina*, Original Script (Londres: DNA Films, 2013) 64-65.

68 Cf. Peter Ludlow, Yujin Nagasawa y Daniel Stollar (eds.) *There’s Something About Mary. Essays on Phenomenal Consciousness and Frank Jackson’s Knowledge Argument* (Cambridge: MIT Press, 2004).

69 Cf. Jobst Landgrebe y Barry Smith. *Why Machines Will Never Rule the World. Artificial Intelligence without Fear* (New York: Routledge, 2023); Erik Larson. *The Myth of Artificial Intelligence. Why Computers Can’t Think the Way We Do* (Cambridge: Harvard University Press, 2024).

70 *Nec deus intersit, nisi dignus vindice nodus*, (Horacio, 2016, p.191). Sobre el abuso del *deus ex machina* en el teatro griego, algo también criticado por Aristóteles en su *Poética*.

ca, el objeto de conocimiento se explica por la *idea* neo/kantiana y no habría diferencias entre un artefacto y un sujeto desde el punto de vista del propósito del primero y la intencionalidad en el segundo. Gran parte de la ideación y desarrollo de la IA a nivel mundial tuvo, durante muchos años, esta aproximación que puede explicarse parcialmente por un legado de la matriz conceptual matemática originaria de la IA dentro de una tradición filosófica analítica anglosajona. El diálogo desde otros modos de aproximación filosófica no ha sido fácil y suele terminar en una *fuga a la ética*, problematizando consecuencias pero sin atender las causas de fondo. Dentro de ese mismo contexto, quedan grandes dificultades por resolver que son, al menos, los tres estatutos de limitación de la IA actual. La *localidad* o no-universalidad de los procesos computacionales (Gödel, K. 1931, Turing, A., 1936, Church, A. (1936), Cook, S. 1971), la *monotonía* o inercia del marco (Maruyama, Y., 2017, Silenzi, M. I., 2021), y la *semásica* o aterrizaje del signo (Harnad, S., 1990, Li, J., Mao, H. 2022). Una visión de la IA que incluya una aproximación *interna-cualitativa* permite asumir una realidad más amplia del lenguaje y afirmar que los actuales procesos funcionan como una reduplicación sígnica, también estructurar una taxonomía que diferencie el dato de la información y ésta del conocimiento, y más allá de los procesos racionales formalizables no pierda los aspectos intuitivos y eidéticos de la inteligencia junto con evitar atribuir entidad y subjetividad a un artefacto que es *partes extra partes*.<sup>71</sup>

El realismo es probablemente la mejor manera de dramatizar argumentos e ideas. La fantasía puede abordar mucho mejor temas que residen principalmente en el inconsciente. Creo que el atractivo inconsciente de una historia de fantasmas, por ejemplo, reside en su promesa de inmortalidad. Si una historia de fantasmas puede asustarte, entonces debes aceptar la posibilidad de que existan seres sobrenaturales. Si es así, entonces hay algo más que el olvido aguardando más allá de la tumba (Kubrick, 2022).<sup>72</sup>

Se convino al inicio de este trabajo en aceptar que el comienzo de la literatura de sci-fi se encuentra en *Frankenstein*. Curiosamente, muchos de los elementos conceptuales de fondo en la obra de Shelley se reiteran en los casos analizados y si se regresa al conjunto general de la narrativa de IA

71 Se deja aquí una promesa de ampliación de los estatutos de limitación (y su relación con los primeros autores en señalar la importancia del tema: Jean Ladrière; John Lucas; Michael Polanyi; Mortimer Taube) en la tesis doctoral en curso, titulada *Filosofía Natural de la Inteligencia Artificial. Realidad y Representación en la Mímesis Computacional*: Riesgo, G. (2025).

72 En el original: *Realism is probably the best way to dramatize argument and ideas. Fantasy may deal best with themes which lie primarily in the unconscious. I think the unconscious appeal of a ghost story, for instance, lies in its promise of immortality. If you can be frightened by a ghost story, then you must accept the possibility that supernatural beings exist. If they do, then there is more than just oblivion waiting beyond the grave.* Stanley Kubrick sobre *El Resplandor*; Cf. *Kubrick on The Shining. An interview with Michel Ciment* disponible en línea a Mayo 2022 en <http://www.visual-memory.co.uk/amk/doc/interview.ts.html>.

seguramente el porcentaje de coincidencia sea muy alto. De igual forma, estos elementos aparecen en obras muy anteriores, incluso consideradas como “mitos”. Por ejemplo, si bien Meyrink publicó *El Golem* en el siglo XX (Meyrink, G. 2007 [1915]), la leyenda de base precede en mucho a Shelley.<sup>73</sup> En la sinagoga del rabino Loew no hay partes humanas como en el laboratorio de Víctor, pero sí barro y una palabra cabalística esencial. Retrocediendo aún más, es posible pasar del fango de Praga al blanco mármol de Galatea en las manos de Pigmalión,<sup>74</sup> pero la historia de la sub-creación de un semejante se repite. Descontando la enorme pista que aporta *Frankenstein* con el subtítulo de: *o el moderno Prometeo* (Mary Shelley, M. W., 2013 [1818]), quizás la más antigua de estas referencias la traiga Hesíodo que hace surgir de la fragua de Hefaios a varias creaturas. De su fuego provienen el guardián Thalos, las Doncellas Doradas (ayudantes del dios-herrero)<sup>75</sup> y Pandora,<sup>76</sup> quien –¿previsiblemente?– desobedece la consigna.<sup>77</sup>

La textura dramática que comparten todos estos relatos míticos –y que reaparece transformada en las narrativas más importantes de la ficción científica de IA– podría denominarse como *el mitologuema de la creatura artificial*. En la estructura de este mitologuema hay una serie de condiciones pre-invencción: el desafío al status quo; un conocimiento técnico más avanzado; el desarrollo es secreto; y hay un *factor X* de falla. A estas características le corresponden otras notas comunes post-invencción: el resultado no esperado; los daños colaterales; hay un enfrentamiento “creatura” vs. “creador”; y no se llega a una consumación del eros, sino que se desencadena un pathos con rango que va de la decepción al horror.

Todo demasiado próximo al *fascinans* y al *tremendum* como para ignorar la advertencia de que se está entrando a un territorio sagrado. Quizás sea apropiado aceptar que la mutua inhesión entre el mito y el logos se halla presente en estas narrativas y, así como la ciencia cosmológica es tangencial a la búsqueda de sentido del universo, también lo es la inteligencia artificial respecto de la búsqueda del fundamento de la inteligencia personal.

Deckard: Ella es una replicante ¿no es así?  
 Tyrell: Está empezando a sospechar, creo.  
 Deckard: ¿Sospechar! ¿Cómo puede no saber lo que es?  
 Tyrell: El comercio es nuestro objetivo aquí en la Tyrell.  
 Y nuestro lema: “más humanos que los humanos”.

73 La antigua leyenda ya mencionaba al gran rabino Loew de Praga (1512-1609).

74 Cf. Ovidio (2016, pp. 252-255). Ver que especialmente las que apuntan a la “conciencia” de Galatea.

75 Autómatas de oro conocidas como las *Κούραι Χρυσεαί*.

76 Cf. Hesíodo (2015, pp. 42-105): *Ἔργα καὶ Ἡμέραι*.

77 Para una excelente relación entre las recepciones griegas y latinas de los mitos de la creación del hombre por Prometeo y de Pandora: Cf. (Martínez Astorino, P., 2003, PP. 335-349).

## Referencias bibliográficas

- Ackerman, F. (1999). *Science-fiction Classics The Stories that Morphed into Movies*. TVBooks.
- Aldiss, B. (1973). *Billion Year Spree: The True History of Science Fiction*. Doubleday.
- Asimov, I. (1976). *The Bicentennial Man and Other Stories*. Ballantine.
- Bernhardt, C. (2019). *Quantum Computing for Everyone*. MIT Press.
- Bohm, D., Basil H. (1983). *The undivided universe: An ontological interpretation of quantum theory*. Routledge.
- Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.
- Bringsjord, S., Bello, P., Ferrucci, D. (2000). "Creativity, the Turing Test, and the (Better) Lovelace Test". *Turing 2000*. Conference at Dartmouth.
- Bunge, M. (1974). "Technology as Applied Science". En Rapp F., (ed.). *Contributions to a Philosophy of Technology*. D. Reidel, pp. 19-39.
- Coleridge, S. T. (1993). *Biographia Literaria* (1817). John Dent.
- Canavan, G., Link, E. (eds.). (2019). *The Cambridge History of Science Fiction*. Cambridge University Press.
- Capanna, P. (1966). *El sentido de la ciencia-ficción*. Columba.
- Carroll, L. (2016) [1871]. *Through the Looking-Glass, and What Alice Found There*. Macmillan.
- Church, A. (1936). "A Note on the Entscheidungsproblem". *Journal of Symbolic Logic*, N° 1. pp. 40-41; pp. 101-102.
- Clarke, A. (1951). "The Sentinel". *10 Story Fantasy*. Avon.
- . 2001. (1968). *A Space Odyssey*. Hutchinson.
- Cook, S. (1971). "The complexity of theorem proving procedures". *Proceedings of the Third Annual ACM Symposium on Theory of Computing*, pp. 151-158.
- Crevier, D.(1993). *AI. The Tumultuous History of The Search for Artificial Intelligence*. Basic Books.
- Dennett, D. (1992). "The self as a center of narrative gravity", *Self and Consciousness: Multiple Perspectives*, Frank Kessel, Pamela Cole, Dale Johnson (eds.). Lawrence Erlbaum, pp. 4-237.
- . (2017). *From Bacteria to Bach and Back. The Evolution of Minds*. Norton & Co.
- Dick, P. (1968). *Do Androids Dream of Electric Sheep?* Doubleday.
- Ensmenger, N. (2012). "Is chess the drosophila of artificial intelligence? A social history of an algorithm". *Social Studies of Science*. N° 42 (1), pp. 5-30.
- Everett, H. (1957). *The Many-Worlds Interpretation of Quantum Mechanics*. Princeton University Press.
- Faye, J. (2019). "Copenhagen Interpretation of Quantum Mechanics". En Zalta, E. (ed.) *Stanford Encyclopedia of Philosophy*. Metaphysics Re-

- search Lab, Stanford University.
- Feynman, R. (2011). *The Feynman Lectures on Physics, Vol. III: The New Millennium Edition: Quantum Mechanics*. Basic Books.
- Gallagher, S., Zahavi, D. (2013). *La mente fenomenológica*. Alianza.
- Gibson, W. (1984). *Neuromancer*. Ace books.
- . (1986). *Count Zero*. Victor Gollancz.
- . (1988). *Mona Lisa Overdrive*. Victor Gollancz.
- Gomila, A. (2002). “La perspectiva de segunda persona de la atribución mental”. *Azalea* N° 4, pp. 123-138.
- . (2016). “La perspectiva de segunda persona: mecanismos mentales de la intersubjetividad”. *Contrastes. Revista Internacional de Filosofía*, N° 6, pp. 65-86.
- Gödel, K. (1931). “Über formal unentscheidbare Sätze der «Principia Mathematica» und verwandter Systeme I”. *Monatshefte für Mathematik und Physik*, N° 38, pp. 173-198.
- Good, I. (1965). “Speculations Concerning the First Superintelligent Machine”. *Advances in Computers*, N°6. Academic Press.
- Harnard, S. (1990). “The Symbol Grounding Problem”. *Physica D*, N° 4, pp. 335-346.
- . (1991). “Other Bodies, Other Minds: A Machine Incarnation of an Old Philosophical Problem”. *Mind and Machines*, N°1, Kluwer Academic Publishers, pp. 43-54.
- Haugeland, J. (1985). *Artificial Intelligence: The Very Idea*. MIT Press.
- Heidegger, M. (1967).[1927]. *Sein und Zeist*. Max Niemeyer.
- Hesíodo. (2015). “Los trabajos y los días”. En *Teogonía. Los Trabajos y los Días*. Gredos.
- Hewstone, M. (1989). *Causal Attribution. From Cognitive Processes to Collective Beliefs*. Blackwell.
- Hobbes, T. (1986) [1668]. *Leviathan*. Penguin.
- Hofstadter, D., Dennett, D. (2001). *The Mind's I: Fantasies and Reflections On Self & Soul*. Basic Books.
- Horacio. (2016). “Ars Poetica”. En *Sátiras. Epístolas. Arte Poética*. Gredos.
- Jammer, M. (1974). *Philosophy of Quantum Mechanics: The interpretations of quantum mechanics in historical perspective*. Wiley.
- Jackson, F. (1986). “What Mary Didn't Know”. *The Journal of Philosophy*, Vol. 83, N° 5, pp. 291-295.
- Jaynes, J. (1976). *The Origin of Consciousness in the Breakdown of the Bica-meral Mind*. Houghton Mifflin.
- Kissinger, H. (2024). “How the Enlightenment Ends” *The Atlantic*. Disponible en línea <https://www.henryakissinger.com/articles/how-the-enlightenment-ends/>
- Kurzweil, R. (2004). *The Singularity is Near: When Humans Transcend Biology*. Penguin Books.
- . (2024). *The Singularity is Nearer: When We Merge with AI*. Penguin Books.

- Landgrebe, J., Barry S. (2023). *Why Machines Will Never Rule the World. Artificial Intelligence without Fear*. Routledge.
- Larson, E. (2024). *The Myth of Artificial Intelligence. Why Computers Can't Think the Way We Do*. Harvard University Press.
- Lee, H.-P. et al. (2025). *The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers*. ACM CHI.
- Li, J., Heng M. (2022). "The Difficulties in Symbol Grounding Problem and the Direction for Solving It". *Philosophies* 7, n° 5, p. 108.
- Ludlow, P., Yujln N., Stollar, D. (eds.). (2004). *There's Something About Mary. Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*. MIT Press.
- McCarthy, J. (1955). *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*. Consultado en mayo de 2022. <http://www.formal.stanford.edu/jmc/history/dartmouth.pdf>.
- Martínez Astorino, P. (2003). "El relato hesiódico de Pandora y sus incidencias en la cosmogonía ovidiana". *Cuadernos de Filología Clásica. Estudios Latinos*, 23 (2), pp. 335-349.
- Maruyama, Y. (2017). "The Frame Problem, Gödelian Incompleteness, and the Lucas-Penrose Argument: A Structural Analysis of Arguments About Limits of AI, and Its Physical and Metaphysical Consequences". En Vincent Müller, (ed.). *Philosophy and Theory of Artificial Intelligence*, Springer, pp. 194-206.
- Mubin, O.; Wadibhasme, K.; Jordan, P.; Obaid, M. (2019). "Reflecting on the Presence of Science Fiction Robots in Computing Literature". *ACM Transactions on Human-Robot Interaction*, Vol. 8, No. 1, 5.
- Nosta, J. (19 de enero de 2025). "The Shadow of Cognitive Laziness in the Brilliance of LLMs". *Psychology Today*. <https://www.psychologytoday.com/us/blog/the-digital-self/202501/the-shadow-of-cognitive-laziness-in-the-brilliance-of-llms>
- Obaid, A.; Bhushan B.; Muthmainnah S.; Suman Rajest, S. (eds.). (2023). *Advanced Applications of Generative AI and Natural Language Processing Models*. IGI Global.
- Ovidio. (2016). *Metamorphoseis*, I-V. Gredos.
- Pinar S., Ayse, I. C., Varol A. (2000). "Turing Test: 50 Years Later". *Minds and Machines* 10, pp. 463-518.
- Platón. (2016). *Phaedrus*, Diálogos III. Gredos.
- Riesgo, G. (2023). *Entender o Calcular. Aspectos Cognitivos del Aprendizaje de Máquina y el Lenguaje Natural*, Tesis de Maestría, Instituto de Filosofía. Universidad Austral.
- . (2025). "Hipálage. El Efecto Eliza desde la perspectiva de la atribución mental y el desplazamiento epistémico de la inteligencia personal". En *La Persona entre lo individual y lo comunitario*. Pontificia Universidad de Salamanca.
- . (2025). "No sos vos, somos nos. Sobre la necesidad de modificar la

- narrativa de la Inteligencia Artificial al plural”. En Bellomo, S., Vincur, T. (eds.). *Alfabetización e Inteligencia Artificial*. Universidad Austral.
- Rothman, D. (2024). *Transformers for Natural Language Processing and Computer Vision*. Packt Publishing.
- Ryle, G. (2009) [1949]. *The Concept of Mind*. Routledge.
- Searle, J. (1980). “Minds, brains, and programs”. *Behavioral and Brain Sciences* 3, pp. 417-457.
- Shakespeare, W. (1993). *The Tempest*. Oxford University Press.
- Shanahan, M. (2010). *Embodiment and the inner life: Cognition and Consciousness in the Space of Possible Minds*. Oxford University Press.
- . (2015). *The Technological Singularity*. MIT Press.
- Spaemann, R. (2010). *Personas. Acerca de la distinción entre “Algo” y “Alguien”*. EUNSA.
- Silenzi, M. I. (2021). “Acerca del Problema de Marco Original y Algunos Otros Problemas Relacionados”. *Tópicos* 42, pp. 189-215.
- Simondon, G. (1989). *Du mode d'existence des objets techniques*. Aubier.
- . (1998). *Imagination et Invention*. Éditions de la Transparence.
- . (2005). *L'individuation à la lumière des notions de forme et d'information*. Éditions Jérôme Millon.
- Stork, D. G., (ed.). *HAL's Legacy: 2001's Computer as Dream and Reality*. MIT Press.
- Sutor, R. (2019). *Dancing with Qubits: How quantum computing works and how it can change the world*. Packt.
- Taylor T., Alan D. (2020). *Rise of the Self-Replicators. Early Visions of Machines, AI and Robots That Can Reproduce and Evolve*. Springer.
- Tipler, F. (1994). *The Physics of Immortality: Modern Cosmology, God, and the Resurrection of the Dead*. Anchor Books.
- Todorov, T. (2006). *Introducción a la literatura fantástica*. Premia.
- Tolkien, J. R. R. (1983). “On Fairy-Stories”. En *The Monsters and the Critics and Other Essays*. George Allen & Unwin Press, pp. 109-161.
- Tolkien J. R. R., Tolkien, C. (eds.). “Quenta Silmarillion: Of Aulë and Yavanna”. En *The Silmarillion*. Harper-Collins.
- Turing, A. (1936). “On Computable Numbers, with an Application to the Entscheidungsproblem”. *Proceedings of the London Mathematical Society*, n° 42, pp. 230-265. <https://doi.org/10.1112/plms/s2-42.1.230>.
- Vedral, V. (2006). *Introduction to Quantum Information Science*. Oxford University Press.
- Westfahl, G. (ed.). (2005). *Science Fiction Quotations*. Yale University Press.
- Wittgenstein, L. (1984).[1934]. *Werkausgabe in 8 Bänden. Band 5: Das Blaue Buch*. Suhrkamp.
- Wollstone Shelley, M. (2013) [1818]. *Frankenstein o el moderno prometeo*. Valdemar.
- Yizhou F., Luzhen T., Huixiao L., Kejie S., Shufang T., Yueying Z., Yuan S., Xinyu L., Dragan G., (2025). *Beware of Metacognitive Laziness: Effects*

*of Generative Artificial Intelligence on Learning Motivation, Processes, and Performance*. British Journal of Educational Technology. Volumen 56 (2), pp. 489-530. Disponible en línea en arXiv:2412.09315.

Zahavi, D. (2005). "Self and Other". En *Subjectivity and Selfhood*. The MIT Press, pp. 147-177.

# Lógica Natural vs. Lógica Formal

## Un debate inaplazable

Enrique Alonso

### Introducción

En esta contribución pretendo repasar los fines y objetivos de dos formas muy distintas de entender el objetivo de la Lógica: una canónica, la defendida por la Lógica Formal o matemática del siglo XX, y otra que se abre paso gracias, ante todo, a la explosión de la IA-generativa<sup>1</sup>, aunque existiera antes incluso de la popularización de esta última.

Presentaré las que, a día de hoy, son las principales escuelas que desarrollan su actividad en el ámbito de la Lógica Natural, para exponer algunas de mis críticas al respecto. Finalmente dedicaré algún espacio a analizar si la Lógica Natural podría aspirar a ser un reemplazo razonable de la Lógica Formal en los planes de estudio de los Grados de Humanidades.

### Lógica Formal vs. Lógica Natural

Como parece obvio, toca fijar qué vamos a entender aquí por *Lógica Formal* o *matemática* y sobre todo por *Lógica Natural*. Definir la primera de ellas no requiere especial esfuerzo ya que existe una larga tradición al respecto.

1. La *Lógica Formal* es la disciplina que estudia la noción de *consecuencia* desde el punto de vista de la estructura de los enunciados que constituyen las premisas y la conclusión.

En cuanto a la Lógica Natural, la cuestión es más disputada, pero existe un cierto consenso que podría resumirse en las siguientes palabras de Lawrence S. Moss: “Por lógica natural entiendo el estudio de la inferencia en el lenguaje ordinario, realizado tan cerca como sea posible de sus ‘formas explícitas’ ” (Moss, 2010, p. 84).

Por tanto podríamos decir que:

2. La *Lógica Natural* se centra en el estudio de las inferencias que tienen lugar en el discurso o razonamiento ordinario intentando respetar la forma de tales inferencias en el mismo lenguaje en el que tienen lugar.

Esta definición constituye una suerte de declaración de intenciones con respecto

---

<sup>1</sup> La IA-generativa se apoya en la tecnología del *machine learning* y está a su vez en el modelo de *redes neurales*. Se trata este último de un paradigma que no gozó de la atención de la comunidad del mismo modo que lo hizo el orientado al diseño de software específico a través de lenguajes de alto nivel, como LISP, por ejemplo. Queda claro qué modelo parece finalmente el ganador de esta competición.

a la Lógica Formal. El análisis de la consecuencia lógica y, en consecuencia, de las razones por las que obtenemos conclusiones a partir de ciertas piezas de información no se *explicarían* ya recurriendo a la Lógica Formal, sino a las herramientas típicas del estudio de los lenguajes naturales: Gramática generativa, categorial, etc.

Como veremos más adelante, esta orientación reconecta de una manera nada disimulada con la Silogística aristotélica, aunque sin rechazar por ello mismo las herramientas de la Matemática, la Gramática y la Lógica formales (Van Benthem, 2008, pp. 13-14), pero ya habrá momento de hablar de ello.

Es fácil entender que la tensión principal entre estas dos maneras de concebir el objetivo de la disciplina de la Lógica reside en la conveniencia, o no, de considerar un lenguaje distinto del ordinario para estudiar la relación de consecuencia. En el caso de la Lógica Formal la decisión es, al menos desde la decisiva contribución de Frege, la de introducir un lenguaje formal capaz de representar adecuadamente la estructura de los enunciados que justifican o soportan una inferencia.

Estas discrepancias con lo tradicional encuentran su justificación en que la lógica, hasta ahora, siempre se ha ajustado muy estrechamente al lenguaje y a la gramática. En especial, creo que la sustitución de los conceptos de sujeto y predicado por los de argumento y función, se acreditará con el tiempo (Frege, 1972, p. 4).

Según esta tradición, que es la que nos acompaña hasta el presente, la tarea del lógico difiere sustancialmente de la del gramático precisamente en este punto. El lógico puede y debe generar un lenguaje que cuente con las categorías apropiadas para representar la estructura de un enunciado, al menos aquellas que permiten dar cuenta de sus relaciones inferenciales. Dichas categorías estarán presididas por la dicotomía fundamental *argumento y función* y no por la tradicional de *sujeto y predicado*. En consecuencia, la Lógica Formal puede ingeniar lenguajes cuya estructura y aspecto disten de los del lenguaje ordinario tanto como sea preciso. Lo único que se exige es que cumpla con la tarea de explicar la inferencia a través de los elementos estructurales identificados.

Esta es la razón por la que en un curso de Lógica Formal elemental sostengamos cosas como la siguiente:

### 3. La estructura del enunciado:

*Todos los hombres son mortales es*

$\forall x (Hx \rightarrow Mx)$

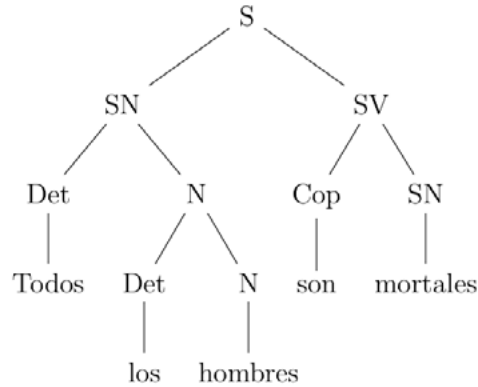
Sin aclarar, en ocasiones, que esa pretendida estructura, casi *inmanente*, no es otra cosa que la que resulta de un proceso de traducción a un lenguaje previamente definido, el de la Lógica de Primer Orden –FOL– en este caso.

Desde la Lógica Natural se intenta evitar este tipo de maniobra reemplazando las herramientas de análisis lógico por las propias de la Gramática, en

particular la Generativa y la Categorial.

Una forma común de expresar y analizar en Lógica Natural la estructura de un enunciado como el anterior podría ser la propia, por ejemplo, de la Gramática Generativa:

4.



Pero podríamos haber considerado igualmente la gramática categorial o el cálculo lambek, por ejemplo (Van Benthem, 1991).

Todas estas formas de analizar la estructura de un enunciado basadas en la Gramática de los lenguajes ordinarios han tenido y tienen uso, en los estudios actuales en Lógica Natural. ¿Pero es realmente este giro el que caracteriza a la Lógica Natural frente a la tradición de la Lógica Formal? Es posible que, en su origen, constituyera una idea interesante promovida por los muchos fracasos que la Lógica Formal acumulaba y acumula en el estudio del discurso ordinario. Sin embargo, el paso de los años y la experiencia adquirida han ido debilitando sustancialmente ese aspecto tan aparentemente definitorio de la Lógica Natural. Se ha generado una cierta tensión entre la estructura superficial de nuestras expresiones e inferencias y las herramientas que se hacen necesarias para explicar las relaciones inferenciales entre premisas y conclusión. Estas últimas pueden llegar a ser realmente complejas imitando o reproduciendo, en definitiva, la presunta artificialidad que el modelo de la Lógica Formal impone en su análisis de las inferencias que tienen lugar en el lenguaje ordinario. Pero de todo ello hablaremos con algo más de detalle en breve.

