



10.15446/fyf.v38n1.114947

Artículos

## LAS NUEVAS IRREGULARIDADES DEL LENGUAJE. DESAFÍOS DE INTERPRETACIÓN Y MECANISMOS DE SIMPLIFICACIÓN DISCURSIVA DE LA INTELIGENCIA ARTIFICIAL\*

## THE NEW IRREGULARITIES OF LANGUAGE. CHALLENGES OF INTERPRETATION AND MECHANISMS OF DISCURSIVE SIMPLIFICATION OF ARTIFICIAL INTELLIGENCE

*Pablo von Stecher*<sup>1</sup> ORCID <https://orcid.org/0000-0001-7588-817X>

### **Cómo citar este artículo:**

Von Stecher, P. (2025). Las nuevas irregularidades del lenguaje. Desafíos de interpretación y mecanismos de simplificación discursiva de la inteligencia artificial. *Forma y Función*, 38(1). <https://doi.org/10.15446/fyf.v38n1.114947>

Este es un artículo publicado en acceso abierto bajo una licencia Creative Commons.

Recibido: 2024-06-08, aceptado: 2024-09-07

---

\* Esta investigación forma parte del proyecto: «La simplificación del lenguaje en la comunicación con máquinas: entre los inicios de la computación en Argentina y los desarrollos actuales del Procesamiento del Lenguaje Natural. Estudio glotopolítico», financiado por el Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET, Argentina).

<sup>1</sup> Universidad de Buenos Aires, Buenos Aires, Argentina, [pablovonstecher@gmail.com](mailto:pablovonstecher@gmail.com)

## Resumen

Se abordan los documentos en español de empresas tecnológicas que explican y promocionan productos del Procesamiento del Lenguaje Natural (componente de la Inteligencia Artificial que estudia cómo las computadoras procesan el lenguaje humano), tales como sistemas de transcripción, uso de *chatbots*, traducción automática. El objetivo es analizar qué consideraciones lingüísticas priman en estos documentos y qué fenómenos del lenguaje deben ser regulados para un buen funcionamiento del PLN. El trabajo se inscribe en un enfoque glotopolítico, perspectiva que analiza las distintas intervenciones sociales sobre el lenguaje. Sobre un *corpus* conformado por los documentos de seis compañías, se estudian tanto los ideologemas y las representaciones sociolingüísticas, como la dimensión enunciativo-argumentativa. Se mostrará que la recurrente caracterización del lenguaje a partir de su complejidad opera discursivamente para justificar las limitaciones del PLN; y que resultan categorizados como «irregularidades» una serie de fenómenos muy disímiles entre sí en cuanto a su naturaleza lingüístico-discursiva, que involucra tanto aspectos semántico-polisémicos como discursivo-polifónicos, pero que también incluye el uso de recursos retóricos y de variedades lingüísticas.

**Palabras clave:** *regulación del lenguaje; procesamiento del lenguaje natural; polisemia; polifonía; variedades sociolingüísticas; glotopolítica.*

## Abstract

Documents written in Spanish by technology companies that explain and promote products of Natural Language Processing (NLP, a component of Artificial Intelligence that studies how computers process human language), such as transcription systems, chatbots, automatic translation, etc., are discussed in this work. The purpose of this study is to analyze which linguistic considerations prevail in these documents and which language phenomena should be regulated for the proper functioning of NLP. This study is conducted within a glottopolitical approach, a perspective that analyzes the different social interventions on language use. From a corpus consisting of documents from six companies, we studied ideologemes and sociolinguistic representations, as well as their enunciative-argumentative dimension. Our analysis shows that the recurrent characterization of language on the basis of its complexity operates discursively to justify the limitations of NLP, and that a series of phenomena can be categorized as “irregularities”, which are very dissimilar in terms of their linguistic-discursive nature, involving both semantic-polysemic and discursive-poliphonic aspects, as well as including the use of rhetorical resources and linguistic varieties.

**Keywords:** *language regulation; natural language processing; polysemy; polyphony; sociolinguistic varieties; glottopolitics.*

## 1. Introducción: tema y objeto

Si bien el uso masivo y consciente de las aplicaciones de la Inteligencia artificial (IA) por parte de la sociedad se ha venido consolidando en las últimas décadas, los primeros adelantos de esta rama avanzada de la informática datan de mediados del siglo pasado, cuando el matemático inglés Alan Turing (1912-1954) publicó un artículo titulado «Computing Machinery and Intelligence». Allí proponía el *test de Turing*, que consistía en la diferenciación, por parte de un sujeto que operaba como «interrogador», entre las respuestas otorgadas por un ser humano y las proporcionadas por una máquina (Turing, 1950; Esparza Urzúa, 2021). Históricamente, se ha definido la IA como el conjunto de estudios y aplicaciones orientados a dotar la computadora de las capacidades que habitualmente se le atribuyen a la inteligencia humana, siguiendo procesos análogos al razonamiento (adquisición de conocimiento, percepción, toma de decisiones), con el fin de resolver problemas de cierta complejidad (Rolston, 1990; Haton & Haton, 1991; Morales Ascencio, 1997).

Entre las distintas ramas de la IA, el Procesamiento del Lenguaje Natural (PLN) estudia cómo las computadoras entienden y procesan datos del lenguaje humano («natural») con el propósito de diseñar y aplicar sistemas informáticos que faciliten la comunicación entre personas y máquinas, es decir, se busca imitar artificialmente ciertos aspectos de las capacidades humanas para el lenguaje (Gelbukh, 2010; Camacho-Álvarez & Navarro Álvarez, 2020). Sus tareas actuales más representativas son: el desarrollo de sistemas de diálogo y conversación por medio de *chats* y el reconocimiento acústico-fonético por parte de la máquina, la traducción automática o semiautomática, las técnicas para convertir sonidos en textos (y viceversa), la corrección automática de textos, los sistemas de transcripción automatizada, la extracción y recuperación de información, la generación y entendimiento de lenguaje natural, entre otros (Alias & Cassanelli, 2019; Moreira et al., 2021; Báez et al., 2022).

La arquitectura del sistema de PLN se divide en cinco niveles: el *fonológico*, que aborda la estructuración de los sonidos de una lengua; el *morfológico*, que revisa los morfemas en