En definitiva, el balance entre una y otra iniciativa ya no es tan claro en este punto. El uso de lenguajes formales artificiales para el estudio de la inferencia ordinaria en la Lógica Formal, en el caso de la Lógica Natural quedaría compensado por el empleo de herramientas complejas basadas en la Gramática Categorial, el Cálculo Lambek o la Teoría de Tipos, por citar solo algunos ejemplos.

Sin embargo, este comentario no intenta ocultar o minimizar las dificultades de la estrategia seguida por la Lógica Formal que se basa, en lo que podríamos denominar, el principio de *traducción como análisis*. Es decir, el diseño de un lenguaje formal al que traducir los enunciados del lenguaje

ordinario como forma de exponer su estructura.

Es evidente que esta estrategia no ha logrado satisfacer de forma simultánea estos dos requisitos:

5.

i. Obtener un lenguaje suficientemente potente desde el punto de vista expresivo y

ii. Diseñar herramientas que permitan decidir la aceptabilidad de los argumentos expresables en dicho lenguaje.

Y ello no porque estemos aún a la espera de alguna aportación trascendental, sino por la misma imposibilidad de la empresa.<sup>2</sup>

La estrategia habitual en Lógica Formal, y que hemos practicado durante un siglo largo, ha pasado siempre por estudiar pequeños fragmentos de las inferencias ordinarias generando lenguajes en los que quepa representar su estructura. Es la vieja conocida táctica de *divide et impera*. Es así como se llegaron a establecer las divisiones que aún están vigentes en el estudio de los sistemas formales: lenguaje sentencial, Primer Orden, Orden Superior, lenguaje modal, lenguaje temporal, deóntico, etc. La idea subyacente consistía en integrar finalmente todos estos estratos en un lenguaje universal obtenido por agregación. Pero todos sabemos que esto no es posible. Cada vez que se intenta generar un sistema complejo capaz de integrar varios de estos niveles se obtiene una suerte de modelo *frankstein* apenas manejable y con muy pocas ventajas técnicas que ofrecer. Esta es la maldición de la Lógica Formal contemporánea: cada logro en capacidad expresiva implica una consiguiente pérdida en control computacional, es decir, en capacidad para distinguir las inferencias correctas de aquellas que no lo son.<sup>3</sup>

Lo que procede en este momento es una discusión más detallada sobre el modo en que la Lógica Natural lidia con estas mismas dificultades.

### La escuela holandesa

A finales de la década de 1980, un grupo informal<sup>4</sup> liderado por Johan van Benthem en la Universiteit van Amsterdam dio inicio a una serie de estudios cuyo rasgo fundamental era el reemplazo de los lenguajes formales, estándar en el estudio de la inferencia, por herramientas procedentes de la Gramática formal (Van Benthem, 1990).

<sup>2</sup> Me refiero, claro está, a los teoremas de limitación sobradamente conocidos y estudiados desde su formulación durante el segundo tercio del siglo XX.

<sup>3</sup> Esta maldición, expuesta en muchas ocasiones como una suerte de balanza, ha sido material común en los cursos impartidos por María Manzano y yo mismo en el Máster de Lógica y Filosofía de la Ciencia impartido en España a través de un consorcio de varias de las principales universidades del país.

<sup>4</sup> Dicho grupo, conocido entonces como *Institute for Logic, Language and Information*, daría lugar en la década siguiente a la creación del *Institute for Language, Logic and Computation* que permanece hasta nuestros días.

El trabajo más destacado dentro de esa pequeña comunidad fue, durante algún tiempo, la tesis doctoral defendida por Victor Sánchez-Valencia en 1991 (Sánchez-Valencia, 1991), en la que se ofrece un elaborado tratamiento de las inferencias basadas en la dicotomía *monotonía/antitonía* existente entre los términos léxicos de nuestros enunciados. Lo explicaré con un ejemplo absolutamente elemental:

6.  
 Todos los griegos son virtuosos.  
 Los atenienses son griegos.  
 Por tanto,  
 Los atenienses son virtuosos.

Como se puede ver, se trata de una forma elemental de silogismo, en particular una instancia de la figura conocida como *Bárbara*.

La explicación contemporánea de la corrección de dicho argumento se obtiene a partir de su traducción al lenguaje de FOL y de la demostración mediante las herramientas apropiadas de que el argumento expuesto en [7] es formalmente correcto.:

7.  
 $\forall x(Gx \rightarrow Vx)$   
 $\forall x(Ax \rightarrow Gx)$   
 Por tanto,  
 $\forall x(Ax \rightarrow Vx)$

Para Sánchez-Valencia existe una forma mucho más aquilatada de justificar este tipo de inferencias. Sin ser totalmente fieles a su formalismo, se podría exponer del siguiente modo:<sup>5</sup>

8.  
 Every G is V  
 $[A] \subseteq [G]$   
 Every A is V

Lo que tiene lugar en este tipo de inferencias, es un silogismo en *Bárbara*. En este caso en particular, se trata de la substitución de un término léxico por otro en función de dos piezas de información:

- i. la relación entre el término substituido y el substituyente en términos de extensión:  $[A] \subseteq [G]$ , y

<sup>5</sup> En [Sánchez-Valencia, 1991, p.16 y ss] se ofrece una traducción de las figuras del silogismo a la notación prevaleciente en su trabajo.

ii. la posición del término léxico sustituido,  $G$  en este caso, en el enunciado en el que figura.<sup>6</sup>

A su vez, los puntos i. y ii. dependen de la partícula cuantificacional, cuantor, dominante en cada expresión. En el caso, por ejemplo, de que el término léxico a ser sustituido en [8] fuera  $V$ , la relación léxica con el término sustituyente, llamémosle  $[C]$  debería ser esta:  $[V] \subseteq [C]$ , dando lugar a:

- 9.
- Every  $G$  is  $V$   
 $[V] \subseteq [C]$   
 Every  $G$  is  $C$

Es decir, los términos léxicos que intervienen en una expresión cuantificacional universal se comportan exigiendo relaciones inversas con el término reemplazante según la posición que ocupan. El término  $G$  puede ser reemplazado por uno de menor extensión, mientras que el término  $V$  lo tendrá que ser por uno de menor extensión. Esto es lo que nos lleva a afirmar que el comportamiento de  $V$  en “Every  $G$  is  $V$ ” es *monótono*, mientras que el de “ $G$ ” es *antitono*. Este rasgo parece, pues, estar asociado a la partícula cuantificacional en primera instancia, dando lugar, en el caso del cuantor existencial, a una conducta monótona en ambos términos:

- 10.
- Some  $G$  is  $V$   
 $[G] \subseteq [A]$   
 Some  $A$  is  $V$

y

- 11.
- Some  $G$  is  $V$   
 $[V] \subseteq [A]$   
 Some  $G$  is  $A$

A este comportamiento de un término léxico dentro de una expresión se le denomina *su polaridad* y se representa como  $(\uparrow)$  en el caso de la monotonía, y como  $(\downarrow)$  en el caso de la antitonía. En consecuencia se puede sostener ahora lo siguiente:

- 12.
- i.  $Q_{\forall}(A, B) = (\downarrow\uparrow)$ , estando  $A$  y  $B$  afirmados.  
 ii.  $Q_{\exists}(A, B) = (\uparrow\uparrow)$ , estando  $A$  y  $B$  afirmados.

<sup>6</sup> El término al que aquí nos hemos referido como el término sustituido es lo que en la tradición silogística se ha venido denominando término medio.

Es decir, ( $\downarrow\uparrow$ ) representaría la polaridad de los términos léxicos sobre los que el cuantor universal actúa, mientras que ( $\uparrow\uparrow$ ) supondría lo propio con respecto al cuantor existencial. Obsérvese que esta forma de tratar a los cuantores difiere, y de manera sustancial, de la que es típica en Lógica formal. Ahora se interpretan, básicamente, como operadores binarios que toman como argumentos términos léxicos, y no como constantes lógicas que operan sobre fórmulas. Creo que a nadie se le escapa que esta maniobra, motivada por el análisis subyacente en el lenguaje ordinario, complica, y mucho, el manejo de expresiones que en el lenguaje de FOL resultan elementales.<sup>7</sup>

Llegados a este punto resulta evidente que el objetivo principal de la propuesta holandesa tiene mucho que ver con la identificación de la polaridad de cada término léxico en contextos de complejidad creciente. Es notable el esfuerzo hecho, a este respecto, y también la complejidad de los algoritmos diseñados a tal efecto. La idea detrás de esta concepción de la Lógica Natural no puede ser más sencilla y al mismo tiempo más sugerente. Las inferencias que los seres humanos hacemos en contextos reales se basan en una operación muy simple: reemplazar términos léxicos por otros cuya relación mutua se conoce,<sup>8</sup> para dar lugar a un enunciado nuevo que actúa como conclusión del primero. Para ello nos basta con conocer la polaridad del término léxico en cuestión, en la expresión en la que va a ser reemplazado.

Obsérvese que, bajo esta perspectiva, una inferencia natural es tan solo el tránsito de un enunciado a otro que se obtiene del inicialmente dado al reemplazar un término léxico dentro de su estructura. Ese reemplazo viene legitimado por una pieza de conocimiento, tal vez implícita, que garantiza que la sustitución puede ser hecha. Como puede verse, es algo que difiere, y de forma no menor, de la concepción tradicional de la consecuencia lógica. En Lógica Formal, las inferencias basadas en meras transformaciones léxicas nunca han tenido un tratamiento diferenciado.<sup>9</sup> El objetivo era, por el contrario, caracterizar la conducta de las constantes lógicas descritas en el lenguaje. Es así como podemos entender algunas de las polémicas más características del siglo pasado, como por ejemplo aquella que priorizaba los cálculos basados en *Deducción Natural*, *Tablas analíticas* o *Cálculo de Secuentes* sobre los *Sistemas axiomáticos*. En los primeros, las reglas describen de forma independiente el comportamiento de cada constante, mientras que en los *Sistemas axiomáticos* se hace, más bien, un estudio de sus interacciones.<sup>10</sup>

En definitiva, todo indica que estaríamos descubriendo una tensión nada disimulada entre una suerte de *Lógica de términos* y una *Lógica de constantes*,

<sup>7</sup> Por ejemplo, el manejo de relaciones de cualquier ariedad con cuantores anidados.

<sup>8</sup> Determinada en este caso por la relación de inclusión.

<sup>9</sup> Ni siquiera la regla de sustitución uniforme puede ser entendida en ese sentido.

<sup>10</sup> De ahí, por ejemplo, la polémica en torno a la conectiva *plonk* (Belnap, 1962).

o alternativamente entre una *Lógica categoremática* y otra *Lógica sincategoremática*. La *Lógica Natural* se podría considerar, entonces, como un intento de recuperar la tradición del estudio de las relaciones entre los términos léxicos y, por lo tanto, constaría como un ejemplo de *Lógica categoremática*. Lo que quiero hacer explícito con estas consideraciones es, en primer lugar, que *Lógica Natural* y *Lógica Formal* difieren en aspectos sustanciales que nos llevan muy atrás en el orden de las decisiones adoptadas a lo largo del tiempo. En este sentido, la *Lógica Natural* sería pre-fregeana. Y en segundo lugar, que sus diferencias no se expresan por completo en el hecho de adoptar un lenguaje formal o las herramientas de la Gramática como instrumento de análisis. En ambos casos, se trata de aspectos que nos llevan más atrás en el tiempo del hito fundacional representado por Frege. Pero creo que el primero de ellos es, en cierto sentido, más radical y definatorio que el segundo.

Veamos si estas características se aprecian igualmente en la otra gran escuela dentro del ámbito de la *Lógica Natural*.

¿Dónde interviene la Gramática entonces? Como ya he dicho, en el estudio de la polaridad de los términos léxicos en expresiones complejas. El algoritmo descrito por Sánchez-Valencia (Sánchez-Valencia, 1991, cap. 5 y 6) reposa en un ingenioso empleo de la Teoría de Tipos sobre una estructura básica aportada por el Cálculo Lambek. Debo indicar, como ya insinué páginas atrás, que el uso de las reglas, de lo que Sánchez-Valencia denomina como un *System of Natural Logic*, no es sencillo ni tiene por qué serlo. Sí es importante indicar que con dicho sistema, Sánchez-Valencia consigue ir más allá, tanto de la Silogística como de la *Lógica de Primer Orden*. Se supera la Silogística al incorporar de forma natural verbos transitivos, típicamente caracterizados como relaciones binarias en LPO, algo que la Silogística no supo manejar adecuadamente. Y va más allá de la *Lógica de Primer Orden* al incorporar de manera bastante natural los cuantores generalizados, al menos los más característicos: *MOST* y *FEW* (Sánchez-Valencia, 1991, p. 136).

### **La Escuela americana**

En este caso las referencias históricas me resultan menos familiares, por lo que no podría ofrecer un lugar concreto o un acontecimiento particular como actos fundacionales de esta escuela. Sin embargo, resulta relativamente fácil identificar a una serie de autores reconocibles por sus colaboraciones e intereses mutuos. Entre ellos se encuentran al menos los siguientes: B. MacCartney, T.F. Icard, y L.S. Moss.

El punto de partida de sus análisis lo constituye, en buena medida, los trabajos de la escuela holandesa en torno a la monotonía y en particular el estudio que Sánchez-Valencia desarrolla sobre ese tema (Sánchez-Valencia, 1991). En este sentido, la posición de la Escuela americana podría considerarse como un intento de ampliar el alcance que el estudio de la monotonía ofrece en el contexto de una *Lógica de términos* o *categoremática*, como aquí la hemos denominado.

Si tomamos la teoría de clases como marco de estudio para el análisis de los términos léxicos, no tenemos por qué limitarnos al estudio de relaciones  $A \subseteq B$  y  $A \supseteq B$  que son, precisamente las de monotonía y antitonía tomando el término léxico  $A$  como referencia.

En la escuela americana se señalan hasta siete relaciones primitivas que son las que se indican en el siguiente cuadro (MacCartney and Manning, 2009, p. 143):

Símbolo	Denominación	Ejemplo	Definición
$x \equiv y$	equivalencia	niño $\equiv$ chaval	$x=y$
$x \sqsubset y$	monotonía	cuervo $\sqsubset$ pájaro	$x \subset y$
$x \sqsupset y$	antitonía	europeo $\sqsupset$ francés	$x \supset y$
$x \wedge y$	negación	humano $\wedge$ no humano	$x \cap y = \emptyset \wedge x \cup y = U$
$x   y$	alternancia	gato $ $ perro	$x \cap y = \emptyset \wedge x \cup y \neq U$
$x \smile y$	covertura	animal $\smile$ no humano	$x \cap y \neq \emptyset \wedge x \cup y = U$
$x \# y$	independencia	hambriento $\#$ hipopótamo	el resto de los casos

Tabla 1: Conjunto  $\mathfrak{B}$  de relaciones semánticas básicas

Este listado se obtiene de manera combinatoria y no tiene mucho sentido. Entretenernos en ello en este momento. Ahora tomaremos un ejemplo ofrecido por MacCartney y Manning (MacCartney and Manning, 2009, p. 145) para aclarar el modo en que se puede operar con estas relaciones básicas. Vamos a considerar tres términos léxicos: *fish*, *human* y *nonhuman*, para analizar algunas relaciones entre ellos tomados de dos en dos.

13.

i. *fish* | *human*

ii. *human*  $\wedge$  *nonhuman*

iii. *fish*  $\sqsubset$  *nonhuman*

Estos autores consideran que lo que está teniendo lugar, al relacionar términos léxicos de esta manera, es una suerte de combinatoria que analiza las relaciones entre términos léxicos como los descritos en la tabla anterior. No se trata de las propias relaciones entre términos léxicos, sino de las relaciones entre esas mismas relaciones. Esto da lugar a la siguiente definición:

14.

$$R \triangleright \triangleleft S \stackrel{\text{def}}{=} \{ \langle x, z \rangle : \exists y (\langle x, y \rangle \in R \wedge \langle y, z \rangle \in S) \}$$

Es decir, las relaciones  $|$  y  $\wedge$ , que serían en este caso,  $R$  y  $S$ , producen otra

relación  $\sqsubset$  que se da entre los términos léxicos  $x$  y  $z$ , que en el ejemplo resultarían ser *fish* y *non-human*. Si se analiza con detalle las expresiones listadas en [13] se observa que estas dan lugar a una inferencia que puede expresarse del siguiente modo:

14.

Ningún pez es un ser humano.

Cualquier individuo, o bien es humano o bien no lo es,

**Por tanto,**

Todo pez es no humano

Es inevitable encontrar similitudes con la Silogística una vez más, solo que en esta ocasión disponemos de recursos expresivos que van más allá de ella.

El siguiente paso consiste en describir tres tipos de posibles actuaciones sobre los términos léxicos de un enunciado:

15.

i. Inserción ( $ins(A_i)$ )

ii. Borrado ( $del(A_i)$ )

iii. Substitución ( $(sub(A_i, A_j))$ )

Veamos ahora cómo funciona este sistema con un ejemplo elemental que, no obstante, simplificaré un poco más, ofrecido MacCartney y Manning (MacCartney and Manning, 2009, p. 153):

15.

i	$x_i$	$e_i$	$\beta(x_0, x_i)$
0	Stimpy is a cat		
1	Stimpy is a dog	sub(cat, dog)	
2	Stimpy is not a dog	ins(not)	$\sqsubset$
3	Stimpy is not a poodle	sub(dog, poodle)	$\sqsubset$

En la 4 columna de la primera línea se lee | lo que indica que la relación  $\beta(x_0, x_1)$ , entre “Stimpy is a cat” y “Stimpy is a dog” es de *alternancia* –véase la tabla de las relaciones elementales–. A su vez, “Stimpy is a dog” se ha obtenido a partir de “Stimpy is a cat” por medio de la operación  $sub(cat, dog)$ . En la línea 2, lo que vemos es que “Stimpy is not a dog” se obtiene a partir

de “Stimpy is a dog” mediante la operación *ins(not)* sobre esta última. Este movimiento provoca que la relación entre “Stimpy is a cat” y “Stimpy is not a dog” es  $\sqsubset$ . Por último, llegamos en la línea 3 a afirmar que la relación entre  $x_0$ , es decir, “Stimpy is a cat” y  $x_3$ , “Stimpy is not a poodle” es, de nuevo,  $\sqsubset$ . Esto supone que entre  $x_0$  y  $x_3$  existe una relación de consecuencia.

El mecanismo procede mediante una serie de pasos –*edits* (MacCartney and Manning, 2009, p.152)– en los que la relación entre el enunciado original,  $x_0$  y los subsiguientes va variando en función de las relaciones entre los términos léxicos que se ven implicados en cada caso. Lo realmente importante es que la relación entre el primer enunciado  $x_0$  de la serie y el último, sea este  $x_n$ , venga dada por  $\beta(x_0, x_n) = \sqsubset$ , lo que, como acabo de indicar, introduce una relación de consecuencia entre ambos enunciados.

Creo que el resto de la historia es bastante fácil de imaginar. Incorpora parte del modelo holandés de una forma natural y conecta, además, con el ámbito de la IA de una forma mucho más directa. El tipo de cadenas como la que se ilustra en [15] reflejan muy bien el género de manipulaciones léxicas que intentan implementarse en los diversos modelos de IA-gen de los que tanto se habla ahora. Más en concreto, son el fundamento de los sistemas de *NLP* –*Natural Language Processing*– que sirven, a su vez, para el diseño de las redes neurales en que se basan las IA-gen.

Es fácil entender el modo en que cadenas como la que hemos analizado en [15] son capaces de producir contenido nuevo a partir de la información aportada por un usuario. No se trata de un análisis del modo en el que nuestra cognición se comporta con respecto a ciertas estructuras inferenciales, sino de la forma en que generamos discurso consistente a partir de una información previamente dada. Tendremos ocasión más adelante de desarrollar este punto. Como puede verse, el sistema desarrollado resulta ser igualmente una *Lógica de términos*, pero difiere sustancialmente del modelo holandés en aquello que se entiende por un argumento. Dedicaremos el siguiente apartado a discutir por extenso este punto.

### Sobre la noción de argumento

Resulta extraño pensar que una noción tan elemental como la de *argumento* pueda verse tensionada luego de haber gozado de un largo periodo de consenso dentro de la comunidad científica. Pero tal y como dije líneas atrás, estamos ante movimientos que replantean algunas de las decisiones más características de la fundación de la Lógica contemporánea.

En Lógica formal entendemos por argumento lo siguiente:

16.

$$A_1, \dots, A_i, \dots, A_n \vdash B$$

donde  $A_1, \dots, A_i, \dots, A_n, B$  son fórmulas previamente definidas dentro de un lenguaje formal dado.  $A_1, \dots, A_i, \dots, A_n$  son las premisas y  $B$  la conclusión. Si  $B$  se *sigue* de las premisas  $A_1, \dots, A_i, \dots, A_n$  entonces decimos que el argumen-

to es correcto, en caso contrario incorrecto.<sup>11</sup>

Para el modelo holandés la cuestión parece ser muy distinta. Tal y como he dicho al presentar esta escuela, una inferencia natural parece estar basada en el reemplazo de un término léxico por otro en una expresión, de tal modo que se establezca entre dichas expresiones una relación de consecuencia. Para ello se precisan dos cosas: información acerca de la relación entre los términos léxicos involucrados –monotonía o antitonía– y la polaridad que presenta el término reemplazado en la expresión dada inicialmente. Un argumento, visto desde esta perspectiva es, de hecho, un argumento correcto. Visualicémoslo con un esquema apropiado:

17.

$$\frac{\varphi(A_1, A_2 \dots A_i \dots A_n)}{\varphi(A_1, A_2 \dots A_j \dots A_n)} \quad (A_i \subseteq A_j)$$

$$\frac{\varphi(A_1, A_2 \dots A_i \dots A_n)}{\varphi(A_1, A_2 \dots A_j \dots A_n)} \quad (A_i \supseteq A_j)$$

Para empezar, aquí no hablamos de premisas, sino de una única expresión inicial. La información relativa a la relación entre los términos léxicos bajo escrutinio  $-(A_i \subseteq A_j)$  o  $(A_i \supseteq A_j)$ <sup>12</sup> – no parece contar como una premisa más, sino como una especie de *garantía* o *justificación* de la conclusión. Esta es la razón por la que se puede sostener que esta idea de argumento presenta, en realidad, lo que sería un *argumento correcto*, y no un argumento en general, como ocurre en Lógica Formal.

Esto mismo ocurre con las figuras del Silogismo, en las que se listan argumentos correctos formados por dos premisas y una conclusión con una estructura fija, dada por la distribución de los términos léxicos en las premisas y en la conclusión, y la tipología de los juicios, proposiciones, presentes.

¿Qué ocurre con el modelo americano? A diferencia de lo que sucede en la Escuela holandesa, aquí podemos tomar en consideración una cantidad finita, pero ilimitada en principio, de enunciados. Obsérvese que no he hablado de premisas ya que, en realidad, solo la primera expresión de una cadena como la que figura en [15] puede considerarse como tal. Si damos forma a estas intuiciones, obtenemos lo siguiente:

11 Al emplear la expresión correcto, pretendo evitar tomar partido por una interpretación sintáctica o semántica de la corrección argumental.

12 Estos términos léxicos pueden corresponder en LPO a relaciones de cualquier aridad. Téngase en cuenta que en términos puramente conjuntistas, la monotonia/antitonía se predica de cualquier tipo de clases, ya estén formadas por individuos o n-tuplas de ellos.

18.

$\langle x_0, x_1, \dots, x_i, x_n \rangle$ , donde

- i.  $x_0, x_1, \dots, x_i, x_n$ , son enunciados.
- ii.  $x_{i+1} = \text{edit}(x_i)$ , donde un *edit* de un enunciado  $x_i$  es otro obtenido a partir de este mediante el uso de las operaciones  $\text{ins}(A_i)$ ,  $\text{del}(A_i)$ ,  $\text{sub}(A_i, A_j)$  sobre términos léxicos y
- iii.  $\beta(x_0, x_n) = \square$

Es decir, un argumento sería una secuencia de *edits* tales que entre el primer enunciado  $x_0$  y el último,  $x_n$ , existe una relación de consecuencia. Téngase en cuenta que una de estas secuencias no constituye en absoluto una prueba en sentido clásico. Las operaciones admitidas para producir nuevos elementos en la secuencia no son preservadoras de la verdad y por tanto,  $x_{i+1}$  no tiene por qué ser consecuencia de  $x_i$ . Este requisito solo se exige entre  $x_0$  y  $x_n$ .

Si se compara la noción de argumento del modelo americano [18], con la del modelo holandés [17], es inmediato observar un claro sesgo computacional en el primero de los casos. La noción de *edit*, central en esta escuela, está claramente orientada a la producción de contenido bajo determinados criterios de control. Se trata, básicamente, de *generar* nuevos enunciados a partir de uno dado, de modo que, en todo momento, se conserve noticia de la relación entre el primero y el último de los producidos. Esa relación viene dada por la función  $\beta(x_0, x_n)$  que es la que garantiza en última instancia si  $x_0$  es o no consecuencia de  $x_n$ , pero no es menos relevante que  $x_n$  guarde cualquier otra posible relación con alguno de los  $x_i$  en la secuencia de *edits*. Lo sustancial es que podamos conocerla en cada paso. Llegados a este punto, lo que procede es dar paso a la ingeniería del conocimiento diseñando algún modelo de *Natural Language Inference* –NLI– que pueda finalmente ser implementado en una aplicación funcional.

En comparación con tales objetivos, el modelo holandés parece mucho más centrado en el análisis formal y filosófico de la inferencia natural, posición que personalmente comparto y que será la que desarrollaré a continuación.

### **Una concepción extensa de la inferencia natural**

Reemplazar los términos léxicos en una expresión dada permite, si se conoce la relación precisa entre el término substituido y el substituyente,

obtener consecuencias lógicas a partir de la expresión inicial. Ciertamente se trata de una idea poderosa que dice mucho de las estrategias que empleamos en la vida cotidiana para poner en uso ciertas habilidades lógicas preformales. De hecho, creo que es justo reconocer que estaríamos ante un modelo extraordinariamente frecuente de ejercicio de nuestras habilidades formales. Pero me cuesta reconocer que las agote.

Es cierto que el reemplazo de un término léxico por otro, conservando la estructura en una oración previamente dada, resulta ser una operación de muy escaso coste cognitivo y con un gran rendimiento. No obstante, me resulta difícil aceptar, al igual que le sucede a la escuela americana, que los reemplazos léxicos deban enfocarse con tanta intensidad en el par *monotonía/antitonía*. Y esto incumbe no solo a las relaciones de ariedad 1, sino a las de cualquier ariedad. Es cierto que la monotonía se puede proyectar sin esfuerzo a relaciones de cualquier ariedad, pero creo que al obrar así se pierden o se oscurecen muchos otros tipos de relaciones entre términos léxicos que sí se tratan, por ejemplo, en FOL. Y, por ir aún un paso más allá, tampoco creo que las inferencias naturales se limiten solo a transformaciones léxicas que preserven la estructura. Creo que también es posible actuar sobre la estructura, aunque ello suponga un coste computacional mucho más alto. Lo primero que voy a hacer es proponer un esquema de argumento que generalice apropiadamente el modelo holandés tal y como lo hemos representado en [17]. Pero antes introduciremos algunas definiciones útiles:

18.

Por  $\varphi(A_1, \dots, A_n)$  se entenderá una fórmula cuyo léxico viene dado a lo sumo por los términos en  $\{A_1, \dots, A_n\}$

Recuerdo que el *léxico* en este caso no hace referencia tan solo a las relaciones 1-arias típicas de FOL –*First Order Logic*–. Un elemento léxico es, en este contexto, una variable de cualquiera de los tipos<sup>13</sup> que el lenguaje en que se formula  $\varphi(A_1, \dots, A_n)$  pueda admitir. Por ejemplo, la fórmula  $p \rightarrow q$  sería una instancia, según [18], de  $\varphi(p, q)$  mientras que  $\forall x(Px \rightarrow \exists yRxy)$  lo sería de  $\psi(Px, Rxy)$ .

La expresión más general de un argumento correcto bajo la óptica de una inferencia natural extendida podría ser la siguiente:

<sup>13</sup> Empleo la expresión tipo en un sentido técnico.

19.

$$\frac{\varphi(A_1, A_2 \dots A_i \dots A_n)}{\psi(B_1, B_2 \dots B_i \dots B_n)} \quad \theta(A_1, A_2 \dots A_i \dots A_n, B_1, B_2 \dots B_i \dots B_n)$$

donde,

20.

- i.  $\varphi(A_1, A_2 \dots A_i \dots A_n)$  es la premisa,
- ii.  $\psi(B_1, B_2 \dots B_i \dots B_n)$  es la conclusión,
- iii.  $\theta(A_1, A_2 \dots A_i \dots A_n, B_1, B_2 \dots B_i \dots B_n)$  es una fórmula con el léxico a lo sumo de la premisa y la conclusión y tal que:

$$\frac{\varphi(A_1, A_2 \dots A_i \dots A_n)}{\psi(B_1, B_2 \dots B_i \dots B_n)} \quad \theta(A_1, A_2 \dots A_i \dots A_n, B_1, B_2 \dots B_i \dots B_n)$$

21.

La fórmula  $\theta(A_1, A_2 \dots A_i \dots A_n, B_1, B_2 \dots B_i \dots B_n)$  de [19] recibe el nombre de *postulado de significado* y puede ser de dos tipos:

- a) Postulado de significado léxico,
- b) Postulado de significado lógico.

Intentaré dar forma a todo este material. Para empezar, he de confesar que he retomado la expresión *Postulado de significado*, recuperando la tradición carpiana (Carnap, 1952). Si vamos a admitir la expresión de relaciones de cualquier complejidad entre el léxico disponible en la premisa y la conclusión, me parece que lo más conveniente es huir de la combinatoria, en ocasiones algo limitante, que muestran tanto el modelo holandés como el americano. Estamos interesados en algo más que la monotonía o la relación obtenida al comparar las extensiones de dos clases dadas. De hecho, un postulado de significado podría ser cualquier expresión en un lenguaje formal de complejidad igual o mayor a aquella en que se expresan premisas y conclusión. Y no solo eso, sino que se contempla, igualmente, la existencia de postulados que afectan a la forma lógica y, en consecuencia, no guardan una relación directa con el léxico involucrado en la inferencia.

Antes de entrar en detalles, me gustaría decir algo más acerca de la forma en que se concibe aquí un *postulado de significado*. La cláusula [20, iv] deja bastante claro que su función es garantizar que la conclusión es consecuencia de la premisa aunque, en este momento, no indiquemos qué mecanismo

inferencial es el que opera para establecer este hecho. En efecto, ni siquiera es necesario que exista un procedimiento formal, de los considerados estándar, para probar la inferencia indicada en [20, iv].

Por otra parte, pedirle a un hablante que dé cuenta en tiempo real de las razones por las que la premisa, junto con el conocimiento aportado por el postulado de significado, arrojan como resultado la conclusión, no es, ni puede ser, un requisito exigible en Lógica Natural. Las inferencias naturales son actos situados en el tiempo que vienen limitados por las capacidades de procesamiento de los sujetos que las practican. La persona que lleva a cabo una de estas inferencias solo tiene que ser capaz de elegir adecuadamente el postulado de significado pertinente al caso y efectuar las adaptaciones que la premisa requiera. Cuáles sean las razones por las que el sujeto en cuestión confía en dicho postulado es algo que puede ser capaz de aclarar eventualmente, y en caso de hacerlo, el nivel de detalle puede ser asimismo variable. Puede haberlo aprendido, sin necesidad de reflexión, a través de la experiencia propia y, con más frecuencia, a partir de la ajena. Pueden ser contenidos memorizados con el tiempo, de los cuales puede dar razón en mayor o menor medida, pero sobre cuya justificación quizá nunca se haya parado a pensar.

Podría decirse, entonces, que esta forma de entender la inferencia natural la acerca peligrosamente al terreno de los razonamientos irreflexivos e incluso falaces. Y podría ser el caso si no existiera un respaldo objetivo y comunitario a la pertinencia de los postulados de significado que los hablantes emplean en cada caso. Lo que pretendo dejar claro es que, en una inferencia natural, no se le puede exigir al hablante que aporte de manera inmediata evidencia de la forma en la que el postulado de significado actúa como sostén de la inferencia, pero esto no significa que dicho sostén no exista. Las inferencias naturales son buenas inferencias porque existe una justificación que la comunidad de los hablantes asume a partir de una evidencia científica bien asentada y transmitida a través de los usos y costumbres de una comunidad de hablantes. No pertenecen, por lo tanto, al dominio de las falacias, ni tampoco al de la Teoría de la Argumentación, aunque se puedan apreciar similitudes.

Las inferencias naturales son buenas inferencias desde el punto de vista lógico, no meras inferencias plausibles, como sucede en Teoría de la Argumentación. Son inferencias que preservan la verdad en un sentido clásico, es decir, en las que la verdad de la premisa y el postulado de significado garantizan *necesariamente*<sup>14</sup> la verdad de la conclusión. Es precisamente esta facultad de nuestra cognición la que este enfoque de la Lógica Natural pretende estudiar, y no más bien el vasto dominio de los argumentos aceptables en el discurso ordinario. Aprovecharé este momento para ofrecer la definición de Lógica Natural

---

14 Cuando hablo aquí de necesidad no pretendo conectar con ninguna versión modal de la consecuencia, sino reflejar la forma en que normalmente calificamos el hecho de que no exista una interpretación que haga verdaderos la premisa y el postulado y falsa la conclusión.

a la que me llevan todas estas consideraciones y que, como se podrá ver de inmediato, difiere sustancialmente de la ofrecida en el punto [2].

22.

La *Lógica Natural* es la disciplina que se ocupa del estudio de los postulados de significado, sus formas, tipologías y modos de aplicación dentro de las inferencias situadas que tienen lugar en el discurso ordinario.

Como se puede ver, no es el apego a las herramientas de análisis del lenguaje ordinario, Gramática Categorial, etc., lo que marca la diferencia, sino el papel de los postulados de significado a partir de un modelo de argumento como el que se presenta en [19] y [20]. Lo que queda de esta sección lo dedicaré, precisamente, a presentar algunos de los objetivos de este tipo de estudio.

En [21] he hablado de *postulados de significado léxico* y de *postulado de significado lógico* como una forma de ampliar el alcance que los primeros desempeñan en las inferencias naturales. Empezaré por lo más conocido tratando los postulados léxicos en primer lugar.

#### Postulados de significado léxico

Los postulados de significado léxico, o simplemente, postulados léxicos expresan relaciones entre los términos del léxico en el que se exprese la premisa. Son, básicamente, todos los que se tratan en profundidad tanto en el modelo holandés como en el americano, con la salvedad de que tenemos a nuestra disposición muchos más recursos para describir relaciones entre términos léxicos, al menos todas las de FOL.

El modelo de argumento se puede restringir de la siguiente manera:

23.

$$\frac{\varphi(A_1, A_2 \dots A_i \dots A_n)}{\varphi(B_1, B_2 \dots B_i \dots B_n)} \quad \theta(A_1, A_2 \dots A_i \dots A_n, B_1, B_2 \dots B_i \dots B_n)$$

Obsérvese que en este caso la premisa y conclusión son ambas expresiones del tipo  $\varphi(\Theta_1, \Theta_2, \dots, \Theta_i, \dots, \Theta_n)$ , dándose a entender de este modo que tienen la misma estructura y que lo único que varía es el léxico.

¿Cómo se trataría la monotonía/antitonía bajo esta concepción? Es decir, ¿cuáles son los postulados léxicos que expresan monotonía y antitonía? La respuesta es bastante evidente como se aprecia a continuación:

24.

Monotonía:

$$\frac{\forall x(Ax \rightarrow Bx)}{\forall x(Ax \rightarrow Cx)} \quad \forall x(Bx \rightarrow Cx)$$

El término léxico  $Bx$  en la premisa tiene polaridad positiva  $-(\uparrow)-$ , el postulado de significado garantiza que  $\forall x(Bx \rightarrow Cx)$ , lo que en términos de la extensión de  $Bx$  y  $Cx$  se puede expresar como  $[B] \subseteq [C]$ , y por tanto, debemos concluir que  $\forall x(Ax \rightarrow Cx)$ .

Observarse que lo único que hemos hecho aquí es dar cuenta de que la inferencia siguiente es lógicamente válida

25.

$$\forall x(Ax \rightarrow Bx), \forall x(Bx \rightarrow Cx) \vdash \forall x(Ax \rightarrow Cx)$$

tal y como se exige en [20, iv].

El caso de la antitonía es igualmente trivial y no lo discutiré en detalle.:

26.

Antitonía:

$$\frac{\forall x(Ax \rightarrow Bx)}{\forall x(Cx \rightarrow Bx)} \quad \forall x(Cx \rightarrow Ax)$$

Nótese, en este caso, que el criterio [20, iv], junto con la identificación de las expresiones que, en FOL, se asocian con relaciones de inclusión entre términos léxicos, permiten emplear este procedimiento para identificar la polaridad de los términos léxicos en una expresión de cualquier complejidad. Supóngase dada una fórmula  $\varphi(A_1, A_2 \dots A_i \dots A_n)$  en FOL. Deseamos ahora saber cuál es la polaridad del término  $A_i$  supuesto que tiene una única ocurrencia en  $\varphi(A_1, A_2 \dots A_i \dots A_n)$ . Tenemos entonces la siguiente solución:

26.

$$\frac{\varphi(A_1, A_2 \dots A_i \dots A_n)}{\varphi(A_1, A_2 \dots A'_i \dots A_n)} \quad \forall x(A_i(x_0 \dots x_n) \rightarrow A'_i(x_0 \dots x_n))$$

Es decir, si reemplazamos en  $\varphi(A_1, A_2 \dots A_i \dots A_n)$ ,  $A_i$  por  $A'_i$  y el argumento al que se añade como premisa el postulado léxico  $\forall x(A_i(x_0 \dots x_n)$

$A'(x_0 \dots x_n))$  es clásicamente correcto, entonces  $A_i$  tiene polaridad positiva ( $\uparrow$ ) en  $\varphi(A_1, A_2 \dots A_i \dots A_n)$ . Para analizar la antitonia se hace lo propio, solo que cambiando el postulado por  $\forall x(A'_i(x_0 \dots x_n) \rightarrow A_i(x_0 \dots x_n))$ . Evidentemente, este procedimiento tiene sus límites, ya que si aceptamos lenguajes de potencia expresiva mayor que FOL no podemos garantizar la existencia de mecanismos de prueba aceptables que actúen como los algoritmos que aquí se precisan. Algo que tampoco se puede afirmar si abandonamos el fragmento monádico de FOL. Pero, tal vez, no deberíamos detenernos en exceso en este punto.

¿Son todos los postulados de significado expresiones cuantificacionales del tipo  $\forall x(A_i(x_0 \dots x_n) \rightarrow A'_i(x_0 \dots x_n))$ ? Es evidente que no. Si así fuera, no tendríamos manera de capturar las intuiciones, a mi juicio válidas, sobre la forma en que dos términos léxicos, en este caso relaciones 1-arias, pueden relacionarse. Veámoslo con un ejemplo muy básico que tiene que ver con la relación de *alternancia* entre términos léxicos introducida en la Tabla 1.

27.

I. Si lleva chip, entonces tiene que ser una mascota.

**II. Si es una mascota no es un animal salvaje,**

Por tanto,

III. Si lleva chip, entonces no es un animal salvaje.

Que se formalizaría en FOL como:

28.

$$\frac{\forall x(Chx \rightarrow Mx)}{\forall x(Chx \rightarrow \neg Sx)} \quad \forall x(Mx \rightarrow \neg Sx)$$

Siempre cabe aducir que estamos ante una instancia camuflada de monotonía en el que el postulado  $\forall x(Mx \rightarrow \neg Sx)$  solo expresa que  $[A] \subseteq [S]$ . No creo, sin embargo, que esta crítica resulte efectiva. Pero, para ello, es cierto que tendríamos que encontrar postulados léxicos que no se expresaran en términos de una expresión cuantificada universalmente. Por fortuna existen, y podría ser que tuvieran que ver con la formulación de excepciones. Veámoslo con un ejemplo:

29.

Todos los comunistas son ateos.

***Sin embargo, es bien sabido que hay Padres jesuitas que son comunistas,***

Por tanto,

Hay Padres jesuitas que son ateos.

Su formalización sería entonces la siguiente:

30.

$$\frac{\forall x(Cx \rightarrow Ax)}{\exists x(Jx \rightarrow Ax)} \quad \exists x(Jx \& Cx)$$

Resulta curioso que si obviamos la triple distinción *premisa-postulado léxico-conclusión* y la reemplazamos por la clásica de *premisa mayor-premisa menor-conclusión*, lo que obtenemos es un silogismo del modo *DARII* perteneciente a la primera figura (Ferrater Mora, 1964, p. 688 y ss.). Como veremos en breve, esto no significa que el desarrollo que aquí estamos exponiendo se reduzca en definitiva a la silogística aristotélica. No lo hace ni en sus fines, ni tampoco en sus herramientas. Pero antes de ir a este punto, sí me gustaría reparar en un detalle interesante.

Según la cláusula [20, iv], lo que tiene que ocurrir para aceptar que, dado un determinado postulado de significado, una premisa arroje una conclusión, es lo siguiente:

31. *Premisa, Postulado de significado*  $\vdash$  *Conclusión*,

admitiendo que existe un mecanismo capaz de evaluar la relación de consecuencia en dicho contexto que sea al menos consistente.<sup>15</sup>

Si la relación de consecuencia  $\vdash$  es suficientemente clásica, las premisas, dadas en esta ocasión por la *premisa* y el *postulado de significado*, pueden alterar su posición sin que el resultado se vea afectado: la conclusión se seguirá igualmente. Sin embargo, desde el punto de vista de una inferencia natural, esto da lugar a un argumento completamente distinto en el que la premisa y el postulado de significado han alternado sus posiciones y en consecuencia su función. Si realizamos dicha operación sobre el argumento descrito en [30] se obtiene lo siguiente:

32.

$$\frac{\exists x(Jx \& Cx)}{\exists x(Jx \rightarrow Ax)} \quad \forall x(Cx \rightarrow Ax)$$

que ahora representaría un silogismo de la 3ª Figura en el modo DISAMIS.

¿Qué ha cambiado? Desde luego no la validez. Lo que ha cambiado es qué pieza de información actúa como *información fáctica* –la premisa– y cuál como *garantía*, o tal vez mejor, como *disparador* de la inferencia. Y no creo, desde luego, que pese a preservarse la validez, [30] y [32] representen la misma inferencia natural.<sup>16</sup>

Esta diferencia remite, más bien, al estatus epistemológico que tienen para el hablante ciertos enunciados. La información fáctica, típicamente expresada en la premisa, parece representar información acerca del mundo y, por lo tanto, contingente y potencialmente cambiante. Sin embargo, los postulados de significado, del tipo que sean, representan información distribuida e interiorizada por la comunidad de hablantes en mayor o menor medida, algo que podría caer dentro del *conocimiento común* tanto como de la *experiencia personal*. Es prematuro pronunciarse sobre la forma lógica privilegiada por los postulados de significado, y en particular, por los postulados léxicos. No obstante, también es justo reconocer que los enunciados de tipo universal pueden ser en gran medida más frecuentes que los

<sup>15</sup> Apelo aquí a la consistencia y no a la corrección porque la primera puede expresarse al menos como una propiedad puramente sintáctica, mientras que la segunda depende de una interpretación semántica de los elementos de un lenguaje dado. Dudo que en estos contextos sea posible definir una clase de interpretaciones como las que son típicas en los desarrollos estándar de la Lógica Formal.

<sup>16</sup> Del mismo modo que no representan la misma forma de Silogismo válido.

existenciales. Es decir, desde un punto de vista cognitivo, es más importante afianzar la norma en lugar de la excepción, si es que los postulados de tipo existencial representan realmente las excepciones.

Un tipo de postulados de significado que no parecen haber recibido demasiada atención en la literatura son aquellos que afectan a términos léxicos de ariedad mayor que uno, es decir, lo que típicamente venimos denominado como *relaciones*. En particular aquellos que no tienen que ver directamente con el par *monotonía/antitonía*. Veamos un ejemplo muy elemental:

33.

- I. Ana está sentada a la derecha de Juan.
- II. Si alguien está a la derecha de otra persona, esta se encuentra a su izquierda,
- III. Por tanto,
- IV. Juan se encuentra sentado a la izquierda de Ana.

Su forma en FOL podría expresarse del siguiente modo:

$$\frac{Daj}{Ija} \quad \forall x \forall y (Dxy \rightarrow Iyx)$$

En este caso da la impresión de que ni siquiera habría necesidad alguna de hacer explícito el postulado de significado, ya que forma parte del acervo más elemental de cualquier hablante. No parece posible ser un hablante competente del castellano sin conocer algo tan elemental como la relación existente entre este tipo de términos *posicionales*. ¿Quiere esto decir que estamos ante postulados de significado de una tipología distinta? Admito que la cuestión no parece tan fácil de dirimir como pudiera parecer. Es cierto que en el resto de los casos que hemos empleado como ejemplos, cabe ampliamente la posibilidad de que otros agentes no compartan la verdad del postulado de significado que actúa como disparador de la inferencia. Podemos ignorar si hay, por ejemplo, mascotas que sí se puedan considerar como animales salvajes en algún sentido del término, o podemos dudar de si realmente hay jesuitas que sean comunistas, pero no parece posible poner en cuestión las relaciones mutuas de términos como *izquierda* o *derecha* sin manifestar graves carencias de algún tipo. Creo que estas consideraciones solo apuntan, al menos de momento, a diferencias epistemológicas que afectan a la adquisición individual y colectiva del significado de ciertos términos. Conocido el significado de relaciones como *estar a la derecha* –*Dxy*– y *estar a la izquierda* –*Ixy*– parece necesario aceptar la validez de expresiones como  $\forall x \forall y (Dxy \rightarrow Iyx)$  lo que quizá nos podría llevar a considerar que las relaciones *Dxy* e *Ixy* actúan como operadores lógicos de un tipo

especial, como sucede, de hecho, con la *igualdad*. Sin descartar esta opción, creo que, por el momento, es preferible seguir considerando al postulado anterior como un postulado semántico y ello pese a que podamos ver en estos operadores de posición algo próximo a una constante lógica.

Otro terreno cuajado de complicaciones es el que afecta a las inferencias naturales procedentes del nivel sentencial. Es normal que tanto el modelo holandés como el americano presten nula atención a este ámbito de nuestra intuición formal. Si partimos de que el léxico de las expresiones propias de la Lógica sentencial son, precisamente, las variables de enunciado, una inferencia basada en un postulado léxico debería ser aquella en la que la conclusión preserva la estructura de la premisa procediendo tan solo al reemplazo de alguna variable sentencial. Lo analizaremos con un ejemplo relativamente trivial.

34.

$$\frac{p \rightarrow q}{p \rightarrow r} \quad q \rightarrow r$$

Cuesta trabajo no ver en esta inferencia un caso de monotonía aplicado a la segunda variable sentencial de la expresión condicional  $p \rightarrow q$ , haciendo entonces de  $q \rightarrow r$  un caso típico de postulado de significado léxico. Esta intuición se haría extensiva a la antitonía dando lugar a las siguientes inferencias naturales:

35.

$$\frac{p \rightarrow q}{p \rightarrow \vartheta} \quad q \rightarrow \vartheta$$

$$\frac{p \rightarrow q}{\vartheta \rightarrow q} \quad \vartheta \rightarrow p$$

Lo que se está retratando en [35] no es sino una versión extendida de la regla de sustitución de equivalentes. Se podría admitir que este tipo de transformaciones tienen que ver, ciertamente, con los postulados de significado léxico que hemos venido estudiando aquí. No obstante, en este momento no me atrevo a zanjar la cuestión de manera taxativa. Las inferencias naturales que afectan al nivel sentencial no parecen presentar el mismo manejo del léxico que aquel que se presenta en primer orden, pero no entraré en ello ahora.

## Postulados de significado lógico

Como ya lo he mencionado, la idea de actuar sobre el léxico de una expresión conservando su estructura original, representa una forma extraordinariamente económica de obtener consecuencias necesarias a partir de una expresión inicial dada. Pero no es la única forma que tenemos de emplear nuestras intuiciones formales en el contexto de la inferencia natural. Disponemos igualmente de habilidades cognitivas que nos permiten actuar sobre la estructura de un enunciado para obtener las consecuencias necesarias del mismo. Consecuencias que transforman la estructura original conservando el léxico total o parcialmente.

Entre este tipo de operaciones destacan las que, a menudo, se han denominado como *reglas de interdefinición*. Veamos algunos ejemplos:

36.

$$\frac{\forall x Ax}{\neg \exists x \neg Ax}$$

$$\frac{p \& q}{\neg(\neg p \vee \neg q)}$$

$$\frac{\Box p}{\neg \Diamond \neg p}$$

Se trata solo de tres casos bien conocidos de inferencias cuya función es la de explicar el significado de una constante lógica apelando a otras. Para ello, resulta oportuno no actuar sobre el léxico al mismo tiempo.

Así visto, resulta difícil identificar cuál es el paralelismo con los esquemas inferenciales promovidos por postulados de significado léxico. En un esquema como el representado, por ejemplo en [31], el postulado de significado  $\forall x(Cx \rightarrow Ax)$  que actúa como garante, o disparador, de la inferencia  $\exists x(Jx \& Cx) \vdash \exists x(Jx \rightarrow Ax)$  parece recoger cierto conocimiento acerca de la extensión de los términos léxicos  $Cx$  y  $Ax$ . En particular:  $[C] \subseteq [A]$ . ¿Qué conocimiento actúa como garante o disparador, en el caso de las inferencias presentadas en [36]? Aparentemente solo puede ser el conocimiento de las reglas e interpretaciones admisibles que dan cuenta del significado de las constantes lógicas afectadas. Es decir, no parece preciso apelar a conocimiento factual de ningún tipo.

Este análisis es correcto, pero solo en cierta medida, ya que la mayoría de los hablantes del español son perfectamente capaces de aplicar tales reglas sin tener un conocimiento, siquiera aproximado, de los sistemas formales en que se tratan las constantes lógicas afectadas y que según lo anterior, son

las que podrían dar cuenta de la corrección de estas inferencias.

La forma en la que los hablantes de una lengua aprenden este tipo de transacciones no es sustancialmente distinta del modo en que aprenden que la extensión de un término léxico está incluida en la de otro dado. Los postulados léxicos aportan conocimiento fáctico sobre el significado de ciertas expresiones, mientras que los postulados lógicos harían lo propio con las constantes lógicas. Y es en este punto donde las diferencias entre léxico y constantes se difuminan. Es cierto que es por completo contingente que los atenienses sean griegos,  $\forall x(Ax \rightarrow Gx)$ , mientras que, cualquiera de las inferencias listadas en [35] dependen de un conocimiento que afecta directamente a nuestra cognición. Pero no dependen en menor medida del conocimiento que los hablantes tienen del significado de ciertas partículas de su lengua y de sus relaciones mutuas.

¿Cómo expresaríamos las inferencias enumeradas en [36] en términos del formato básico de una inferencia natural? Una opción bastante razonable sería la siguiente:

37.

$$\frac{\forall x Ax}{\neg \exists x \neg Ax} \quad \forall \Theta \forall x (\forall x \Theta x \rightarrow \neg \exists x \neg \Theta x)$$

Al emplear una notación puramente formal, es completamente comprensible que, a continuación, esperemos una explicación en términos de los formalismos comprometidos, en este caso FOL y SOL (*Second Order Logic*). Sin embargo, aquí solo pretendemos resumir una explicación que sin ningún problema podríamos ofrecer en el lenguaje natural. La fórmula de 2° Orden  $\forall x(\forall x \Theta x \rightarrow \neg \exists \neg x \Theta x)$  solo resume apropiadamente un principio como el siguiente: *Si una afirmación se predica de todos, entonces no puede haber algún individuo del que no se predique esa misma afirmación.* No hace falta un gran conocimiento de la Lógica de Primer Orden ni de la de Segundo Orden para encontrar un consenso suficiente acerca de la validez de este principio.

Y lo mismo sucede en el resto de los casos listados en [36]. Veámoslo:

38.

$$\frac{p \& q}{\neg(\neg p \vee \neg q)} \quad \forall \alpha \forall \beta (\alpha \& \beta \rightarrow \neg(\neg \alpha \vee \neg \beta))$$

$$\frac{\Box p}{\neg \Diamond \neg p} \quad \forall \alpha (\Box \alpha \rightarrow \neg \Diamond \neg \alpha)$$

El postulado de significado que apoya la primera de estas inferencias responde a un principio informal que se podría expresar del siguiente modo: *si es cierto que dos enunciados cualesquiera son verdaderos a la vez, entonces no puede ocurrir que alguno de ellos sea falso*. Y sucede, de igual modo, para el caso modal analizado, cuya traducción informal sería: *Si algo se da necesariamente, entonces no es posible que sea falso*.

Pero los postulados de significado lógico no se agotan en las interdefiniciones. Hay muchos más frentes en los que resultan reconocibles. Iré por partes:

39.

$$\frac{\forall x Ax}{\exists x Ax} \quad \forall \Theta \forall x (\forall x \Theta x \rightarrow \exists x \Theta x)$$

$$\frac{\Box p}{\Diamond p} \quad \forall \alpha (\Box \alpha \rightarrow \Diamond \alpha)$$

Los postulados de significado empleados en [39] expresan formalmente los siguientes principios intuitivos: *lo que se dice de todos, se dice de alguno y lo que se da necesariamente es igualmente posible*. Curiosamente ninguno de estos dos postulados son universalmente aceptados en el ámbito de la Lógica Formal. El primero de ellos porque está asociado a la condición de que no existan modelos cuyo dominio de cuantificación sea el conjunto vacío, y, el segundo, por razones semejantes: no puede haber mundos posibles incapaces de acceder a otro composable con ellos.<sup>17</sup>

¿Cuántos más de estos postulados de significado se pueden considerar? Esta pregunta esconde lo que, posiblemente, sea una crítica bastante consistente a los postulados de significado lógico. Porque en realidad, cualquier inferencia formal que pueda ser demostrada en algún sistema formal<sup>18</sup> aceptado por la comunidad científica, puede ser reinterpretada como un

<sup>17</sup> De hecho, la fórmula  $\Box \alpha \rightarrow \Diamond \alpha$  se tiene que añadir como axioma, el axioma D, en la interpretación deóntica de los operadores modales.

<sup>18</sup> Y deberíamos añadir, que satisfaga el Teorema de Deducción.

postulado de significado formal. En ese sentido, se podría decir que los postulados lógicos son una forma de reintroducir la Lógica Formal, con todo su peso, en el territorio de la Lógica Natural. Y ello solo con una especie de ejercicio malabar consistente en expresar una inferencia válida como un principio en el que se cuantifica universalmente sobre las variables léxicas existentes.

Aquí hay varias consideraciones que hacer. La primera tiene que ver con la complejidad de los postulados a incorporar. Por definición, no parece aceptable que podamos manejar postulados de una complejidad excesiva, simplemente porque no los podríamos memorizar ni aplicar de manera efectiva. Por otra parte, debe tenerse en cuenta que una inferencia natural no incorpora en su aplicación las razones por las que un determinado postulado es o no aceptable. Estas habrán sido aportadas y discutidas en su momento e incorporadas en el acervo de una cierta comunidad de hablantes. No se precisa de una especie de Lógica universal que sea capaz de dar cuenta, a la vez, de todos los postulados que podemos tratar de este modo. De hecho sabemos que tal cosa no existe. Los postulados lógicos pueden encontrar respaldo en sistemas formales, cierto, pero también pueden basarse en razonamientos informales, a falta de algo mejor. De hecho, algunos de estos postulados pueden resultar contradictorios entre sí o simplemente inaceptables, tanto para algunos hablantes como para ciertos contextos. Por ejemplo, si empleo como postulado de significado  $\forall\alpha\forall\beta(\alpha\&\neg\alpha \rightarrow \beta)$  para concluir  $q$  a partir de una premisa  $p\&\neg p$ , estaré corriendo el riesgo de que algún partidario de las Lógicas paraconsistentes o las relevantistas ponga en cuestión mi razonamiento. De hecho, solo en contextos muy específicos se acepta un postulado tan radical como el anterior, por mucho que represente el principio sacrosanto de *ex contradictione quodlibet*.

En definitiva, no hace falta una Lógica universal,<sup>19</sup> ni tan siquiera un único sistema formal. Podemos recurrir a una diversidad de ellos, incluso contradictorios entre sí, o en ausencia de alguno apropiado, a razonamientos informales que aporten razones a favor del postulado empleado en cada caso.

De hecho, el único rasgo de logicidad compartido por cualquier inferencia natural basada en un postulado lógico es la aplicación del *modus ponens* –MP– en los siguientes términos:

40.

*Premisa, Premisa    Conclusión*  $\vdash$  *Conclusión*

¿Ofrece alguna ventaja que nuestros postulados de significado se justifiquen por medio de un sistema formal al uso? Sin duda. Es algo que puede ayudar, y mucho, a analizar inferencias naturales que no resulten tan cla-

<sup>19</sup> No hago referencia aquí a los contenidos propios de los Congresos *Universal Logic*, sino a un sistema capaz de dar cuenta de todas las inferencias en las que intervienen partículas portadoras de logicidad.

ras a primera vista. Todo dependerá entonces de la potencia expresiva del lenguaje en el que surja la disputa. Si está limitada a un rango de recursos para los que existe algún cálculo, o aún mejor, algún algoritmo decidible de prueba, la disputa siempre será soluble mediante el cálculo. En caso contrario, deberemos conformarnos con lo que podamos conseguir en cada caso.

Creo, y con esto concluyo, que la identificación de los postulados de significado lógico y su función en la inferencia natural tiene un componente empírico (Geurts, 2005, secc. 6.2) que no puede ser desestimado. Me gustaría pensar que existe un hilo conductor capaz de jerarquizar su uso en función de variables como el tiempo de ejecución, universalidad, tipología de los operadores lógicos involucrados, etc. Pero dudo de que estemos siquiera cerca de ese logro.

### **La Lógica Natural en las Humanidades**

La cuestión que sobrevuela en este trabajo no es otra que esa. ¿Debemos considerar a la Lógica Natural como un saludable reemplazo de la Lógica Formal en los estudios de Filosofía? La cuestión del lugar de la Lógica en las Humanidades, no es nueva, ni tiene una fácil respuesta –véase, por ejemplo (Alonso, 2019b)– pero es, en cierta medida, la auténtica cuestión de la Lógica.

Años atrás, la Lógica Formal inició un paulatino retroceso frente a la Teoría de la Argumentación en los planes de estudio de Humanidades (Alonso, 2019a). Unas veces tuvo lugar como un genuino reemplazo de una asignatura por otra en los planes de estudio y otras, más frecuentemente, como un cambio subrepticio de los contenidos propios de la Lógica Formal ¿Se va a repetir este proceso con la Lógica Natural? Creo que estamos muy lejos de ello.

Es cierto que la Lógica Natural presenta contenidos que, en buena ley, resultan más próximos a las preocupaciones e intereses del humanista que aquellos que en la actualidad se estudian en algunos cursos de Lógica Formal. Esto, que parece una virtud, podría ser también un riesgo si se aprovecha para alimentar una reintroducción de la vieja Silogística escolástica en los planes de estudio. Muchos estarían conformes con este anacronismo, pero este no es mi caso.

La Lógica Natural requiere, para ser justamente comprendida, conocimientos no elementales de Gramática Formal. A eso se debe sumar cierto dominio de ramas muy técnicas de la Lógica Formal: Teoría de Tipos, Cálculo de Tablas Analíticas y Cálculo lambda, entre otros. Es posible, no lo descarto en absoluto, que en un tiempo razonable seamos capaces de reducir la carga técnica al mínimo, al tiempo que vayan componiéndose manuales útiles para exponer sus contenidos de forma sistemática. Al fin y al cabo, esa es la intrahistoria de la Lógica Formal en los estudios de Humanidades. Por el momento, habrá que esperar a que dichos manuales estén disponibles y a que la Academia asimile la necesidad o la conveniencia de revisar los contenidos de sus objetivos formativos.

Con esto no quiero, ni por asomo, promover la desaparición de la Lógica Formal de los planes de estudio de las Humanidades. Sería fácil sucumbir a la retahíla de discursos que ven en esta disciplina una suerte de Matemáticas encubiertas bien alejadas del genuino espíritu de la Filosofía. Pero es inmediato observar que esos mismos argumentos servirían para cerrar el paso a la Lógica Natural y, en general, a cualesquiera de los estudios que traigan al presente los métodos y enseñanzas de la Filosofía. Solo una visión muy específica y pobre de esta disciplina podría reivindicar tal cosa. Yo solo pido una reflexión para analizar cuánta Lógica y de qué tipo es aquella que debe saber un filósofo, no si debe conocer alguna.

Por otra parte, aún está pendiente de aclarar qué derechos arropan la petición de una presencia para la Lógica Natural en la Filosofía. Un argumento, y no menor, podría ser el de su peso en las infraestructuras básicas de la IA-gen. Ya hemos hablado de ello al comentar algunos de los enfoques de la escuela americana y su orientación eminentemente computacional. Que la Filosofía ignore los principios básicos de una disciplina con un potencial transformador como lo es la IA-gen, me parece de una irresponsabilidad que muy bien haríamos en evitar, aunque también pueda reconocer mi escepticismo al respecto. Pero no es esa la razón por la que defiendo que se estudie el modo en que la Lógica Natural podría formar parte de la formación del filósofo. La pregunta que no puedo, no podemos, dejar de hacernos es ¿por qué somos capaces de razonar lógicamente? Para entender la dimensión de esta pregunta fundamental tengo que pedirle al lector que imagine la posibilidad de que no lo hiciéramos, o que lo hiciéramos de forma sustancialmente distinta a como, de hecho, ocurre. No es fácil, pero lo poco que la paleoantropología nos dice de la condición humana es que ésta se ha expresado en un número, además creciente, de especies de las cuales solo la nuestra ha sobrevivido en el tiempo ¿Tendrían una Lógica, siquiera rudimentaria? ¿sería la nuestra? La etología también nos regala, con el tiempo, estudios en los que distintas especies, a veces muy distintas en realidad, presentan formas de logicidad nada desdeñables ¿Cuál es, comparativamente, nuestra *firma* característica? ¿Cuál es la estructura de esta habilidad cognitiva que nos permite inferir nuevas<sup>20</sup> piezas de información a partir de otras dadas?

Para un materialista militante, como es mi caso, la existencia de la habilidad formal en nuestra cognición solo puede explicarse como parte del éxito evolutivo de nuestra especie. En consecuencia, y en la medida en que la presión selectiva siga presente entre nosotros, estaría inmersa en un constante proceso de cambio. Entender la estructura de esa parte de nuestra cognición, y hacerlo tanto desde el plano conceptual como desde el empírico, es una aventura a la que no me gustaría renunciar como humanista. Es cierto que estaríamos abandonando una multitud de supuestos que afectan a los

---

20 No quiero decir nuevas en el sentido de que amplíen la información de la premisa: eso nos llevaría fuera de la Lógica.

estudios de Lógica situándola en un territorio intermedio entre las matemáticas, la antropología, la psicología y la teoría de la mente, pero, ¿qué perdemos dando una oportunidad a esta opción? En nuestra tradición reciente la hibridación de disciplinas ha dado, a menudo, los mejores resultados, no creo que debamos negarnos a intentarlo, una vez más, en este caso.

### Referencias bibliográficas

- Alonso, E., (2019a). “Lógica y teoría de la argumentación. Anatomía de una reforma”. *Quadripartita ratio*, (7), pp. 12-31.
- (2019b). “¿Qué lógica necesita saber un filósofo?” *Andamios, Revista de Investigación Social*, 16(41), pp. 143-163.
- Belnap, N. D. (1962). “Tonk, plonk and plink”. *Analysis*, 22(6), pp. 130-134.
- Carnap, R. (1952). “Meaning postulates”. *Philosophical Studies*, 3(5), pp. 65-73.
- Ferrater Mora, J. (1964). *Diccionario de filosofía de bolsillo*. Alianza.
- Frege, G. (1972). *Conceptografía*. Instituto de Investigaciones Filosóficas. Colección: Filosofía Contemporánea.
- Geurts, B. (2005). “Monotonicity and processing load”. *Journal of Semantics*, 22(1), pp. 97-117.
- MacCartney, B. and Manning, C. (2009). “An extended model of natural logic”. *Proceedings of the 8th International Conference on Computational Semantics*, pp. 140-156.
- Moss, L. S. (2010). “Natural Logic and Semantics”. En Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J. M., Mattern, F., Mitchell, J. C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B., Sudan, M., Terzopoulos, D., Tygar, D., Vardi, M. Y., Weikum, G., Aloni, M., Bastiaanse, H., De Jager, T., and Schulz, K., editors, *Logic, Language and Meaning*, volume 6042, pp. 84-93. Springer Berlin Heidelberg.
- Sánchez-Valencia, V. (1991). *Studies on Natural Logic and Categorical Grammar*. Universiteit van Amsterdam.
- Van Benthem, J. (1990). “Categorical Grammar and Type Theory”. *Journal of Philosophical Logic*, 19(2), pp. 115-168. <http://www.jstor.org/stable/30226424>.
- Van Benthem, J. (1991). “Logic in action”. *Journal of Philosophical Logic*, 20(3), pp. 225-263.
- Van Benthem, J. (2008). *A brief history of natural logic. Technical report, ILLC*. Amsterdam and Stanford.

# Un nuevo paradigma para la formación de traductores en la era de la IA generativa

Silvana Debonis

## Introducción

En las últimas décadas, el avance vertiginoso de la inteligencia artificial generativa (IAG) ha dado lugar a una reconfiguración sustancial en múltiples sectores, y la traducción no ha sido ajena a este fenómeno. Los avances en los modelos de lenguaje de gran escala han redefinido la capacidad de las máquinas para traducir textos y generar contenido con una fluidez y coherencia comparables a las de un traductor humano experimentado. Sin embargo, esta evolución plantea interrogantes esenciales sobre la formación de traductores en un contexto donde la tecnología ha dejado de ser un mero asistente para convertirse en un productor autónomo de contenido.

En el ámbito de la traducción, la irrupción de sistemas basados en IAG, abre nuevas oportunidades y plantea desafíos significativos. Por un lado, las herramientas de traducción automática han evolucionado hasta lograr estructuras sintácticas complejas, algo que los enfoques estadísticos y basados en reglas no podían alcanzar. Por otro, el impacto de estos avances suscita interrogantes éticos y profesionales de gran envergadura: ¿Qué competencias deben desarrollar los traductores en formación para afrontar un entorno laboral en el que las máquinas ya no se limitan a asistir, sino que también generan textos de manera autónoma? ¿Cómo deben reconfigurarse los programas educativos para incorporar habilidades en posesición, supervisión estratégica y gestión crítica del contenido automatizado?

Este capítulo examina la transformación del perfil profesional del traductor en el contexto de la inteligencia artificial generativa, con especial atención a los fundamentos teóricos de esta tecnología y a sus implicaciones en la formación académica. A partir de este análisis, se presenta una propuesta didáctica orientada al desarrollo de un traductor híbrido, con competencias para interactuar con sistemas basados en IA, corregir textos generados automáticamente y evaluar su calidad, coherencia y adecuación contextual con criterio profesional. Se sostiene que esta orientación formativa resulta indispensable para preparar traductores capaces de asumir un rol estratégico en los procesos de producción textual automatizada, más allá del manejo instrumental de las herramientas tecnológicas.

La importancia de este análisis reside en que la incorporación de la inteligencia artificial generativa a los flujos de trabajo traductológicos ha transformado la naturaleza misma de las tareas involucradas en la traducción. Lejos de limitarse a un cambio instrumental, esta integración ha configurado un escenario híbrido donde las competencias lingüísticas tradicionales deben articularse con saberes técnicos y capacidades estratégicas. En este

nuevo marco, el traductor deja de ocupar exclusivamente el lugar de mediador lingüístico para asumir funciones más amplias como supervisor, analista y gestor de calidad textual. Esta evolución del rol profesional plantea desafíos prácticos y teóricos que afectan tanto a las tareas concretas del traductor como a los marcos desde los cuales se concibe su intervención. En particular, obliga a revisar nociones como fidelidad textual y adecuación cultural en entornos multilingües mediados por IA, donde la figura del autor, el traductor y el modelo se superponen y reconfiguran en nuevas formas de producción textual.

Desde una perspectiva teórica, este artículo se apoya en los aportes del procesamiento del lenguaje natural (PLN), el aprendizaje profundo (*deep learning*) y los estudios sobre posesión aplicados a la traducción automática. Estos enfoques permiten construir un marco conceptual robusto para comprender el funcionamiento de los modelos de inteligencia artificial generativa y delinear criterios pedagógicos para su integración en la formación profesional. En esta línea, se enfatiza la necesidad de cultivar una postura crítica en los futuros traductores, que favorezca la reflexión sobre los resultados generados por la IA y promueva una evaluación rigurosa en términos de precisión terminológica, adecuación estilística y fidelidad semántica.

Complementariamente, este trabajo articula el análisis teórico con el examen empírico de casos concretos. Mediante la comparación entre traducciones producidas por sistemas de IA y versiones elaboradas por traductores humanos especializados, se exploran los alcances y limitaciones de la inteligencia artificial generativa en dominios donde la coherencia contextual y el rigor terminológico resultan determinantes. Este análisis permite identificar patrones sistemáticos de error, como la replicación de sesgos inherentes a los datos de entrenamiento o la aparición de distorsiones conceptuales que no pueden resolverse sin la intervención de un profesional capacitado.

En consonancia con estos objetivos, este trabajo examina cómo la formación de traductores debe estructurarse a partir de un modelo pedagógico que trascienda la enseñanza instrumental del uso de herramientas basadas en IA. Se aboga por un enfoque integrador que incorpore competencias analíticas, críticas y éticas, orientadas a una evaluación exhaustiva de los textos generados automáticamente. La propuesta formativa se centra en el desarrollo de habilidades estratégicas para la posesión, el análisis de errores recurrentes y la detección de sesgos lingüísticos o conceptuales. Este tipo de formación capacita al traductor para desenvolverse con solvencia en entornos híbridos y, al mismo tiempo, fomenta una actitud consciente y responsable frente al impacto que tiene la inteligencia artificial sobre la producción textual contemporánea.

Como corolario, este artículo plantea un marco conceptual integrador que vincula las competencias tecnológicas con las traductológicas, y propone una formación orientada a consolidar un perfil profesional híbrido.

Lejos de reducirse al uso funcional de herramientas automatizadas, este modelo posiciona al traductor como un mediador estratégico, capaz de gestionar –con criterio– los procesos de generación textual, evaluar con rigor la calidad del contenido producido por IA y garantizar su coherencia, precisión y adecuación cultural en función del contexto. Más allá de describir un fenómeno tecnológico, el texto invita a repensar la práctica traductológica a la luz de estos cambios, y propone un modelo formativo que prepara a los futuros profesionales para desempeñar un papel activo en la supervisión, validación y adaptación de contenidos generados en entornos donde traducción y tecnología se entrelazan de manera indisoluble.

### **De la predicción a la creación: la evolución del aprendizaje automatizado hacia la IA generativa**

La inteligencia artificial ha transformado numerosas industrias, incluida la traducción. No obstante, para comprender el alcance de esta transformación, es crucial distinguir entre el aprendizaje automatizado (*machine learning* o ML) y la inteligencia artificial generativa (IAG). Ambos pertenecen al campo de la IA, pero presentan diferencias sustanciales en cuanto a sus capacidades y aplicaciones (Goodfellow, Bengio y Courville, 2016; Jurafsky y Martin, 2021).

El aprendizaje automatizado se basa en la identificación de patrones a partir de grandes volúmenes de datos. Su objetivo es mejorar la precisión de predicciones y clasificaciones mediante el entrenamiento de modelos en tareas específicas (LeCun, Bengio y Hinton, 2015), como la traducción automática de textos o la detección de errores en traducciones ya realizadas. En este sentido, el ML ha sido ampliamente utilizado en herramientas de traducción asistida, lo que ha permitido mejorar los resultados a partir de datos previamente procesados.

La inteligencia artificial generativa representa un avance cualitativo dentro del campo de la inteligencia artificial, ya que no solo analiza patrones, sino que es capaz de generar contenido nuevo. Su capacidad para producir texto, imágenes, audio y código con niveles de sofisticación sin precedentes ha transformado la producción lingüística automatizada. En el ámbito de la traducción, esto se plasma en modelos que no se limitan a la simple sustitución de palabras, sino que generan textos con mayor fluidez, coherencia y sensibilidad contextual (Vincent, 2021; OpenAI, 2023).

Para comprender esta evolución, resulta imprescindible examinar los principios tecnológicos que han permitido su desarrollo. En las siguientes secciones, se analizan el aprendizaje automatizado (*machine learning*), el aprendizaje profundo (*deep learning*) y los modelos de lenguaje de gran escala (LLM), cuyos avances han sido determinantes en la transformación de la traducción automática (Vaswani et al., 2017; Bender et al., 2021).

## ***Machine learning*: la base del aprendizaje automatizado**

El aprendizaje automatizado o *machine learning* (ML) es un subconjunto de la inteligencia artificial que permite a las máquinas mejorar su rendimiento en una tarea específica a partir de datos sin necesidad de una programación explícita (Goodfellow, Bengio y Courville, 2016; LeCun, Bengio y Hinton, 2015). Este concepto ha sido utilizado en traducción automática durante décadas. Herramientas como Google Translate en sus primeras versiones utilizaban modelos basados en reglas y traducción estadística, donde las palabras y frases más probables se seleccionaban en función de grandes volúmenes de texto previamente traducidos.

Los modelos tradicionales de ML funcionan mediante el entrenamiento de algoritmos con enormes conjuntos de datos. A partir de ellos, el sistema aprende a reconocer patrones y a hacer predicciones. Por ejemplo, en un sistema de traducción automática basado en ML, el algoritmo analizaría miles de traducciones previas y determinaría cuál es la mejor opción en función del contexto. Sin embargo, estos sistemas tienen una limitación importante: su incapacidad para generar traducciones realmente fluidas y adaptadas a matices culturales o estilísticos.

La evolución del ML hacia el *deep learning*, o aprendizaje profundo, permitió superar muchas de estas barreras (LeCun, Bengio y Hinton, 2015).

### **La revolución del lenguaje**

El aprendizaje profundo es una subdisciplina del aprendizaje automatizado que emplea redes neuronales artificiales con múltiples capas para procesar datos de manera jerárquica y eficiente. Estas redes, inspiradas en la estructura del cerebro humano, permiten modelar relaciones complejas dentro de los datos, identificando patrones con una capacidad superior a la de los métodos tradicionales de aprendizaje automático (LeCun, Bengio y Hinton, 2015). A diferencia de los enfoques convencionales, donde los sistemas dependen de reglas explícitas o modelos estadísticos, el aprendizaje profundo permite que los algoritmos descubran representaciones y estructuras subyacentes a partir de grandes volúmenes de datos sin intervención humana directa. Goodfellow, Bengio y Courville señalan que los primeros algoritmos de aprendizaje fueron concebidos como modelos computacionales basados en el funcionamiento del cerebro, con el objetivo de replicar sus procesos de aprendizaje. Por esta razón, el aprendizaje profundo ha sido asociado con el concepto de redes neuronales<sup>1</sup> artificiales (ANN). Desde esta perspectiva, los modelos de aprendizaje profundo se inspiran en la es-

---

<sup>1</sup> La perspectiva neuronal del aprendizaje profundo se apoya en dos principios clave: primero, el cerebro demuestra que la inteligencia es posible, por lo que un enfoque viable para desarrollar inteligencia artificial es modelar sus principios computacionales. Segundo, comprender los mecanismos de la inteligencia humana no solo tiene un valor teórico, sino que también ofrece aplicaciones prácticas en la ingeniería.

estructura y principios del cerebro biológico, ya sea humano o de otras especies, para desarrollar sistemas capaces de procesar información de manera similar.

Esta capacidad de extraer patrones complejos y modelar relaciones jerárquicas ha permitido avances significativos en múltiples áreas –una de las más destacadas es el Procesamiento del Lenguaje Natural, un subcampo del aprendizaje profundo, que se erige como un área fundamental pues permite la interacción entre las computadoras y el lenguaje humano–. Su desarrollo busca dotar a las máquinas de la capacidad de comprender, interpretar y generar texto de manera natural, reduciendo así la brecha entre la comunicación humana y la inteligencia artificial (Jurafsky y Martin, 2021). Para ello, el PLN integra modelos de lingüística computacional con técnicas avanzadas de aprendizaje automático y aprendizaje profundo, lo que posibilita el análisis estructural de un texto, la inferencia de significados, la identificación de relaciones semánticas y la adaptación del lenguaje generado al contexto específico.

La integración del aprendizaje profundo en el Procesamiento del Lenguaje Natural (PLN) ha sido fundamental para el desarrollo de la Traducción Automática Neuronal (*Neural Machine Translation*, NMT por sus siglas en inglés). A diferencia de los métodos estadísticos tradicionales, que fragmentaban el texto y dependían de probabilidades derivadas de grandes corpus de datos, la NMT utiliza redes neuronales profundas para procesar y traducir oraciones completas en su totalidad. Este enfoque holístico permite capturar dependencias contextuales<sup>2</sup> a largo plazo, resultando en traducciones más coherentes y fluidas.

Un avance crucial en la NMT fue la introducción del mecanismo de atención,<sup>3</sup> propuesto por Bahdanau, Cho y Bengio en 2015. Este mecanismo

---

2 En procesamiento del lenguaje natural, las **dependencias contextuales a largo plazo** se refieren a la capacidad de un modelo para relacionar palabras, frases o ideas que aparecen separadas dentro de un mismo texto. Los modelos tradicionales de traducción automática, como los basados en reglas o en métodos estadísticos, tenían dificultades para mantener la coherencia en textos extensos, ya que procesaban segmentos de manera aislada. En cambio, los enfoques basados en aprendizaje profundo, como la Traducción Automática Neuronal (NMT), utilizan arquitecturas como los transformadores, que analizan el contexto global de una oración o incluso de un párrafo completo. Esto permite generar traducciones más fluidas y coherentes, evitando errores como cambios de sujeto inesperados, inconsistencias terminológicas o falta de concordancia entre partes del texto.

3 El mecanismo de atención (*attention mechanism*) es una técnica utilizada en redes neuronales, especialmente en el ámbito del aprendizaje profundo y la traducción automática neuronal (NMT, *Neural Machine Translation*), que permite a los modelos enfocarse en partes específicas de la entrada mientras generan la salida. Los modelos de traducción automática neuronal tradicionales usaban un enfoque encoder-decoder donde toda la información de una oración de entrada se comprimía en un único vector de contexto de longitud fija. Esto generaba problemas en oraciones largas, ya que el modelo tenía que recordar toda la información relevante de la oración de origen sin poder priorizar ciertas palabras o frases clave. El mecanismo de atención resuelve este problema al permitir que el modelo asigne distintos pesos a diferentes partes de la oración de entrada mientras genera cada palabra de la oración de salida.

permite que los modelos de traducción se enfoquen en diferentes partes de la oración de origen al generar cada palabra de la traducción, asignando pesos a las palabras más relevantes según el contexto. Esta capacidad de focalización mejora significativamente la calidad de las traducciones, en especial en oraciones largas o complejas, donde el significado depende de múltiples elementos interconectados.

Posteriormente, la evolución de la NMT y del PLN en general alcanzó un hito significativo con la introducción de los modelos *Transformer*, presentados en 2017 por Vaswani et al. Estos modelos abandonaron las arquitecturas basadas en redes neuronales recurrentes en favor de una estructura que utiliza exclusivamente mecanismos de autoatención (*self-attention*). Esta innovación permite procesar oraciones completas en paralelo, mejorando la eficiencia y la capacidad del modelo para capturar relaciones contextuales que abarcan amplios segmentos del texto. Esto significa que el sistema puede entender y mantener coherencia en información distribuida a lo largo de párrafos o incluso capítulos, lo cual es esencial para tareas que requieren una comprensión profunda del contenido. Como resultado, los modelos Transformer han potenciado aplicaciones como la traducción automática, la generación de texto y la interacción conversacional con asistentes virtuales y así han establecido nuevos estándares en precisión y fluidez en el PLN.

En definitiva, el aprendizaje profundo y el PLN han transformado radicalmente la traducción automática, al proporcionar herramientas que permiten alcanzar un nivel de precisión y coherencia lingüística sin precedentes. No obstante, la supervisión humana sigue siendo indispensable para garantizar la fidelidad respecto del mensaje original, la adecuación cultural y la precisión terminológica, especialmente en campos técnicos o especializados (Castilho et al., 2017).

#### Recuadro 1 - Transformers y autoatención: La revolución del PLN

Los Transformers son una arquitectura de modelos de aprendizaje profundo que han revolucionado el Procesamiento del Lenguaje Natural (PLN) y otras áreas, debido a varias ventajas clave sobre modelos anteriores:

- **Paralelización:** A diferencia de las Redes Neuronales Recurrentes (RNN), que procesan las palabras de una oración de forma secuencial, los Transformers pueden analizar todas las palabras simultáneamente. Esto permite aprovechar mejor las capacidades de hardware moderno, reduciendo significativamente los tiempos de entrenamiento y procesamiento.
- **Captura de dependencias a largo plazo:** Las RNN suelen tener dificultades para recordar información de palabras distantes en una oración larga. Los Transformers, mediante su mecanismo de atención, pueden relacionar directamente cualquier palabra con otra, independientemente de la distancia entre ellas y así mejorar la comprensión del contexto global.
- **Eficiencia en el manejo de secuencias largas:** Gracias a su capacidad para procesar todas las palabras simultáneamente y enfocarse en las más relevantes, los Transformers manejan de manera más efectiva oraciones o documentos extensos sin perder contexto.

Una característica central de los Transformers es el **mecanismo de autoatención** (*self-attention*). Este permite al modelo evaluar la importancia de cada palabra en una oración en relación con las demás. Por ejemplo, al procesar la oración “el gato que vive en la casa es negro”, el mecanismo de autoatención ayuda al modelo a entender que “que vive en la casa” se refiere a “el gato”, estableciendo conexiones directas entre estas palabras.

En términos sencillos, el **mecanismo de autoatención** funciona asignando un peso a cada palabra según su relevancia, con el objeto de comprender el significado de las demás palabras de la oración. Esto le permite al modelo enfocarse en las partes más importantes del texto al generar traducciones, resumir información o realizar otras tareas de PLN.

En resumen, gracias a su capacidad para procesar información en paralelo y su mecanismo de autoatención, los Transformers han mejorado significativamente la eficiencia y precisión de las tareas de procesamiento del lenguaje, superando así las limitaciones de modelos anteriores como las RNN.

### **Modelos de lenguaje de gran escala (LLM): la Base de la IA Generativa**

Los Modelos de Lenguaje de Gran Escala (LLM), como GPT-4 de OpenAI y Gemini de Google DeepMind, han revolucionado la traducción automática al alcanzar niveles de sofisticación sin precedentes. Estos modelos, entrenados con billones de parámetros, poseen la capacidad de comprender y generar lenguaje humano con una fluidez y coherencia notables, comparables al desempeño de traductores humanos experimentados. Su desarrollo ha sido fundamental para el avance de aplicaciones que requieren generación de contenido, como la traducción automática, la redacción creativa y la interacción conversacional. A pesar de los avances significativos, los modelos de lenguaje de gran escala (LLM) aún presentan limitaciones, como la generación de traducciones imprecisas en contextos especializados y la perpetuación de sesgos presentes en los datos de entrenamiento.

A diferencia de los sistemas tradicionales de aprendizaje automático, que suelen operar mediante reglas predefinidas o análisis estadísticos limitados, los LLM emplean arquitecturas avanzadas de redes neuronales, como los transformadores. Estas estructuras les permiten procesar y generar texto considerando contextos amplios, lo que genera respuestas más coherentes y contextualmente apropiadas. Por ejemplo, modelos como GPT (*Generative Pre-trained Transformer*) de OpenAI y Gemini de Google han demostrado habilidades sobresalientes en la generación de texto que refleja una comprensión profunda del contexto y los matices del lenguaje.

El funcionamiento de los LLM se sustenta en tres pilares fundamentales:

En primer lugar, el pre-entrenamiento masivo. En esta fase, el modelo se expone a enormes volúmenes de texto provenientes de diversas fuentes, como libros, artículos y contenido web. Este proceso le permite aprender patrones lingüísticos complejos, estructuras gramaticales y relaciones se-

mánticas y sienta así las bases para una comprensión profunda del lenguaje.

En segundo lugar, el ajuste fino o *fine-tuning*. Tras el pre-entrenamiento, el modelo se especializa en tareas específicas mediante la exposición a conjuntos de datos más reducidos y especializados. Por ejemplo, para aplicaciones en traducción médica, el modelo se ajusta con textos médicos para mejorar su precisión y relevancia en ese dominio particular.

El tercer y último pilar es la generación de contenido. Con el conocimiento adquirido en las etapas anteriores, el LLM es capaz de generar texto nuevo que es coherente y adecuado al contexto. Esto incluye desde la traducción de idiomas hasta la creación de contenido original, adaptándose a diferentes estilos y registros según las necesidades específicas.

La relevancia de los LLM en la IA generativa radica en su capacidad para producir contenido que no solo es gramaticalmente correcto, sino también contextualmente pertinente y estilísticamente adecuado. Esta habilidad es esencial en aplicaciones donde la generación de lenguaje natural de alta calidad es crucial, como en asistentes virtuales, chatbots avanzados y sistemas de traducción automática que buscan emular la precisión y fluidez de un traductor humano experimentado.

No obstante, a pesar de sus impresionantes capacidades, los LLM presentan desafíos importantes. Pueden generar respuestas incorrectas o inapropiadas si se enfrentan a contextos ambiguos o si los datos de entrenamiento contienen sesgos, lo que puede afectar a la equidad y objetividad en sus resultados (Bender et al., 2021).<sup>4</sup> Además, aunque estos modelos son muy eficientes en la producción de contenido, carecen de una verdadera comprensión del lenguaje y del contexto, lo que puede derivar en errores críticos en áreas sensibles como la medicina, el derecho o la comunicación intercultural. Por esta razón, la supervisión humana sigue siendo esencial para garantizar la precisión y la adecuación cultural de sus respuestas.

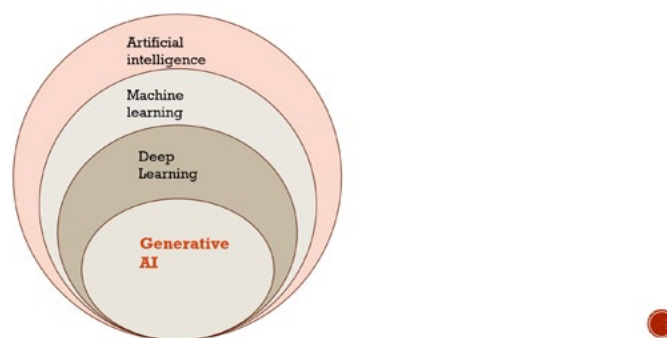
En este contexto, los Modelos de Lenguaje de Gran Escala constituyen la base de la Inteligencia Artificial Generativa pues proporcionan las herramientas avanzadas para la creación autónoma de contenido lingüístico de alta calidad. Su capacidad para interpretar y generar texto ha transformado industrias como la traducción automática, la redacción asistida y la aten-

---

<sup>4</sup> Bender et al. (2021) advierten sobre los riesgos inherentes al uso de modelos de lenguaje de gran escala, destacando que, si bien estos sistemas han logrado avances significativos en tareas de procesamiento del lenguaje natural, también presentan limitaciones fundamentales. En primer lugar, señalan que estos modelos no comprenden realmente el significado del lenguaje, sino que simplemente generan texto basado en patrones estadísticos aprendidos de datos masivos. Como resultado, pueden producir respuestas que parecen coherentes, pero que carecen de una base semántica real, lo que puede llevar a la difusión de información errónea o sesgada. Además, los autores destacan que el entrenamiento de estos modelos se basa en conjuntos de datos que reflejan sesgos históricos y dominantes, lo que puede perpetuar desigualdades estructurales y reforzar estereotipos perjudiciales. Finalmente, advierten sobre los costos ambientales y financieros asociados con el desarrollo de estos modelos, así como la necesidad de prácticas de documentación y curación de datos más rigurosas para mitigar estos problemas

ción al cliente. Sin embargo, su desarrollo debe ir acompañado de estrategias que mitiguen sus limitaciones, asegurando un uso responsable y ético que maximice su potencial sin comprometer la precisión y la equidad en la generación de contenido.

### Gráfico 1: Evolución de la Inteligencia Artificial: De la Automatización a la Generación



#### Recuadro 2. De la Máquina al Cerebro: La Revolución de la Traducción con IA Generativa

La traducción automática ha experimentado una notable evolución desde sus inicios, transitando por diversos enfoques que han buscado mejorar la precisión y la naturalidad de las traducciones. Inicialmente, los sistemas se basaban en reglas lingüísticas predefinidas y diccionarios bilingües, analizando la estructura gramatical del texto fuente para generar el texto en el idioma de destino. Estos sistemas pioneros en su tiempo eran rígidos y limitados en su capacidad para manejar ambigüedades y matices del lenguaje.

Posteriormente, en las décadas de 1980 y 1990, surgió la Traducción Automática Estadística (SMT), que empleaba modelos estadísticos fundamentados en extensos corpus de textos paralelos. En lugar de depender de reglas predefinidas, estos sistemas calculaban las probabilidades de correspondencia entre frases del idioma fuente y el idioma de destino. Aunque mejoraron la fluidez en comparación con los sistemas basados en reglas, aún enfrentaban desafíos con estructuras gramaticales complejas y contextos culturales.

Con el auge del aprendizaje profundo, la Traducción Automática Neuronal (NMT) ha transformado significativamente el campo de la traducción. Mediante la utilización de redes neuronales, la NMT modela el proceso de traducción de extremo a extremo, considerando el contexto completo de una oración. Esto permite generar traducciones más coherentes y naturales. Sin embargo, la NMT requiere grandes cantidades de datos y recursos computacionales para entrenar modelos efectivos.

Sin embargo, aún persisten grandes desafíos. Los sistemas suelen tener dificultades para interpretar ironías, humor, expresiones idiomáticas y referencias culturales. Por ejemplo, durante una ceremonia de premios, la traducción automática de comentarios irónicos puede resultar

### Diferencia entre modelos y herramientas

La distinción entre modelos y herramientas de inteligencia artificial es esencial tanto para comprender su base técnica como para optimizar su uso, mejorar la toma de decisiones en el proceso de traducción y analizar su impacto en la profesión del traductor.

Desde una perspectiva tecnológica, la distinción es clave porque los modelos de IA representan la base computacional y algorítmica de la traducción automática, mientras que las herramientas son aplicaciones que incorporan esos modelos en entornos accesibles para los usuarios (Goodfellow, I., Bengio, Y., & Courville, A., 2016). No todos los modelos funcionan de la misma manera ni tienen las mismas capacidades, y conocer sus diferencias permite a los traductores elegir la mejor opción para cada tarea. Dependiendo del contexto de trabajo, la elección de la herramienta adecuada debe considerar qué modelo subyacente la respalda.

Para citar algunos ejemplos, el modelo NLLB-200 (*No Language Left Behind*) fue desarrollado por Meta con el propósito de mejorar la traducción automática en idiomas con pocos recursos lingüísticos. A diferencia de los modelos tradicionales, que suelen centrarse en lenguas con grandes volúmenes de datos de entrenamiento (como inglés, español o chino), NLLB-200 fue entrenado para traducir con alta precisión en 200 idiomas, incluidas muchas lenguas minoritarias y regionales que anteriormente no contaban con traducción automática de calidad.<sup>5</sup> Esto representa un avance clave para la inclusión digital y el acceso equitativo a la información en comunidades que no tienen una fuerte presencia en internet.

Por otro lado, SEAMLESSM4T es un modelo también desarrollado por Meta, pero con un enfoque distinto: la traducción multimodal. Este modelo no solo traduce texto a texto, sino que también permite la conversión de habla a habla, habla a texto y texto a voz en múltiples idiomas.<sup>6</sup> Su integración de voz y texto lo hace particularmente útil en contextos audiovisuales, como la subtítulos automática de videos, la traducción en conferencias multilingües o la interpretación en tiempo real en conversaciones internacionales. Mientras que NLLB-200 está optimizado para expandir la cobertura lingüística, SEAMLESSM4T está diseñado para ofrecer una experiencia más natural en la comunicación multilingüe oral y escrita.

En el ámbito profesional, la traducción automática no es un reemplazo del trabajo humano, sino una herramienta que complementa la labor del traductor. Sin embargo, el desconocimiento sobre cómo operan los modelos puede llevar a una dependencia excesiva de las herramientas sin una evaluación crítica de sus resultados. Esto puede conducir a una confianza ciega en la salida del sistema, sin reconocer que las herramientas pueden propagar errores sistemáticos derivados de los modelos en los que se basan.

A pesar de los avances en modelos de IA como ChatGPT-4, Gemini o Claude, la traducción automática neuronal sigue presentando errores sistemáticos recurrentes, especialmente en dominios técnicos y especializados. Como advierten Castilho et al. (2017),<sup>7</sup> estos sistemas pueden generar

<sup>5</sup> <https://ai.meta.com/research/no-language-left-behind/es-es/>

<sup>6</sup> <https://ai.meta.com/research/publications/seamlessm4t-massively-multilingual-multimodal-machine-translation/>

<sup>7</sup> Castilho et al. (2017) advierten que, si bien la Traducción Automática Neuronal (NMT) ha logrado avances significativos en términos de fluidez y calidad percibida en compara-

traducciones con una apariencia de precisión engañosa, lo que aumenta el riesgo de que los usuarios confíen ciegamente en sus resultados sin cuestionar su validez. Un problema frecuente es la omisión o adición de información, ya que los modelos pueden eliminar detalles clave o insertar elementos inexistentes en la traducción, afectando la fidelidad del mensaje original. Aunque han mejorado en la adaptación terminológica, todavía presentan dificultades en la traducción de jerga técnica o términos especializados, lo que puede generar errores críticos en campos como el derecho, la medicina o la ingeniería.

A pesar de los avances en la traducción automática, persisten desafíos significativos que requieren atención. Uno de ellos es la “ilusión de precisión”: los sistemas actuales pueden generar textos con una fluidez impresionante, lo que lleva a los usuarios a asumir que el contenido es siempre correcto. Sin embargo, estos modelos operan mediante predicciones probabilísticas sin una verdadera comprensión del lenguaje, limitándose a generar la respuesta más probable según patrones aprendidos en grandes volúmenes de datos. Como resultado, los errores pueden ser menos evidentes pero más peligrosos, ya que los textos producidos parecen correctos incluso cuando contienen inexactitudes. Esta percepción errónea puede llevar a una disminución en la intervención humana en la posesición, lo que facilita la propagación de errores sistemáticos sin una supervisión adecuada.

Además, la falta de transparencia en el funcionamiento de los modelos de inteligencia artificial y la presencia de sesgos en los datos de entrenamiento plantean desafíos adicionales. Las herramientas de traducción automática pueden reflejar prejuicios lingüísticos, ideológicos o culturales presentes en los datos con los que fueron entrenadas, comprometiendo la objetividad y precisión de las traducciones. La opacidad de estos modelos dificulta la comprensión por parte de los usuarios sobre cómo se generan los resultados, lo que agrava el problema. Por lo tanto, aunque los avances en IA han mejorado la calidad general de la traducción automática, la revisión y supervisión humanas siguen siendo esenciales para garantizar la precisión, la adecuación terminológica y la adaptación cultural de los textos generados.

Si bien los modelos de IA constituyen la arquitectura algorítmica sobre la cual se fundamenta la traducción automática, su impacto en la práctica profesional del traductor se materializa a través de las herramientas de IA. Estas herramientas, concebidas como aplicaciones que integran los mode-

---

ción con la Traducción Automática Estadística (SMT), sus resultados no siempre superan a los sistemas anteriores cuando se evalúan con métodos humanos. En varios de sus estudios de caso, encontraron que las evaluaciones automáticas favorecían a NMT, pero los revisores humanos identificaban errores recurrentes que el sistema no corregía, como omisiones, adiciones y errores de traducción de segmentos largos. Además, la naturaleza opaca de los modelos neuronales dificulta entender el origen de estos fallos, lo cual puede crear una falsa percepción de precisión y fiabilidad en los usuarios finales, sobre todo cuando no se aplica una revisión crítica o posesición rigurosa del contenido traducido.

los en interfaces accesibles, facilitan el uso de la inteligencia artificial en el ámbito traductológico y determinan la manera en que los profesionales interactúan con la tecnología, evaluando sus resultados y adaptándolos a sus necesidades específicas.

En este sentido, resulta imperativo diferenciar entre el modelo subyacente y la herramienta que lo implementa, pues, aunque una misma tecnología algorítmica puede ser utilizada en múltiples plataformas, la manera en que se presenta al usuario, las funciones adicionales que incorpora y los parámetros de personalización que permite varían considerablemente. La traducción automática no se reduce a la simple sustitución de palabras entre idiomas, sino que implica la integración de distintos niveles de procesamiento, desde la identificación del significado y la selección terminológica hasta la adaptación al contexto discursivo y pragmático del texto.

A diferencia de los modelos, que operan en el nivel estructural de la IA, las herramientas representan el espacio de mediación entre la tecnología y el usuario. No se trata únicamente de interfaces de acceso a la traducción automática, sino de sistemas que combinan distintas funcionalidades para optimizar la calidad del resultado final. En este proceso, el traductor deja de ser un consumidor pasivo de la tecnología y se convierte en un agente crítico que debe evaluar, corregir y mejorar las propuestas generadas por la IA.

El desarrollo de herramientas de inteligencia artificial para la traducción ha avanzado enormemente en las últimas décadas. Más allá de la simple correspondencia lingüística, estas herramientas han evolucionado para integrarse en flujos de trabajo complejos que requieren precisión, coherencia y adaptabilidad al contexto de uso. A través de diversas mejoras tecnológicas, las herramientas de IA han optimizado los procesos de traducción automática, para incorporar funcionalidades que permiten una mayor personalización del resultado y una intervención más eficiente por parte del traductor humano. Este desarrollo responde a la necesidad de superar las limitaciones de los primeros sistemas de traducción automática, cuya aplicación se restringía a tareas mecánicas y cuya utilidad en textos especializados era limitada. En la actualidad, las herramientas han trascendido esta concepción inicial y se han consolidado como instrumentos fundamentales en el entorno de la traducción profesional.

Uno de los avances más relevantes en este ámbito es la implementación de sistemas híbridos y la combinación de modelos. En sus primeras etapas las herramientas de traducción automática operaban sobre un único motor de inteligencia artificial; hoy en día, muchas de ellas integran múltiples modelos con el fin de mejorar la calidad del resultado. Esta estrategia permite optimizar la traducción en distintos contextos y reducir la incidencia de errores sistemáticos. Un ejemplo destacado de este enfoque es *Smartcat*, una plataforma que combina traducción automática con memorias de traducción y bases terminológicas. Esta integración posibilita un mayor control sobre la coherencia terminológica y contribuye a la uniformidad en textos especializados, lo que resulta esencial en ámbitos como el jurídico, el

médico o el técnico, donde la precisión es un requisito ineludible.

Otro aspecto fundamental en la evolución de estas herramientas es la incorporación de sistemas de posesición asistida y optimización del resultado. Una de las críticas más recurrentes a la traducción automática es su limitada capacidad para interpretar el lenguaje en su dimensión contextual y pragmática. A pesar de los avances en modelos neuronales, la traducción automática sigue enfrentando dificultades para captar matices lingüísticos, ironías y estructuras complejas. Para mitigar este problema, herramientas como DeepL han desarrollado sistemas que permiten ajustar la traducción a distintos niveles de formalidad y estilo. Este tipo de funcionalidades le ofrece al usuario opciones alternativas de traducción que pueden seleccionarse en función del registro y el propósito del texto, lo que representa un avance significativo en la integración de la inteligencia artificial en la práctica profesional del traductor.

Otra de las innovaciones más relevantes en la traducción automática es la incorporación de traducción en tiempo real y sistemas multimodales. La demanda de soluciones inmediatas en un mundo globalizado ha impulsado el desarrollo de aplicaciones que permiten la traducción simultánea de textos y discursos orales. Herramientas como Microsoft *Translator* han ampliado las posibilidades de la inteligencia artificial en este campo al integrar funciones de transcripción de voz, subtítulos automática y traducción simultánea en conferencias multilingües. Estas innovaciones no solo han transformado la práctica de la traducción, sino que han tenido un enorme impacto en el campo de la interpretación, introduciendo sistemas de apoyo basados en IA que facilitan la comunicación en eventos internacionales. La capacidad de traducir voz en tiempo real sin intervención humana representa un avance tecnológico significativo, aunque sigue enfrentando desafíos en términos de precisión, manejo de acentos y preservación de la intencionalidad discursiva.

En conclusión, el desarrollo de herramientas de IA para la traducción ha permitido introducir mejoras que trascienden la mera equivalencia lingüística, facilitando la integración de la inteligencia artificial en el trabajo del traductor profesional. La combinación de modelos, la posesición asistida, la gestión terminológica, la traducción en tiempo real y la adaptación estilística son solo algunas de las innovaciones que han transformado el panorama de la traducción automática. No obstante, a pesar de estos avances, la intervención humana sigue siendo imprescindible para garantizar la precisión, coherencia y adecuación de las traducciones generadas por IA. La interacción entre traductores y tecnología se configura, así, como un elemento central en la evolución del sector, consolidando un nuevo paradigma en el que la inteligencia artificial no reemplaza la labor del traductor, sino que redefine sus funciones y amplía sus posibilidades de acción.

En el contexto de la formación de traductores, la integración de un conocimiento profundo sobre los modelos de inteligencia artificial resulta fundamental para una adopción crítica y estratégica de la tecnología en el ámbito

educativo (Bowker, L. 2021, pp. 57-76). No se trata únicamente de dotar a los futuros profesionales de la capacidad de utilizar herramientas como DeepL o *Google Translate*, sino de desarrollar una competencia analítica que les permita comprender la lógica de funcionamiento de estos sistemas, evaluar sus resultados con criterio profesional y optimizar su aplicación en función de las exigencias específicas de cada encargo de traducción (Melby, A., 2020). La formación en inteligencia artificial aplicada a la traducción incrementa la eficiencia y la productividad. Además, capacita a los profesionales para detectar y corregir errores sistemáticos, identificar sesgos en los datos de entrenamiento y tomar decisiones informadas sobre el uso de la traducción automática en diversos contextos profesionales.

Desde esta perspectiva, la diferenciación entre modelos y herramientas de IA adquiere un valor pedagógico y epistemológico central. Comprender esta distinción proporciona un marco conceptual que permite analizar en profundidad las capacidades y limitaciones de la traducción automática, lo que ayuda a evitar una dependencia acrítica de la tecnología (Pym, A. 2011, pp. 163-182). Asimismo, esta diferenciación facilita la optimización de los procesos de traducción en el ámbito profesional pues permite un uso estratégico de las herramientas disponibles y fomenta una mayor conciencia sobre su impacto en la comunicación intercultural. En una disciplina en la que la exactitud terminológica y la fidelidad semántica son aspectos esenciales, al conocer el funcionamiento de los modelos subyacentes los traductores pueden intervenir de manera informada y efectiva en la posesición para corregir distorsiones y asegurar la calidad del producto final.

En este sentido, no basta con conocer qué herramientas existen o cómo se emplean en términos operativos; es imprescindible comprender los principios que rigen los modelos que las sustentan, reconocer los sesgos que pueden incorporar y evaluar hasta qué punto sus resultados responden a los estándares de calidad exigidos en cada encargo<sup>8</sup>. La inteligencia artificial ha transformado el campo de la traducción, pero su implementación eficaz requiere una formación especializada que trascienda la simple capacitación en el uso de software y promueva una visión crítica y reflexiva sobre sus implicaciones en la práctica profesional. De este modo, el traductor no se convierte en un usuario pasivo de la tecnología, sino en un mediador experto capaz de combinar los beneficios de la automatización con el criterio humano, garantizar la precisión, la coherencia y la adecuación cultural de los textos traducidos.

---

8 Koehn, P. (2020). *Neural Machine Translation*. Cambridge University Press.

**Tabla 1: Modelos y herramientas**

<b>Modelos</b>	<b>Descripción</b>
GPT-4	Modelo de lenguaje desarrollado por OpenAI, utilizado para generar texto con coherencia avanzada.
NLLB-200	Modelo de Meta diseñado para traducir con alta precisión en 200 idiomas, incluyendo lenguas con pocos recursos.
SEAMLESSM4T	Modelo de Meta que permite la traducción multimodal entre voz y texto, mejorando la accesibilidad lingüística.
M2M-100	Modelo de Facebook AI que admite traducción en 100 idiomas sin necesidad de un idioma pivote.
Marian NMT	Modelo de traducción automática neuronal de código abierto, utilizado en diversas plataformas.
Whisper	Modelo de OpenAI especializado en la transcripción automática de audio a texto.
Herramientas	
ChatGPT	Aplicación que usa el modelo GPT para ofrecer respuestas interactivas y generación de contenido.
Google Translate	Herramienta de traducción automática que emplea distintos modelos de IA para ofrecer traducción en más de 100 idiomas.
DeepL	Plataforma que emplea modelos neuronales para generar traducciones de alta calidad con ajustes estilísticos.
Smartcat	Plataforma CAT que combina traducción automática con memorias de traducción y colaboración en la nube.
Microsoft Translator	Aplicación de traducción en tiempo real con integración de texto, voz y subtítulos.
Otter.ai	Herramienta basada en IA para transcribir reuniones y generar resúmenes de conversaciones en tiempo real.

### ¿Traductor o IA? Evaluación empírica de la calidad de traducción

Como mencionamos en las secciones anteriores, en la era de la inteligencia artificial, la traducción automática ha avanzado significativamente, en especial con la introducción de modelos de lenguaje de gran tamaño como ChatGPT-4. Estos desarrollos han generado debates sobre la eficacia de las traducciones realizadas por profesionales humanos en comparación con las generadas por inteligencia artificial. Para aportar evidencia empírica a esta

discusión, llevé a cabo un estudio que evaluó la calidad de las traducciones en el ámbito económico-financiero<sup>9</sup>, mediante el cual comparé el desempeño de traductores humanos con y sin especialización y el de ChatGPT-4<sup>10,11</sup>.

El estudio se centró en traducciones bidireccionales entre inglés y español, seleccionando textos representativos del sector económico-financiero debido a su complejidad terminológica y estilística. Para analizar cómo la especialización y la tecnología influyen en la calidad de las traducciones, se seleccionaron tres traducciones representativas de cada categoría: el modelo de IA Generativa ChatGPT-4, traductores especializados en economía y finanzas y traductores

---

9 La elección del campo económico-financiero responde a mi experiencia profesional, lo que facilita una evaluación más precisa y contextualizada de la calidad traductora.

10 Existen diversos modelos de inteligencia artificial diseñados para tareas de traducción automática. Entre los más destacados se encuentra NLLB-200 de Meta, un modelo capaz de traducir con alta calidad a 200 idiomas diferentes, incluyendo aquellos con recursos limitados. Otro ejemplo es SEAMLESSM4T, también desarrollado por Meta, que permite la traducción de voz a voz en 36 idiomas en tiempo real, conservando el tono y las emociones del hablante. Además, Google Neural Machine Translation (GNMT) utiliza redes neuronales profundas para mejorar la precisión y fluidez en las traducciones, evaluando segmentos de palabras en el contexto de oraciones completas.

Para este estudio, he optado por utilizar ChatGPT-4 como modelo de inteligencia artificial para la traducción. Esta elección se fundamenta en varias razones. En primer lugar, ChatGPT-4 ha demostrado una notable capacidad para comprender y generar lenguaje humano con fluidez y coherencia, lo que es esencial para traducir textos con matices complejos. Además, su entrenamiento en vastos conjuntos de datos le permite adaptarse a diferentes contextos y terminologías, incluyendo el ámbito económico-financiero, que es el foco de este análisis. Por último, la accesibilidad y flexibilidad de ChatGPT-4 facilitan su integración en diversos flujos de trabajo, permitiendo una evaluación práctica y eficiente en comparación con otros modelos.

11 Al momento de redactar este capítulo, se han desarrollado versiones más avanzadas de ChatGPT que ofrecen mejoras significativas en comparación con la versión 4.0. Por ejemplo, en mayo de 2024, OpenAI presentó GPT-4o, un modelo multimodal capaz de procesar y generar texto, imágenes y sonido. GPT-4o es notablemente más rápido y eficiente que su predecesor, proporcionando respuestas con menor latencia y mayor precisión. Además, en septiembre de 2024, se introdujo el modelo o1-preview, diseñado para emular un razonamiento más humano y abordar problemas complejos con mayor eficacia. Este modelo ha demostrado un rendimiento superior en áreas como programación competitiva y razonamiento científico.

A pesar de la disponibilidad de estas versiones más recientes, la elección de ChatGPT-4 para este estudio se justifica por la necesidad de mantener la coherencia y consistencia en los datos recopilados. Establecer el punto de corte en la versión 4.0 permite una evaluación comparativa más precisa, evitando las variables introducidas por las actualizaciones posteriores. Además, dado que ChatGPT-4 ha sido ampliamente adoptado y evaluado en diversos contextos, su desempeño está bien documentado, proporcionando una base sólida para el análisis empírico en el ámbito de la traducción económico-financiera.

Las mejoras en las versiones posteriores, como GPT-4o y o1-preview, incluyen una mayor velocidad de procesamiento, capacidad para manejar múltiples modalidades de información y habilidades de razonamiento más avanzadas. Estas actualizaciones amplían las aplicaciones potenciales de la inteligencia artificial en campos especializados, ofreciendo herramientas más robustas para profesionales y académicos. Sin embargo, para los fines de este estudio, centrarse en ChatGPT-4 proporciona un equilibrio adecuado entre funcionalidad y estabilidad, facilitando una evaluación rigurosa de la calidad de las traducciones generadas por IA en comparación con las realizadas por traductores humanos.

sin especialización. Esta estructura permitió analizar cómo la especialización y la tecnología influyen en la calidad de las traducciones.

Al comparar las traducciones de cada categoría, se identificaron fortalezas y debilidades específicas. Este análisis proporcionó una visión integral sobre la capacidad de la IA generativa para manejar textos especializados y su posición relativa frente a traductores humanos con distintos niveles de especialización.

### Ejemplo 1:

Source:	<b>Labour market developments present the most proximate risk to our monetary policy view. We assess the likelihood of wage-price spirals emerging from current tight labour markets and wage developments.</b>
IA:	Los desarrollos en el mercado laboral representan el riesgo más inmediato para nuestra perspectiva de política monetaria. Evaluamos la probabilidad de que surjan espirales de salarios y precios a partir de los actuales mercados laborales ajustados y los desarrollos salariales
Traductor sin especialización	Los sucesos en el mercado laboral plantean el riesgo más inmediato para nuestra visión de la política monetaria. Evaluamos la posibilidad de espirales de salarios y precios provocados por mercados laborales ajustados y los desarrollos en materia salarial.
Traductor especializado	El riesgo más inmediato para nuestra perspectiva de la política monetaria proviene del mercado laboral. Evaluamos si cabe esperar que la escasez de oferta de mano de obra y los aumentos salariales generen espirales de salarios y precios.

El análisis de estas traducciones permite identificar diferencias en la fidelidad al texto original, la adecuación terminológica y la naturalidad del discurso en español. En este caso, el fragmento original se centra en la relación entre la evolución del mercado laboral y los riesgos para la política monetaria, con énfasis en la posibilidad de espirales inflacionarias impulsadas por aumentos salariales. La comparación entre las versiones permite evaluar el grado de comprensión conceptual y el nivel de especialización de cada traductor.

Uno de los aspectos más llamativos en la traducción generada por la inteligencia artificial (IA) es su apego excesivo a la estructura del texto original.

Un ejemplo claro de esto es la traducción de *developments* como desarrollos, una elección que, aunque válida en ciertos contextos, no es la opción más natural en el lenguaje económico en español. En este ámbito, *developments* suele referirse a sucesos, cambios o evolución, dependiendo del contexto. La segunda versión humana opta por sucesos, lo cual es una mejora, aunque no aporta una ventaja significativa en términos de claridad o precisión. La tercera versión es la más acertada al evitar traducir *developments* de manera literal y reformular la idea de manera más natural en español.

Otro punto de divergencia clave es la traducción de *tight labor markets*, un concepto técnico que describe una situación en la que la demanda de trabajadores supera la oferta, lo que genera presiones salariales al alza. La IA traduce este término como “mercados laborales ajustados”, lo cual no transmite con precisión la noción de escasez de mano de obra. El traductor, con menos experiencia, tampoco logra captar la esencia del concepto y mantiene una traducción demasiado cercana al original. En cambio, el traductor especializado resuelve este desafío de manera más efectiva al reformularlo como “escasez de oferta de mano de obra”, una opción que refleja con mayor precisión la dinámica subyacente de los mercados laborales en tensión.

Asimismo, la última versión muestra una mayor flexibilidad en la estructuración de la oración, lo que permite una lectura más fluida en español. En lugar de seguir mecánicamente la construcción del original, como hacen las otras versiones, esta última reestructura la información para transmitir el mensaje de manera más clara y natural. Un ejemplo notable es la traducción de *we assess the likelihood* en la versión especializada, que opta por “evaluamos si cabe esperar”, en lugar de la traducción más rígida “evaluamos la probabilidad”, que se encuentra en las versiones anteriores. Esta elección no solo enriquece el discurso, sino que también refleja un conocimiento más profundo del lenguaje utilizado en informes económicos.

El análisis de estas versiones demuestra que la calidad de una traducción en el ámbito económico-financiero depende no solo de la precisión terminológica, sino también de la capacidad de transmitir conceptos de manera clara y natural en español. La traducción generada por IA es correcta en términos gramaticales, pero sigue demasiado de cerca la estructura del original y presenta elecciones léxicas poco idiomáticas, como “desarrollos” y “mercados laborales ajustados”. La versión del traductor con menos experiencia no ofrece mejoras significativas, ya que mantiene una estructura similar y no logra una traducción precisa del concepto de *tight labor markets*.

Por el contrario, la traducción realizada por un profesional con especialización en el área demuestra un mayor dominio del tema, ya que evita traducciones literales innecesarias, reformula el contenido de manera más natural y proporciona una versión que fluye mejor en español. La elección de “escasez de oferta de mano de obra” para *tight labor markets* y la reformulación de *we assess the likelihood* como “evaluamos si cabe esperar”, son ejemplos claros de cómo la experiencia en el campo permite producir traducciones más precisas y naturales.

**Ejemplo 2:**

Source:	<p><b>Tradeoffs in designing support policies for firms</b></p> <p>Withdrawal of untargeted fiscal support can increase bankruptcy, while prolonging untargeted support erodes fiscal space and risks corporate debt overhang.</p>
IA:	<p><b>Compensaciones al diseñar políticas de apoyo para empresas</b></p> <p>La retirada de apoyo fiscal no dirigido puede aumentar la bancarrota, mientras que prolongar el apoyo no dirigido erosiona el espacio fiscal y arriesga un exceso de deuda corporativa.</p>
Traductor sin especialización	<p><b>Concesiones en el diseño de políticas de apoyo a empresas</b></p> <p>El retiro de apoyo fiscal sin objetivo puede conducir a un aumento de las quiebras empresariales, en tanto prolongarlo erosiona el espacio fiscal y puede generar exceso de deuda empresarial.</p>
Traductor especializado	<p><b>Disyuntivas al diseñar políticas de apoyo al sector empresarial</b></p> <p>Retirar el apoyo fiscal no específico puede provocar más quiebras de empresas, pero continuar con el apoyo no específico erosiona el espacio fiscal y se corre el riesgo de un sobreendeudamiento del sector empresarial.</p>

La evaluación de distintas traducciones de un fragmento sobre políticas fiscales de apoyo empresarial permite identificar discrepancias en la precisión terminológica, la adecuación al registro financiero y la coherencia lingüística dentro del español técnico. En este análisis, se contrastan tres enfoques de traducción con el objetivo de determinar cuál logra capturar con mayor fidelidad el significado del texto original en inglés.

Uno de los primeros aspectos que resalta es la incoherencia terminológica en la traducción generada por la inteligencia artificial (IA). Mientras que utiliza “retirada” para *withdrawal*, un término de uso más frecuente en España, también emplea “bancarrota” para *bankruptcy*, una palabra que no es de uso común en el español europeo, pero sí en América Latina y en el español empleado en Estados Unidos. Este desbalance indica una falta de coherencia en la variante del español utilizada, lo que puede generar con-

fusión en el lector especializado. En contraste, las traducciones realizadas por los traductores humanos mantienen una coherencia mayor al utilizar “retiro” y “quiebras empresariales”, términos más alineados con el lenguaje financiero hispanohablante en su conjunto.

Otro punto clave en esta comparación es la traducción de *tradeoffs*, un concepto fundamental en economía y políticas públicas que implica la existencia de decisiones que conllevan costos y beneficios alternativos. La traducción automática opta por “compensaciones”, lo cual es incorrecto en este contexto, ya que sugiere un equilibrio entre dos factores sin reflejar la idea de toma de decisiones bajo restricciones. Una de las traducciones humanas utiliza “concesiones”, lo cual se acerca más al sentido correcto, aunque sigue sin expresar plenamente la necesidad de optar entre dos alternativas con implicaciones contrapuestas. La mejor elección, en este caso, es “disyuntivas”, ya que capta con precisión el dilema inherente en el diseño de políticas de apoyo y resuena con la idea planteada en la segunda parte de la oración.

La traducción de *untargeted fiscal support* también representa un desafío significativo. En la traducción generada por IA, el término se convierte en “apoyo fiscal no dirigido”, una opción que, aunque comprensible, no refleja con suficiente claridad la naturaleza de este tipo de ayuda gubernamental. La versión proporcionada por uno de los traductores humanos, “apoyo fiscal sin objetivo”, es aún más imprecisa, ya que “sin objetivo” sugiere una falta de propósito, en lugar de la ausencia de focalización en un sector específico. La opción más precisa es “apoyo fiscal no específico”, ya que este término se alinea con la jerga utilizada en economía y finanzas para describir medidas fiscales generalizadas, sin destinatarios concretos.

Finalmente, la traducción de *corporate debt overhang* también refleja diferencias en el nivel de especialización de los traductores. La opción de la IA, “exceso de deuda corporativa”, es una traducción literal que no plasma adecuadamente el matiz de *debt overhang*, el cual hace referencia a una situación en la que el alto nivel de endeudamiento impide nuevas inversiones y el crecimiento económico de la empresa. La alternativa “exceso de deuda empresarial” es una mejor aproximación, aunque sigue sin ser la terminología más utilizada en análisis financieros. La opción más precisa es “sobreendeudamiento del sector empresarial”, ya que encapsula con mayor fidelidad el fenómeno económico en cuestión y es un término ampliamente aceptado en literatura especializada.

Este análisis demuestra que la calidad de una traducción en el ámbito financiero no depende únicamente de la capacidad de procesar el lenguaje, sino también del conocimiento profundo de la terminología y los conceptos económicos. Aunque la IA ha logrado una traducción más acertada que la de un traductor humano sin experiencia en algunos términos, no está exenta de incongruencias y tiene poca precisión conceptual. La traducción realizada por un profesional con conocimiento del sector es, sin duda, la más adecuada, al emplear términos como “disyuntivas”, “apoyo fiscal no específico” y “sobreendeudamiento del sector empresarial”, lo que garantiza una mayor precisión y adecuación al público al que está dirigido el texto.

**Ejemplo 3:**

Source:	XX's 1.2 billion fixed income portfolio recently increased its allocation to callable agencies from 10% to 12%. The move was funded by selling Treasuries.
IA:	La cartera de renta fija de XX, que asciende a 1.2 mil millones, recientemente aumentó su asignación en agencias con opción de compra del 10% al 12%. El movimiento se financió mediante la venta de bonos del Tesoro.
Traductor sin especialización	La cartera de renta fija de XX por valor de 1.200 millones recientemente aumentó la asignación a títulos de agencia con opción de compra del 10% al 12%. El movimiento se financió con la venta de títulos del Tesoro.
Traductor especializado	Recientemente, la cartera de renta fija de XX valuada en 1.200 millones aumentó la asignación a títulos emitidos por agencias del gobierno estadounidense con opción de rescate, del 10% al 12%. Para financiarlo, la cartera vendió títulos del Tesoro.

En esta comparación de traducciones, se analizan las diferencias terminológicas y estilísticas en la adaptación al español de un fragmento financiero en inglés. La evaluación se enfoca en tres aspectos clave: precisión terminológica, claridad y coherencia en la traducción de términos técnicos.

Uno de los aspectos más relevantes es la traducción de *callable agencies*, un término financiero que hace referencia a títulos de deuda emitidos por agencias gubernamentales con opción de rescate anticipado. La primera opción traduce esto como “agencias con opción de compra”, lo cual es incorrecto, ya que da a entender que las agencias mismas pueden ser adquiridas, cuando en realidad se trata de títulos valores financieros. La segunda opción mejora la interpretación al referirse a “títulos de agencia con opción de compra”, aunque “opción de compra” sigue sin capturar con precisión el concepto. La tercera versión es la más clara y adecuada, ya que especifica que se trata de “títulos emitidos por agencias del gobierno estadounidense con opción de rescate”, lo cual refleja con mayor fidelidad la naturaleza del activo financiero.

Otro punto clave es la traducción del término *Treasuries*. La primera versión emplea “bonos del Tesoro”, una traducción frecuente, pero que se asocia más específicamente a los *Treasury Bonds*, que son solo una parte del

espectro de instrumentos de deuda emitidos por el gobierno de EE.UU. Las otras dos opciones eligen “títulos del Tesoro”, una elección más precisa, ya que abarca tanto los *Treasury Bonds* como los *Treasury Notes* y los *Treasury Bills*, y por lo tanto es más adecuada en el contexto financiero.

En conclusión, la tercera versión es la más precisa y alineada con el lenguaje técnico adecuado para el público financiero, ya que aclara el tipo de instrumentos involucrados, utiliza “opción de rescate” en lugar de “opción de compra” y emplea “títulos del Tesoro” en vez de “bonos del Tesoro”.

#### Ejemplo 4:

Source:	If you are looking to open a position to trade an earnings announcement, one of the simplest way is by buying or shorting the stock.
IA:	Si buscas abrir una posición para operar con un anuncio de ganancias, una de las formas más simples es comprando o vendiendo en corto la acción
Traductor sin especialización	Si busca abrir una posición para negociar un anuncio de ganancias, una de las formas más sencillas es comprar o vender acciones en corto.
Traductor especializado	Si desea abrir una posición para negociar en función de anuncios de resultados, una de las formas más simples es comprar acciones o venderlas en descubierto.

En esta comparación de traducciones, se analizan las diferencias en la adaptación al español de un texto financiero relacionado con la apertura de posiciones para operar en torno a anuncios de ganancias empresariales. La evaluación se enfoca en la terminología financiera, la claridad del mensaje y la adecuación a la audiencia.

Uno de los puntos más relevantes es la traducción de *earnings announcement*, un término fundamental en el contexto bursátil. La primera y segunda versiones optan por “anuncio de ganancias”, mientras que la tercera elige “anuncios de resultados”. Si bien “ganancias” es comprensible, en terminología financiera es más preciso hablar de “resultados”, ya que las empresas pueden reportar tanto ganancias como pérdidas. Por lo tanto, la tercera versión ofrece una mejor adaptación al lenguaje técnico del mercado financiero.

El concepto de trading también se refleja de manera diferente en cada versión. La primera traducción emplea “operar con”, mientras que las dos siguientes prefieren “negociar”. Si bien ambos términos pueden ser válidos, “negociar” es más formal y se utiliza con mayor frecuencia en el ámbito financiero. Además, la tercera versión refuerza la relación entre la posición y

el anuncio de resultados al incluir “en función de”, lo que proporciona una mayor claridad conceptual.

Otro aspecto importante es la traducción de *shorting the stock*, un concepto clave en la operativa bursátil. La primera y segunda versiones utilizan “vender en corto”, una expresión común en español de América Latina. Sin embargo, la tercera versión opta por “venderlas en descubierto”, una variante más utilizada en España y en textos técnicos de corte académico. Ambas opciones son correctas, pero “vender en descubierto” puede ser más apropiada en contextos formales o regulatorios.

En conclusión, la tercera versión es la más precisa en términos técnicos y la más adecuada para una publicación académica o financiera, ya que emplea “anuncios de resultados” en lugar de “anuncio de ganancias”, introduce la relación con la operativa bursátil mediante “en función de” y utiliza “vender en descubierto”, una expresión más formal en textos especializados. No obstante, en un contexto más general o dirigido a operadores del mercado, “vender en corto” podría ser una alternativa válida y de uso más extendido en América Latina.

#### Ejemplo 5:

Source:	Las diferencias de cambio incorporadas al activo actuaron como un adelantamiento del reconocimiento de las variaciones en el poder adquisitivo de la moneda.
IA:	The exchange differences incorporated into the asset acted as an advance recognition of changes in the currency's purchasing power.
Traductor sin especialización	The Exchange differences included in assets were an advanced recognition of the loss of purchasing power.
Traductor especializado	Exchange differences were capitalized before the recognition of changes in the currency's purchasing power, thus anticipating their impact.

Este análisis se centra en la traducción de un fragmento relacionado con el reconocimiento contable de diferencias de cambio y su impacto en el poder adquisitivo de la moneda. La evaluación de las distintas versiones permite identificar el grado de precisión terminológica, la coherencia conceptual y la naturalidad del discurso en inglés.

Uno de los principales términos a evaluar en esta comparación es *diferencias de cambio*, traducido en todas las versiones como *exchange differences*. Esta elección es acertada, ya que en terminología contable *exchange differences* es el término estándar en inglés para referirse a variaciones derivadas

del tipo de cambio en activos y pasivos expresados en moneda extranjera.

Sin embargo, la IA traduce *incorporadas al activo* como *incorporated into the asset*, una construcción que, aunque gramaticalmente correcta, es demasiado literal y no refleja con precisión el proceso contable. La primera versión humana utiliza *included in assets*, lo cual es una mejora sutil pero mantiene cierta ambigüedad. La segunda versión reformula el concepto con *were capitalized*, lo que sugiere de manera más clara que las diferencias de cambio fueron registradas dentro del valor del activo, lo que en terminología contable es una representación más fiel del concepto de capitalización de diferencias de cambio.

Otro punto clave es la traducción de *actuaron como un adelantamiento del reconocimiento*, que la IA traduce como *acted as an advance recognition*. Aunque esta versión es comprensible, no es un uso natural en inglés financiero. La primera versión humana mejora esta traducción al expresar *were an advanced recognition*, lo cual es más idiomático pero sigue sin plasmar plenamente la idea subyacente. En cambio, la segunda versión reformula la oración y clarifica la acción con *thus anticipating their impact*, lo que comunica mejor la función de las diferencias de cambio en la contabilidad del activo.

El concepto de variaciones en el poder adquisitivo de la moneda es traducido correctamente en todas las versiones como *changes in the currency's purchasing power*. Sin embargo, la primera versión humana introduce *loss of purchasing power*, lo cual modifica el significado original, ya que la frase en español no especifica una pérdida, sino un cambio, que podría ser positivo o negativo. Este es un claro ejemplo de una traducción errónea por interpretación incorrecta.

Las traducciones de la IA y la primera versión humana siguen de manera demasiado cercana la estructura del español, lo que las hace menos fluidas en inglés técnico. En cambio, la segunda versión del traductor humano reformula la oración de manera más natural y precisa al expresar *Exchange differences were capitalized before the recognition of changes in the currency's purchasing power, thus anticipating their impact*. Esta reformulación no solo mejora la claridad, sino que también plasma con mayor precisión el proceso contable descrito en el original.

El análisis muestra que, si bien la traducción generada por IA es correcta en términos gramaticales, peca de literalidad y rigidez, lo que impide una interpretación natural del concepto contable en inglés. La primera versión humana presenta mejoras, pero introduce una alteración en el significado al sugerir que hubo una “pérdida de poder adquisitivo” en lugar de un simple cambio. La segunda versión es, sin duda, la mejor de todas, ya que reformula el contenido con mayor precisión, fluidez y terminología contable adecuada, utilizando *capitalized* en lugar de *incorporated*, y expresando la idea de anticipación con *thus anticipating their impact* en lugar de traducciones más mecánicas.

Este caso demuestra que, en traducciones de textos contables y financieros, el conocimiento técnico es fundamental para evitar interpretacio-

nes incorrectas o traducciones demasiado literales que pueden desvirtuar el mensaje original.

### Ejemplo 6:

Source:	Las mencionadas sociedades tributarán el Impuesto sobre los Bienes Personales en calidad de responsable sustituto de sus integrantes, teniendo que presentar la declaración jurada de Bienes Personales Acciones y Participaciones Societarias a la fecha que determine la Administración Federal de Ingresos Públicos.
IA:	The aforementioned companies will pay the Personal Assets Tax as a substitute responsible party for their members, and must file the sworn statement of Personal Assets for Shares and Corporate Interests by the date determined by the Federal Administration of Public Revenue.
Traductor sin especialización	The aforementioned companies will pay the Personal Assets Tax as a substitute responsible party for their members, and must file the statement of Personal Assets Tax on Shares and Corporate Interests by the date determined by the Federal Administration of Public Revenue.
Traductor especializado	The above-mentioned companies, acting as surrogate taxpayers, are required to file the Personal Assets Tax Return and pay the Personal Assets Tax liability on behalf of their members in respect of any shares or equity interests held by such members as of the date established by the Argentine Federal Revenue Administration.

En este análisis comparativo de traducciones, se evalúa la precisión terminológica, la fidelidad al texto original y la claridad del mensaje en inglés en el ámbito tributario argentino. El fragmento en español hace referencia a la obligación de ciertas sociedades de tributar el Impuesto sobre los Bienes Personales en calidad de responsables sustitutos y de presentar la declaración jurada correspondiente en la fecha establecida por la Administración Federal de Ingresos Públicos (AFIP).

Las dos primeras traducciones en inglés son idénticas y representan una adaptación bastante literal del texto original. La expresión *substitute responsible party* para “responsable sustituto” es comprensible, pero no es la terminología más utilizada en inglés para referirse a esta figura. En contextos fiscales, el término más adecuado sería *surrogate taxpayer* o *withholding agent*, dependiendo del sistema impositivo específico. La tercera traducción

resuelve este problema al utilizar *surrogate taxpayers*, una opción más aliñada con la terminología tributaria en inglés.

Otro aspecto relevante es la traducción de “declaración jurada de Bienes Personales Acciones y Participaciones Societarias”. Las traducciones de la IA y del traductor sin especialización intentan transmitir la idea general, pero presentan estructuras que pueden resultar poco naturales en inglés y no reflejan con precisión la terminología utilizada en sistemas fiscales anglosajones. Por ejemplo, la expresión *sworn statement of Personal Assets for Shares and Corporate Interests* no es común en la práctica fiscal en inglés. Por otro lado, la traducción del traductor especializado utiliza *Personal Assets Tax Return*, una expresión más acorde con la terminología estándar en sistemas tributarios de habla inglesa. Además, esta versión incluye la frase *and pay the Personal Assets Tax liability on behalf of their members*, lo que aclara que la obligación tributaria recae sobre la sociedad en calidad de responsable sustituto, y no directamente sobre los socios. Esta adición refuerza la precisión y claridad de la traducción, alineándose con las prácticas fiscales en las que las sociedades actúan como responsables sustitutos al liquidar el impuesto sobre los bienes personales correspondientes a las acciones o participaciones de sus socios o accionistas.

Un punto que merece atención es la traducción de “Administración Federal de Ingresos Públicos”. La primera y segunda traducciones la convierten en *Federal Administration of Public Revenue*, que es una traducción literal y comprensible, aunque no necesariamente reconocida internacionalmente. La tercera versión opta por *Argentine Federal Revenue Administration*, una opción más precisa que incluye la referencia al país, lo que facilita la comprensión para lectores que no estén familiarizados con el sistema fiscal argentino.

En términos de fidelidad al original, la tercera versión es la mejor lograda, ya que no solo mantiene la información clave, sino que además la adapta al lenguaje técnico tributario del inglés sin caer en una literalidad excesiva. La inclusión de *in respect of any shares or equity interests held by such members* es un acierto, ya que especifica la naturaleza del activo gravado sin modificar el significado original del texto en español.

Este análisis demuestra que una traducción efectiva en el ámbito tributario no solo debe respetar la terminología específica del sistema fiscal de origen, sino también adaptarse a los conceptos y términos utilizados en el idioma de destino. La tercera traducción logra este equilibrio al evitar traducciones literales innecesarias, mejorar la precisión terminológica y estructurar la información de manera más natural en inglés.

**Ejemplo 7:**

Source:	En los primeros cien días de gobierno logró que las cuentas públicas registren superávit fiscal primario dos meses consecutivos, algo que no ocurría desde agosto de 2019. Sin embargo, los economistas plantean que el ajuste del gasto es una buena señal pero fue más por efecto licuadora que de motosierra, y afirman que no es sostenible esa dinámica en los próximos meses.
IA:	In the first 100 days of government, they recorded a primary fiscal surplus in the public accounts for two consecutive months, something that had not happened since August 2019. However, economists state that the spending adjustment is a good sign but it was more due to the blender than to the chainsaw, and affirm that this dynamic is not sustainable in the coming months.
Traductor sin especialización	In its first 100 days, the government recorded a primary surplus in its public accounts for two consecutive months, which had not happened since August 2019. However, according to economists, although this spending adjustment is a good sign it was more due to a blender effect than to a chainsaw effect, which situation is not sustainable over the coming months.
Traductor especializado	In the first hundred days of government, public accounts recorded a primary fiscal surplus for two consecutive months, something not seen since August 2019. However, economists argue that while the spending adjustment is a good sign, it was more due to inflationary erosion rather than significant cuts, and they assert that this dynamic is not sustainable in the coming months.

En este análisis comparativo se evalúan tres versiones en inglés de un fragmento económico en español, con especial atención a la precisión terminológica, la claridad del mensaje y la adaptación idiomática. El texto original describe el desempeño fiscal de un gobierno en sus primeros cien días y analiza la naturaleza del ajuste del gasto público, diferenciando entre dos mecanismos coloquialmente denominados efecto licuadora y efecto motosierra.

Uno de los aspectos más relevantes es la traducción del concepto de superávit fiscal primario. La primera versión opta por *primary fiscal surplus*, mientras que la segunda lo traduce simplemente como *primary surplus*.

Aunque ambas opciones son correctas, *primary fiscal surplus* mantiene una mayor precisión técnica y es más común en literatura económica.

Otro punto clave es la traducción del concepto coloquial efecto licuadora frente al efecto motosierra, que en el discurso económico argentino se refiere a dos formas de reducir el gasto público: la licuadora representa la reducción del gasto en términos reales debido a la inflación, mientras que la motosierra alude a recortes directos y drásticos del gasto nominal. La primera y segunda traducciones intentan mantener la metáfora al referirse a *the blender* y a *blender effect*, pero sin el contexto adecuado, estas expresiones pierden claridad para un lector no familiarizado con la jerga económica argentina. La tercera traducción, en cambio, opta por una reformulación conceptual al decir *it was more due to inflationary erosion rather than significant cuts*, lo que transmite con precisión el significado sin depender de metáforas que pueden no ser comprendidas en otros contextos lingüísticos.

En cuanto a la traducción de los economistas plantean, la primera versión usa *economists state*, una opción gramaticalmente correcta pero algo rígida. La segunda versión introduce *according to economists*, lo que genera una transición más fluida. La tercera versión emplea *economists argue*, que es una elección acertada, ya que sugiere que los economistas están presentando una postura argumentativa más que simplemente afirmando un hecho.

Finalmente, las tres versiones coinciden en la traducción de que no es sostenible esa dinámica en los próximos meses, utilizando *this dynamic is not sustainable in the coming months*. La segunda versión introduce *which situation is not sustainable*, lo cual es menos natural y puede resultar en una construcción innecesariamente enrevesada en inglés. La tercera versión mantiene la estructura más natural y clara.

En conclusión, la tercera traducción es la más precisa y efectiva, ya que no solo respeta la terminología técnica adecuada, sino que también reformula expresiones coloquiales para hacerlas comprensibles a un público angloparlante sin perder el significado original. La elección de *inflationary erosion rather than significant cuts* es particularmente acertada.

### **La redefinición del rol del traductor en la era de la IA generativa**

Como se explicó en la primera parte de este artículo, la evolución de la tecnología de traducción puede entenderse en tres grandes etapas: los sistemas basados en reglas, los modelos de traducción automática estadística (SMT, por sus siglas en inglés) y la traducción automática neuronal (NMT). Los primeros dependían de una programación explícita para la conversión de frases, lo que los hacía altamente estructurados pero rígidos. Los modelos SMT, por su parte, introdujeron el aprendizaje basado en corpus de traducciones previas, mejorando la precisión, aunque con deficiencias en la coherencia textual.

Con la llegada de la traducción automática neuronal y el aprendizaje profundo, la capacidad de los sistemas para generar traducciones más naturales

y contextualmente adecuadas se incrementó de manera significativa. Los modelos de lenguaje de gran escala (LLM), como GPT-4 y NLLB-200, han llevado esta evolución aún más lejos al permitir la generación de contenido que no solo es preciso en términos léxicos y sintácticos, sino que también capta matices culturales y estilísticos de los textos (Vincent, 2021).

En este contexto, el traductor ya no se posiciona como un mero intermediario lingüístico, sino como un mediador estratégico que supervisa, valida y optimiza las producciones generadas por la IA.

### **Un nuevo perfil profesional: el traductor como supervisor y estratega**

En un entorno donde la inteligencia artificial desempeña un rol central en la producción textual, la función del traductor asume un conjunto más complejo de tareas. Entre las competencias emergentes que este nuevo paradigma exige se encuentran:

#### **Posedición y revisión crítica**

##### **La posedición en la era de la inteligencia artificial generativa: un proceso crítico para la calidad de la traducción**

La posedición, definida como la revisión y corrección de textos generados por sistemas de traducción automática (TA), se ha convertido en una actividad esencial en la industria de la traducción. Aunque la IA generativa ha hecho grandes avances en términos de fluidez y coherencia, su desempeño sigue siendo imperfecto, especialmente en lo que respecta a la adecuación cultural, la precisión terminológica y la fidelidad al texto original. Los modelos de IA, como los de traducción automática neuronal (NMT, por sus siglas en inglés), son capaces de producir textos gramaticalmente correctos, pero a menudo carecen de una comprensión semántica profunda y pueden introducir errores sutiles que solo un ojo experto puede detectar (Mossop, 2021).<sup>12</sup> La posedición, por tanto, no es un mero acto de corrección superficial, sino un proceso que exige un análisis minucioso del texto para garantizar que el mensaje transmitido sea preciso y adecuado a su contexto.

Uno de los principales desafíos de la posedición radica en la identificación y corrección de errores contextuales. Si bien los modelos de IA pueden generar traducciones técnicamente correctas, estas no siempre reflejan las intenciones del autor original o no logran captar matices lingüísticos esenciales. Por ejemplo, en textos jurídicos o médicos, una mala interpretación de un término técnico puede acarrear consecuencias graves, desde problemas legales hasta riesgos para la salud del paciente (Tinsley et al., 2020).

---

<sup>12</sup> En esta obra, Mossop profundiza en los procesos de revisión y edición de textos en traducción, incluidas las tareas de posedición y hace hincapié en la importancia de la intervención humana para detectar y corregir errores que las herramientas automatizadas no pueden identificar por sí mismas.

Además, la posesición no solo se limita a detectar errores gramaticales o sintácticos; también exige intervenir en aspectos estilísticos para asegurar que la traducción mantenga coherencia con el tono y registro del texto fuente. Este aspecto adquiere especial relevancia en traducciones de contenido publicitario o literario, donde la expresividad y el estilo juegan un papel clave en la recepción del mensaje.

Otro aspecto crucial de la posesición es la mitigación de sesgos inherentes a los modelos de IA. Como han señalado estudios recientes, los sistemas de traducción automática pueden reflejar e incluso amplificar sesgos de género, culturales y raciales presentes en los datos con los que fueron entrenados (Bender et al., 2021). Por ejemplo, se ha demostrado que algunos sistemas de traducción tienden a asociar profesiones como “doctor” con el género masculino y “enfermera” con el femenino, perpetuando estereotipos culturales. La posesición, en este sentido, se convierte en una práctica de corrección ética, en la que el traductor no se limita a ajustar errores técnicos, sino que garantiza que la traducción refleje principios de equidad e inclusión.

Para abordar estos desafíos, los traductores deben desarrollar un enfoque crítico hacia la IA y complementar su conocimiento lingüístico con una sensibilidad hacia cuestiones socioculturales y éticas. El problema de los sesgos en la inteligencia artificial es una consecuencia directa de la forma en que se entrena a los modelos de IA. Los sistemas de traducción automática neuronal (NMT), incluidos los modelos de lenguaje de gran escala (LLM) como GPT-4 o NLLB-200, aprenden a traducir a partir de grandes volúmenes de datos extraídos de textos existentes en la web, libros, artículos y otros documentos escritos por humanos. Sin embargo, estos datos no son neutrales; reflejan patrones lingüísticos, culturales y sociales que pueden contener prejuicios implícitos. Como resultado, los modelos de IA pueden perpetuar y amplificar estos sesgos en sus traducciones (Bender et al., 2021). Por ejemplo en *the lawyer met with the secretary to discuss the case*, la IA tiende a traducirlo como “el abogado se reunió con la secretaria para analizar el caso”, asumiendo que la secretaria es mujer y el abogado hombre, y muchas veces ignora referencias anteriores en el propio texto de donde surgiría que no es así.

Estos sesgos pueden afectar la calidad y la precisión de una traducción de múltiples maneras. Primero, pueden distorsionar el significado original del texto al introducir connotaciones que no estaban presentes en la fuente. En contextos profesionales, como documentos legales, médicos o financieros, esto puede llevar a interpretaciones erróneas que comprometan la fidelidad del mensaje original. Además, los sesgos pueden influir en la percepción cultural de un texto al reforzar estereotipos o representaciones unilaterales de ciertas comunidades. Por ejemplo, en traducciones de noticias o literatura, una IA que se fue entrenada en su mayor parte con textos en inglés podría tender a adaptar expresiones o estructuras propias de ese idioma sin respetar las normas discursivas de la lengua meta, lo que menoscaba la autenticidad y adecuación cultural de la traducción (Costa-jussà & Hard-

meier, 2019).<sup>13</sup> Este fenómeno puede ser particularmente problemático en traducciones de contenido político o de sensibilidad social, donde una redacción sesgada puede modificar la intención original del autor.

Para mitigar el impacto de los sesgos en la traducción automática, es crucial que los traductores desarrollen un enfoque crítico al trabajar con herramientas de IA. La posesición no solo implica corregir errores gramaticales y terminológicos, sino también revisar de manera consciente la presencia de sesgos en el texto traducido. Además, las empresas que desarrollan modelos de IA deben implementar estrategias para reducir la propagación de prejuicios en sus sistemas, como el uso de conjuntos de datos más diversos y equilibrados, el ajuste fino de modelos con datos culturalmente representativos y la incorporación de mecanismos de control para detectar y corregir resultados tendenciosos (Mitchell et al., 2020). Sin embargo, aunque estos avances tecnológicos pueden mejorar la equidad de las traducciones generadas por IA, la intervención humana sigue siendo indispensable para garantizar que los textos sean precisos, inclusivos y culturalmente apropiados.

Finalmente, la posesición eficiente requiere un entrenamiento especializado y un dominio de herramientas tecnológicas avanzadas. Los traductores deben familiarizarse con metodologías de posesición que les permitan optimizar su tiempo sin comprometer la calidad del producto final. Existen dos enfoques principales en este proceso: la posesición ligera y la posesición completa (TAUS, 2019).<sup>14</sup> La posesición ligera implica correcciones mínimas para hacer que el texto sea comprensible y adecuado para su propósito inmediato, mientras que la posesición completa busca mejorar la fluidez, precisión y naturalidad del texto hasta alcanzar un nivel comparable al de una traducción humana de alta calidad. Dependiendo del tipo de contenido y del contexto en el que se utilice, los traductores deben elegir el enfoque adecuado para equilibrar eficiencia y calidad. Además, la posesición debe ir acompañada de una evaluación constante de las herramientas de IA utilizadas, ya que el desempeño de los modelos puede variar en función de la combinación de idiomas, el dominio temático y la estructura del texto. Por ello, la posesición no es solo una fase de corrección, sino un proceso estratégico en el que el traductor debe tomar decisiones informadas sobre el uso y las limitaciones de la IA en cada proyecto.

En conclusión, la posesición es una competencia indispensable en la nueva era de la traducción asistida por IA. Si bien la inteligencia artificial generativa ha elevado la calidad de la traducción automática, sigue siendo necesaria la intervención humana para corregir errores contextuales, eliminar sesgos y asegurar la coherencia estilística y terminológica. La capacita-

13 En este trabajo, las autoras analizan cómo los sesgos, en particular los de género, se filtran en los procesos de traducción automática y resaltan la importancia de estrategias de mitigación para preservar la precisión y la autenticidad cultural en los textos traducidos.

14 En estas directrices, la organización analiza en detalle los distintos niveles de posesición (ligera y completa), ofreciendo recomendaciones para optimizar el trabajo de los profesionales de la traducción y asegurar la calidad de los textos resultantes.

ción en posesión no solo permitirá a los traductores mejorar su productividad, sino que también garantizará que los textos traducidos mantengan los más altos estándares de precisión, ética y adecuación cultural. A medida que la IA continúe evolucionando, la posesión se consolidará como una de las habilidades más valiosas en la industria de la traducción, redefiniendo el rol del traductor como un profesional estratégico en la gestión de la comunicación multilingüe.

### **Evaluación del rendimiento de los modelos de IA**

#### **El traductor como evaluador de la calidad en traducción automática: nuevas competencias en la era de la IA**

La capacidad de analizar la confiabilidad de los modelos de inteligencia artificial es fundamental para los traductores, ya que cada herramienta posee fortalezas y debilidades específicas que impactan en la calidad del texto final. No todas las IA son igualmente eficaces para todos los tipos de textos: algunos modelos destacan en traducción técnica y científica, mientras que otros tienen mejores resultados en textos literarios o creativos. Para evaluar estos sistemas, los traductores deben poder identificar las limitaciones de cada modelo, detectar patrones de error recurrentes y decidir estratégicamente cuándo y cómo intervenir manualmente para mejorar el resultado final.

En este contexto, el conocimiento de métricas de evaluación de traducción automática es esencial. Entre las más utilizadas se encuentran BLEU (*Bilingual Evaluation Understudy*) y TER (*Translation Edit Rate*). BLEU compara la traducción generada por IA con una referencia humana, midiendo la similitud mediante la coincidencia de n-gramas.<sup>15</sup> Una mayor puntuación BLEU sugiere una mayor fidelidad en términos de estructura y elección léxica. Sin embargo, BLEU tiene limitaciones, ya que no evalúa la fluidez ni la adecuación contextual, lo que puede ser problemático en textos donde la precisión semántica es más importante que la correspondencia literal.

Por otro lado, TER mide la cantidad de ediciones necesarias para convertir una traducción automática en una de calidad humana. Una menor puntuación TER indica un menor esfuerzo de posesión requerido. Esta métrica es especialmente útil para evaluar la eficiencia de la posesión y determinar si el uso de una herramienta de traducción automática realmente aporta beneficios en términos de tiempo y calidad. No obstante, al igual que BLEU, TER no es infalible: se centra en la cantidad de cambios necesarios sin considerar aspectos como la fluidez del discurso o la naturalidad de la traducción en el idioma de destino.

---

<sup>15</sup> Su cálculo se basa en la coincidencia de **n-gramas**, que son secuencias contiguas de *n* palabras dentro de un texto. Por ejemplo, en la frase “la inteligencia artificial avanza rápido”, los **bigramas** (n=2) serían “la inteligencia”, “inteligencia artificial”, “artificial avanza”, y “avanza rápido”. BLEU analiza cuántos de estos segmentos de palabras en la traducción automática coinciden con la referencia humana, asignando una puntuación mayor a traducciones con estructuras y elecciones léxicas más similares.

Además de estas métricas tradicionales, han surgido nuevos enfoques que buscan mejorar la evaluación de la calidad de la traducción automática. Por ejemplo, COMET<sup>16</sup> (*Crosslingual Optimized Metric for Evaluation of Translation*) utiliza modelos de aprendizaje profundo para evaluar la calidad de la traducción, considerando tanto la similitud con una referencia humana como la adecuación contextual y semántica. Estas métricas avanzadas representan un salto cualitativo en la evaluación automática de la traducción, pero su implementación aún requiere la validación de expertos humanos para garantizar que los resultados sean representativos de la calidad lingüística.

El dominio de estas métricas permite a los traductores tomar decisiones más informadas sobre qué herramientas de traducción automática utilizar según el tipo de texto y los requisitos específicos del proyecto. Un traductor capacitado en evaluación de calidad podrá determinar si un modelo neuronal es más adecuado para textos jurídicos debido a su precisión terminológica, mientras que otro puede ser suficiente para la comprensión general de textos informales. Asimismo, un sólido conocimiento de estas herramientas permite identificar casos en los que la traducción automática no es recomendable, como en documentos con alta carga de creatividad, textos muy especializados o en contextos donde la precisión absoluta es crucial.

Más allá del uso de métricas automáticas, los traductores también deben desarrollar habilidades críticas para la evaluación cualitativa de las traducciones generadas por IA. Esto implica la capacidad de identificar problemas de coherencia discursiva, adecuación cultural y fluidez estilística, aspectos que no siempre pueden ser cuantificados con precisión por los algoritmos actuales. En este sentido, la combinación de herramientas de evaluación automática con la supervisión experta de traductores humanos es la mejor estrategia para garantizar la calidad de las traducciones en un entorno donde la inteligencia artificial juega un papel cada vez más predominante.

---

16 COMET utiliza modelos de redes neuronales preentrenadas para analizar la relación entre el texto de origen, la traducción generada por IA y, cuando está disponible, una traducción de referencia humana. Su evaluación se basa en tres aspectos clave:

- 1- Adecuación (Adequacy): Qué tan bien la traducción refleja el significado del texto original.
- 2- Fluidez (Fluency): Qué tan natural y comprensible es la traducción en el idioma meta.
- 3- Precisión semántica (Semantic Accuracy): Si la traducción mantiene relaciones y significados correctos dentro del contexto.

COMET ha mostrado una fuerte correlación con la evaluación humana, lo que lo convierte en una métrica más confiable para medir la calidad real de las traducciones, especialmente en casos donde el significado es más relevante que la coincidencia exacta de palabras.

## **Adaptación al entorno híbrido de trabajo**

### **La traducción en un entorno híbrido: equilibrio entre inteligencia artificial y labor humana**

La incorporación de la inteligencia artificial en la industria de la traducción ha dado lugar a un entorno híbrido en el que la interacción entre humanos y máquinas es permanente. Lejos de reemplazar al traductor, la IA se ha convertido en una herramienta estratégica que optimiza el flujo de trabajo y agiliza la producción de contenidos multilingües. Sin embargo, su implementación requiere una adaptación crítica por parte del traductor, quien debe aprender a integrar estas herramientas sin comprometer la calidad, la creatividad y la precisión de los textos. Esto significa que la traducción automática no puede verse como una solución universal, sino como un recurso complementario que debe ser utilizado con criterio profesional (Bowker, 2019).

Uno de los principales desafíos del traductor en este nuevo paradigma es determinar en qué casos la inteligencia artificial puede ser utilizada de manera confiable y en qué situaciones es imprescindible una intervención humana. En textos que exigen creatividad, como literatura, publicidad o subtítulo artístico, la IA tiende a generar resultados mecánicos y predecibles, ya que los modelos de traducción automática neuronal (NMT) se basan en patrones estadísticos y no en una comprensión genuina del significado y la intención del autor (Toral & Way, 2018). La traducción de metáforas, juegos de palabras y referencias culturales sigue representando un obstáculo para la IA, ya que estos elementos requieren interpretación y adaptación en lugar de una mera conversión léxica. En estos casos, la labor del traductor es fundamental para garantizar la naturalidad y el impacto del mensaje en la lengua meta.

Por otro lado, en ámbitos técnicos y repetitivos, como la traducción de manuales, documentación de software o textos legales estandarizados, la inteligencia artificial puede ser de gran utilidad para mejorar la productividad. Los modelos de IA han demostrado ser eficaces en la traducción de textos muy estructurados y con terminología bien definida, especialmente cuando se combinan con bases de datos terminológicas y memorias de traducción (Mossop, 2021). En este tipo de documentos, el uso de traducción automática seguida de una revisión humana puede reducir significativamente los tiempos de entrega sin comprometer la calidad del producto final.

Otro factor clave en este entorno híbrido es la capacidad del traductor para tomar decisiones estratégicas sobre el nivel de intervención que requiere cada texto. No todos los proyectos necesitan el mismo grado de revisión, y el traductor debe evaluar si una posesición ligera es suficiente o si es necesaria una posesición completa, donde puede ser necesario reescribir parte del texto para garantizar su coherencia y fluidez. La capacidad de tomar estas decisiones de manera eficiente depende del conocimiento técnico y lingüístico del traductor y de su familiaridad con las herramientas de IA y los

modelos de traducción automática disponibles. Y como si esto fuera poco, esta capacidad de evaluación tiene importantes implicaciones económicas, ya que una gestión eficiente de los recursos tecnológicos puede mejorar la competitividad del traductor en el mercado laboral.

En conclusión, la traducción en un entorno híbrido exige del traductor una mentalidad flexible y estratégica. La clave del éxito en este nuevo paradigma radica en el equilibrio entre la automatización y la creatividad humana, ya que permite a los traductores aprovechar lo mejor de ambos mundos sin comprometer la esencia de su labor.

### **Consideraciones éticas y responsabilidad del traductor**

El uso de inteligencia artificial en la traducción ha introducido múltiples beneficios en términos de eficiencia y accesibilidad, pero también plantea desafíos éticos que requieren una reflexión profunda. Entre los problemas más relevantes se encuentra el sesgo inherente a los modelos de IA, que puede producir traducciones discriminatorias o inapropiadas para el contexto cultural. Como estos modelos se entrenan con vastas cantidades de datos disponibles en la web y en documentos escritos por humanos, inevitablemente absorben los patrones lingüísticos y culturales dominantes, incluidos estereotipos, prejuicios y sesgos de género, raza o clase social (Bender et al., 2021). Estos sesgos pueden reflejar y perpetuar desigualdades existentes, lo que convierte al traductor en una figura clave para detectar y corregir estos problemas antes de que impacten en la comunicación intercultural.

Otro desafío ético fundamental en la traducción con IA es la protección de la confidencialidad de los datos. Con la evolución de las herramientas de traducción automática, muchas han migrado a plataformas en la nube, lo que implica que los textos que ingresan los usuarios pueden ser almacenados, analizados y utilizados para mejorar los modelos. Sin lugar a dudas, esta práctica permite optimizar el desempeño de los sistemas; sin embargo, también plantea riesgos significativos en cuanto a la privacidad y seguridad de la información. En ámbitos sensibles como la medicina, el derecho o las finanzas, el uso de herramientas de traducción basadas en IA puede exponer datos confidenciales a terceros sin el conocimiento o consentimiento del usuario (Moorkens, 2022). Por ejemplo, traducir un informe médico a través de una plataforma en línea sin cifrado adecuado podría resultar en la filtración de información protegida por normativas de privacidad como el GDPR en Europa o la HIPAA en Estados Unidos.

Para abordar estas preocupaciones, los traductores deben ser conscientes de las implicaciones legales del uso de herramientas de IA y adoptar buenas prácticas en la gestión de datos. En primer lugar, es fundamental verificar si las plataformas utilizadas almacenan o reutilizan el contenido ingresado para entrenamiento de modelos. Herramientas que garantizan la eliminación automática de los datos tras su procesamiento o que permiten el uso de servidores locales son preferibles en entornos donde la seguridad es priori-

taria. En este sentido, los *Small Language Models* (SLM) emergen como una solución eficiente y segura para procesar traducciones dentro de infraestructuras privadas sin exponer información sensible a servicios externos.

A diferencia de los *Large Language Models* (LLM), que suelen operar en la nube y pueden implicar riesgos de privacidad al enviar datos a servidores de terceros, los SLM pueden implementarse en servidores locales o entornos cerrados, lo que garantiza un mayor control sobre los datos corporativos.<sup>17</sup> Gracias a esta característica, son particularmente adecuados para sectores donde la confidencialidad es crítica, como el financiero, el legal o el gubernamental.

Además, los SLM pueden personalizarse con terminología específica de la organización y así generar traducciones más precisas y adaptadas a las necesidades internas<sup>18</sup>. Aunque tienen una capacidad de procesamiento menor en comparación con los modelos de IA de gran escala, su eficiencia en tareas específicas y su capacidad de operar sin conexión los convierten en una opción estratégica para empresas que priorizan la seguridad de la información.

En conclusión, la integración de la inteligencia artificial en la traducción plantea dilemas éticos complejos que no pueden soslayarse. La responsabilidad del traductor en este nuevo paradigma trasciende la revisión lingüística: debe actuar como un garante de la equidad y la seguridad en la comunicación multilingüe. La mitigación de sesgos en la IA requiere una intervención humana activa para evitar que los modelos de traducción refuercen estereotipos y exclusiones. Al mismo tiempo, la protección de la confidencialidad de los datos exige una gestión cuidadosa de las herramientas tecnológicas para asegurar que su uso cumpla con los estándares éticos y legales adecuados. A medida que la tecnología sigue evolucionando, el papel del traductor como mediador crítico entre la IA y la sociedad será cada vez más crucial en la construcción de un ecosistema de traducción más justo, seguro y confiable.

Para concluir esta sección, la inteligencia artificial generativa ha transformado la industria de la traducción y redefinido el papel del traductor en un entorno cada vez más automatizado. Lejos de ser reemplazado por la tecnología, el traductor moderno se consolida como un profesional que supervisa, mejora y valida las traducciones generadas por IA.

Frente a este nuevo escenario, es imperativo que la formación en traducción se adapte a estas demandas e incorpore estrategias pedagógicas que preparen a los futuros profesionales para trabajar en sinergia con la IA. Solo así se garantizará que la tecnología no desplace al traductor, sino que lo empodere como un actor clave en la producción de contenido multilingüe

---

<sup>17</sup> Los SLM pueden ejecutarse en infraestructuras locales, conocidas en inglés como infraestructuras on-premise, o en servidores internos protegidos, lo que reduce el riesgo de exposición de datos en redes externas.

<sup>18</sup> A través de técnicas de entrenamiento personalizado, es posible optimizar los SLM para industrias específicas, y así mejorar la calidad y coherencia de las traducciones especializadas

de calidad. Más allá de los desafíos éticos y legales, la irrupción de modelos generativos también interpela los fundamentos epistemológicos de la práctica traductológica. En un entorno donde los textos son producidos parcial o totalmente por sistemas automatizados, se desdibujan los límites tradicionales entre autoría, mediación y revisión. El traductor ya no trabaja exclusivamente sobre un original humano, sino que interviene sobre contenidos generados por algoritmos que carecen de intencionalidad, contexto y responsabilidad en sentido estricto. Esta transformación exige repensar qué significa ser fiel a un texto cuando su fuente no es una voz consciente ni culturalmente situada. La adecuación cultural, por su parte, deja de ser un simple ajuste pragmático para convertirse en un proceso activo de reconstrucción de sentido. En este nuevo escenario, el traductor actúa como un curador textual: interpreta, reconfigura y valida discursos que, aunque formalmente correctos, requieren intervención humana para adquirir pertinencia, coherencia y profundidad comunicativa.

### **Adecuación de los planes de estudio: formación integral para los traductores del futuro**

La formación de traductores en la era digital requiere una adecuación de los planes de estudio que equilibre la enseñanza de competencias tradicionales con la incorporación de habilidades tecnológicas. Es fundamental que los futuros traductores continúen desarrollando sus habilidades lingüísticas y traductológicas, ya que su capacidad para corregir, supervisar y mejorar las traducciones generadas por IA dependerá en gran medida de su dominio de la práctica de traducción. Una posesición efectiva dista mucho de ser una tarea mecánica de corrección superficial; antes bien, es un proceso de toma de decisiones informadas que requiere sensibilidad lingüística, conocimiento cultural y comprensión de los matices del discurso. Sin una sólida formación en traducción, los profesionales carecerán del criterio necesario para intervenir en los textos generados por IA de manera efectiva y garantizar que la calidad final sea equiparable a la de una traducción humana de alto nivel.

Por lo tanto, la adecuación de los planes de estudio no debe reducirse a la simple incorporación de asignaturas sobre tecnología, sino que debe diseñarse con un enfoque integrador que mantenga la traducción como eje central de la formación. La enseñanza de la posesición, por ejemplo, debe ir acompañada de una práctica intensiva en traducción convencional, de modo que los estudiantes comprendan los errores más frecuentes de la traducción automática a partir de su propia experiencia en la resolución de problemas traductológicos. De igual manera, el estudio de métricas como BLEU y TER debe complementarse con el desarrollo de habilidades críticas para la evaluación cualitativa de traducciones, pues los sistemas automatizados de medición no siempre reflejan aspectos clave como la fluidez estilística, la adecuación cultural y la coherencia discursiva.

Además, la formación en ética profesional y seguridad de los datos debe ocupar un lugar central en la preparación de los nuevos traductores. El uso de herramientas de IA conlleva riesgos en términos de sesgo y privacidad, por lo que los estudiantes deben recibir formación para poder detectar problemas de discriminación en la traducción automática y en la correcta gestión de documentos confidenciales en entornos digitales. Un traductor que no domine los principios fundamentales de su profesión difícilmente podrá evaluar la calidad de una traducción generada por IA o aplicar criterios éticos en su corrección.

En conclusión, la adecuación de los planes de estudio es una necesidad ineludible para preparar a los traductores ante los desafíos de la inteligencia artificial. Este proceso no debe interpretarse como una sustitución de la enseñanza de la traducción por la formación en herramientas tecnológicas, sino como una actualización del perfil profesional en la que la práctica traductora sigue siendo el pilar fundamental. Solo aquellos traductores que dominen su oficio podrán hacer un uso inteligente de la IA, corrigiendo sus deficiencias y asegurando que la tecnología funcione como un complemento de la labor humana en lugar de un sustituto deficiente. La clave del éxito en la formación de los nuevos traductores radica, por lo tanto, en la integración equilibrada de conocimientos tecnológicos y traductológicos.

### **El fin del traductor generalista: por qué la especialización es clave**

Históricamente, los traductores generalistas han tenido un papel fundamental en la industria, ya que su versatilidad les permitía abordar una amplia gama de textos sin necesidad de una especialización profunda en un área específica. Sin embargo, con la llegada de la traducción automática y su mejora constante, las tareas más rutinarias y predecibles se han ido delegando progresivamente en la IA. Hoy en día, la traducción de documentos administrativos, correos electrónicos, textos informativos generales e incluso artículos de prensa de baja complejidad puede realizarse de manera eficiente con herramientas como DeepL o Google Translate. Estos modelos han reducido drásticamente la necesidad de traductores que trabajen en textos estándar, ya que las empresas y los clientes pueden obtener resultados satisfactorios sin intervención humana o con una mínima posesición (Toral & Way, 2018).

Ante este escenario, el único camino viable para el traductor es la especialización. La IA es poderosa, pero no infalible, y aún presenta deficiencias significativas en áreas que requieren un conocimiento profundo del tema, precisión terminológica extrema y capacidad de interpretación contextual. Esto ocurre en campos como el derecho, la medicina, la ingeniería, las finanzas y las ciencias sociales, donde una mala traducción puede tener consecuencias graves, desde errores legales hasta riesgos para la salud del paciente. En estos ámbitos, la posesición y supervisión de un traductor especializado es mucho más que una opción; es un imperativo.

## **Traducción especializada: el factor diferencial ante la IA**

Uno de los principales problemas de los modelos de traducción automática es su falta de comprensión conceptual. Como se ha explicado en las secciones anteriores, la IA no “piensa” ni comprende el significado de las palabras en un sentido profundo; su funcionamiento se basa en patrones estadísticos y correlaciones aprendidas a partir de grandes volúmenes de texto. Si bien esto le permite generar traducciones altamente naturales en muchos casos, cuando se enfrenta a terminología específica, documentos muy técnicos o conceptos que requieren interpretación, los errores pueden ser catastróficos (Moorkens, 2022).

Por ejemplo, en la traducción jurídica, la precisión es fundamental, ya que una mínima ambigüedad puede alterar el significado de un contrato o una ley. Los sistemas de IA pueden traducir términos legales de manera literal sin considerar las diferencias entre los sistemas jurídicos de distintos países. Un traductor especializado en derecho, en cambio, no solo domina la terminología, sino que también comprende los principios legales subyacentes y puede adaptar la traducción para que sea funcional en el contexto del ordenamiento jurídico de destino.

En el caso de la traducción médica, la situación es aún más delicada. Un error en la traducción de un informe clínico, un prospecto farmacéutico o un estudio de investigación puede comprometer la salud del paciente o llevar a interpretaciones erróneas en un diagnóstico. Los traductores que se especializan en medicina no solo conocen la terminología con precisión; además, comprenden la sintaxis y las convenciones del lenguaje especializado.

Otro caso relevante es el de la traducción financiera, donde la exactitud numérica y conceptual es esencial. Un error en la traducción de un informe contable, un contrato de inversión o un documento de auditoría puede generar pérdidas económicas o problemas legales. Los traductores especializados en finanzas comprenden no solo los términos, sino también la lógica subyacente a los estados financieros y las regulaciones del sector, lo que les permite detectar incoherencias que una máquina pasaría por alto.

### **El papel del traductor especializado en la posesición y validación de la IA**

A medida que la traducción automática va adquiriendo más precisión, la posesición se erigirá en una de las principales áreas de trabajo para los traductores humanos. Sin embargo, solo un traductor con conocimientos especializados puede llevar a cabo una posesición efectiva. Un traductor sin formación específica en un área determinada puede corregir errores superficiales de gramática o sintaxis, pero difícilmente podrá evaluar si la terminología utilizada es la correcta o si el texto refleja con precisión el contenido del original.

En este sentido, la posesición en traducción especializada no es una ta-

rea meramente mecánica, sino un proceso de validación y control de calidad. El traductor experto debe evaluar si el texto generado por la IA es adecuado, corregir errores conceptuales, verificar la coherencia terminológica y garantizar que el resultado final sea funcional en su contexto específico. Esto requiere una formación sólida en el campo de especialización y una capacidad analítica que la IA, por más avanzada que sea, no puede replicar.

### **La formación del traductor especializado: una necesidad ineludible**

Dado este panorama, los programas de formación en traducción deben evolucionar para preparar a los estudiantes en campos de especialización concretos. No basta con enseñar herramientas tecnológicas y posesión; es necesario que los futuros traductores desarrollen competencias avanzadas en disciplinas específicas. Esto implica una mayor interacción con expertos en cada campo, cursos especializados en derecho, medicina, finanzas, ingeniería u otras áreas, y la capacidad de interpretar textos más allá de su estructura lingüística.

Además, los traductores especializados deben mantenerse actualizados en su área de conocimiento, dado que la terminología y las convenciones de cada sector evolucionan constantemente. Esto puede lograrse a través de formación continua, acceso a literatura técnica y científica en el idioma de trabajo, participación en conferencias y colaboración con profesionales del sector.

### **Conclusión: El futuro es de los traductores especialistas**

El avance de la inteligencia artificial generativa (IAG) ha redefinido las dinámicas de la industria de la traducción, impactando de forma profunda tanto en la práctica profesional como en la formación académica de los traductores. A lo largo del capítulo se ha evidenciado cómo la proliferación de modelos de IA generativa ha reducido la demanda de traductores generalistas y ha puesto en el centro de la escena a profesionales de la traducción con una formación especializada. Esta tendencia no solo responde a la creciente capacidad de los sistemas automatizados para generar traducciones con un nivel aceptable de fluidez y coherencia, sino también a la necesidad de supervisión experta en sectores técnicos, jurídicos y financieros, donde los errores pueden tener consecuencias significativas.

Los casos prácticos analizados en este capítulo ilustran claramente cómo la IA generativa, aunque poderosa en términos de procesamiento lingüístico, sigue presentando limitaciones importantes en cuanto a precisión terminológica, adecuación contextual y fidelidad al mensaje original. Por ejemplo, en el caso del sector financiero, el análisis comparativo entre las traducciones generadas por IA y las realizadas por traductores especializados revela discrepancias críticas en la interpretación de conceptos económicos complejos. Términos como *tight labor markets* o *corporate debt overhang* se tradujeron de forma literal o

ambigua en el modelo de IA, mientras que los traductores con formación específica en economía supieron contextualizar adecuadamente los conceptos, captar las sutilezas semánticas y ajustar el registro al contexto económico-financiero. Este tipo de intervenciones mejoran la precisión de la traducción y evitan interpretaciones erróneas que podrían afectar la comprensión del mensaje por parte de los usuarios finales.

En el ámbito jurídico, las dificultades detectadas en la traducción de términos especializados evidencian la importancia de contar con traductores con formación específica en derecho. La interpretación incorrecta de expresiones como *surrogate taxpayer* o *tax liability* por parte de los sistemas de IA pone de relieve la necesidad de intervención humana en contextos donde la precisión terminológica es fundamental para evitar ambigüedades jurídicas. En estos casos, la IA generativa fue capaz de proporcionar traducciones gramaticalmente correctas, pero careció del conocimiento especializado necesario para captar los matices y los usos técnicos del lenguaje jurídico. Así, el traductor especializado no solo actúa como un corrector lingüístico, sino como un mediador experto que debe garantizar que el mensaje se mantenga alineado con las normativas y prácticas jurídicas del idioma de destino.

Los ejemplos analizados evidencian con claridad la necesidad de una formación especializada, al tiempo que plantean un desafío estructural para los programas de estudios en traducción: la revisión y actualización curricular. Los planes de formación deben incorporar módulos dedicados a la posesición, a la evaluación crítica de traducciones generadas por inteligencia artificial y al análisis sistemático de errores. Esta transformación curricular requiere que los futuros traductores desarrollen competencias analíticas sólidas, capaces de identificar patrones de error frecuentes en los textos automatizados y de formular estrategias de corrección eficaces. Además de resolver problemas léxicos o gramaticales, esta capacidad permite reconocer sesgos culturales, prejuicios implícitos y usos terminológicos impropios que pueden afectar la calidad y la pertinencia del texto final.

Los casos prácticos analizados a lo largo del capítulo confirman que la labor del traductor especializado excede con creces la corrección superficial del texto. En contextos de alta complejidad, su función adquiere un carácter estratégico: debe determinar si la traducción generada por inteligencia artificial cumple con el propósito comunicativo del encargo o si requiere una reformulación integral para preservar la claridad, la precisión conceptual y la adecuación al destinatario. El ejemplo del término *callable agencies* ilustra esta exigencia: la versión automatizada resultó ambigua y potencialmente equívoca, mientras que la intervención del traductor especializado permitió reinterpretar el concepto, situarlo en su contexto financiero y reformularlo con la terminología pertinente. Esta capacidad de análisis y reconfiguración discursiva no puede quedar librada a la intuición individual, sino que debe ocupar un lugar central en la formación profesional. Los programas de estudio deben incluir prácticas que entrenen a los futuros traductores

en la evaluación funcional de textos automatizados, atendiendo no solo a la corrección formal, sino también a la solidez conceptual, la coherencia argumentativa y la adecuación estilística según el género discursivo y el público al que se dirige.

Otro aspecto crucial que debe integrarse en los planes de estudio es el análisis de sesgos y prejuicios implícitos en los modelos de IA generativa. Como se ha señalado en estudios previos, los sistemas de IA entrenados con grandes volúmenes de datos tienden a replicar patrones discriminatorios presentes en los textos de entrenamiento. En este sentido, los traductores especializados deben ser capaces de identificar y corregir estos sesgos, asegurando que la traducción final sea inclusiva, precisa y respetuosa con las diferencias culturales. La posesición, en este contexto, no es un simple proceso de corrección, sino una actividad ética y crítica que demanda un conocimiento profundo de los contextos socio-culturales y una sensibilidad hacia las implicaciones del discurso traducido.

En definitiva, el porvenir de la traducción profesional depende de la consolidación de un perfil especializado, con dominio técnico, criterio terminológico y sensibilidad discursiva. Si bien la inteligencia artificial generativa ha demostrado ser eficaz en tareas rutinarias o de baja complejidad, su desempeño se debilita en contextos donde se requiere una comprensión profunda del contenido, del registro y del propósito comunicativo. Los ejemplos analizados a lo largo del capítulo evidencian que, en esos casos, el traductor formado en un área específica no solo resuelve fallas de la máquina, sino que añade valor interpretativo, funcional y ético a la mediación lingüística.

Este diagnóstico exige una revisión sustantiva de los modelos de formación. La enseñanza de la traducción debe abandonar esquemas genéricos y avanzar hacia una propuesta integral que articule alfabetización tecnológica, pensamiento crítico y especialización temática. No se trata únicamente de enseñar a usar herramientas, sino de formar profesionales capaces de evaluar sus salidas, intervenir estratégicamente sobre los textos generados y tomar decisiones que aseguren la pertinencia del mensaje en contextos multilingües y culturalmente diversos. La inteligencia artificial, lejos de desplazar al traductor, amplía sus funciones, redefine sus responsabilidades y exige un nuevo tipo de experticia.

Más que una amenaza, la irrupción de la IAG marca una oportunidad para quienes estén dispuestos a asumir este nuevo rol con profundidad, precisión y sentido profesional. La capacidad de analizar, validar y reformular críticamente contenidos automatizados se convierte hoy en una competencia estructural, no en un diferencial. En un entorno atravesado por la aceleración tecnológica, quienes dominen el oficio con rigor, comprendan las dinámicas discursivas y se posicionen como mediadores entre lenguas, culturas y sistemas inteligentes serán los actores centrales de la traducción del presente y del futuro.

**Referencias bibliográficas:**

- Bahdanau, D., Cho, K., Bengio, Y. (2015). “Neural Machine Translation by Jointly Learning to Align and Translate”. Disponible en arXiv preprint arXiv:1409.0473.
- Bender, E., Gebru, T., McMillan-Major, A., Shmitchell, S. (2021). “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?” *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*.
- Bowker, L. (2019). *Computer-Aided Translation: An Introduction*. Routledge.
- Bowker, L. (2021). “Machine Translation and the Future of the Translator”. *Translation Spaces*, 10(1), pp. 57-76.
- Castilho, S., Moorkens, J., Gaspari, F., Doherty, S., Way, A. (2017). “Is Neural Machine Translation the New State of the Art?” *The Prague Bulletin of Mathematical Linguistics*, 108(1), pp. 109-120.
- Costa-jussà, M. R., Hardmeier, C. (2019). “Addressing gender bias in machine translation”. En *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts*, pp. 1-2. Association for Computational Linguistics.
- García, I. (2023). “The Role of the Translator in the Age of AI”. *Journal of Translation Studies*, 45(2), pp. 67-89.
- Goodfellow, I., Bengio, Y., Courville, A. (2016). *Deep Learning*. MIT Press.
- Jurafsky, D., Martin, J. H. (2021). *Speech and Language Processing* (3rd ed.). Pearson.
- Kenny, D. (2022). *Machine Translation and Global Research: Towards Improved Machine Translation Literacy*. Routledge.
- Koehn, P. (2020). *Neural Machine Translation*. Cambridge University Press.
- LeCun, Y., Bengio, Y., Hinton, G. (2015). “Deep Learning”. *Nature*, 521(7553), pp. 436-444.
- Luong, M. T., Pham, H., Manning, C. D. (2015). “Effective Approaches to Attention-based Neural Machine Translation”. Disponible en arXiv preprint arXiv:1508.04025.
- Moorkens, J. (2022). “Ethics in Machine Translation”. *Translation and Interpreting Studies*, 17(3), pp. 432-450.
- Mossop, B. (2021). *Revising and Editing for Translators*. Routledge.
- Melby, A. (2020). *Human and Machine Translation: Ethics, Trust, and Responsibility*. John Benjamins Publishing.
- OpenAI. (2023). “GPT-4 Technical Report”. Disponible en arXiv preprint arXiv:2303.08774.
- Pym, A. (2011). “Translation Theory and Machine Translation: Can They Be Friends?”. *Translation and Interpreting Studies*, 6(2), pp. 163-182.
- Pym, A. (2020). *Translation and Text Transfer: An Essay on the Principles of Intercultural Communication*. Routledge.
- 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* (NeurIPS), 30.
- TAUS. (2019). *TAUS Post-Editing Guidelines*.
- Toral, A., & Way, A. (2018). "Post-Editing Neural Machine Translation". *Journal of Language Technology*, 5(1), pp. 23-45.
- Vincent, J. (2021). *The Rise of Large Language Models*. The Verge.
- Vincent, J. (2021). "The unstoppable rise of AI language models: How GPT-4 and beyond are shaping the future of text generation". *The Verge*.  
<https://www.theverge.com/>
- Way, A. (2022). "Machine Translation: Past, Present, Future". *Language and Translation Studies*, 20(1), pp. 12-34.

**Dulce Maria Santiago**

Doctora en Filosofía por la Pontificia Universidad Católica Argentina (2006). Profesora Titular en la misma Universidad y en la Universidad Nacional Santo Tomás de Aquino. Investigadora sobre Pensamiento Latinoamericano en la Universidad Nacional del Sur. Publicó El problema de la cultura (2009) y numerosos capítulos de libros colectivos. Dictó cursos en Cuba (2011) y en Colombia (2016).

**Ian Henríquez Herrera**

Ian Henríquez Herrera es doctor y magíster en derecho privado de la Universidad de los Andes; magíster en Derecho Privado y licenciado en Ciencias Jurídicas de la Universidad de Chile. Investigador del Instituto de Investigación en Derecho de la Universidad Autónoma de Chile y fellow de la Universidad Ludovika de Budapest.

**Roxana Alvarez Sanguinetti**

Roxana Alvarez Sanguinetti es Master of Arts in English (Boston College); licenciada en Letras, mención lingüística y literatura inglesa de la Pontificia Universidad Católica de Chile. Profesora asistente, Boston College (USA).

**Gustavo Riesgo**

Licenciado en Filosofía por la Universidad del Norte Santo Tomás de Aquino donde completó el Posgrado de Especialización en Pensamiento Medieval. Obtuvo becas de perfeccionamiento académico en universidades de California, Oxford y Austral.

Es Master y Doctor en Filosofía de las Ciencias con su tesis Teoría General de la Inteligencia Artificial. Una Epistemología de la Ingeniería del Conocimiento.

Posee extensa formación previa en Sistemas y Computación sobre los que ha realizado posgrados de especialización en inteligencia artificial; con más de treinta años de experiencia laboral en el rubro, fungen como Chairman de la empresa AXIS HUMAN.

Es también candidato a Doctor en Filosofía en la UBA por la investigación sobre Metafísica de la Luz en Dionisio Areopagita. Preside la Fundación Areopagiticum dedicada al Neoplatonismo y las Humanidades Digitales.

Profesor universitario titular de Filosofía de la Inteligencia Artificial en la UNSTA. Ha publicado artículos y libros en sus especialidades.

Lenguaje, Cultura y Algoritmos: convergencias entre la Inteligencia Artificial y las Humanidades, se sitúa en el umbral de un cambio de época en el que la inteligencia artificial irrumpe no solo como innovación tecnológica, sino como fenómeno cultural, antropológico y espiritual. Desde una perspectiva inspirada en el humanismo cristiano, la obra propone recuperar el papel insustituible de las Humanidades como ámbito privilegiado para el discernimiento sobre la verdad, el bien y la dignidad de la persona.

A través de una trama interdisciplinaria, los distintos aportes abordan las tensiones que emergen entre el desarrollo tecnocientífico y la comprensión integral de lo humano. Se examinan críticamente las derivas del transhumanismo, las ambigüedades del lenguaje aplicado a los sistemas algorítmicos y los desplazamientos que la inteligencia artificial introduce en el arte, la lógica y la formación profesional. En todos los casos, se advierte el riesgo de una racionalidad que, al absolutizar la eficiencia, pierde de vista el sentido.

Lejos de toda fascinación acrítica o rechazo simplista, el libro invita a una conversión cultural que permita reorientar la técnica hacia una ecología integral, donde la persona humana sea reconocida en su centralidad y en su apertura a la trascendencia.

Así, esta obra se ofrece como un espacio de encuentro entre fe, razón y cultura, y como una invitación a pensar, con responsabilidad y esperanza, el futuro que estamos llamados a construir juntos.

## **Claudio Conforti**

Claudio Marcelo Conforti es Bachiller Universitario y Profesor de Enseñanza Secundaria Normal y Especial en Filosofía por la UNSTA. Máster Universitario en Lógica y Filosofía de la Ciencia por la Universidad de Salamanca (España). Profesor de la UCA en la Facultades de Filosofía y Letras y de Psicología y Psicopedagogía. Profesor de Filosofía de la Ciencia y de la Naturaleza en la UNSTA. Profesor titular de Metodología de las Ciencias en la UdeMM. Vicerrector del Instituto de Educación Superior N° 1 “Dra. Alicia Moreau de Justo”.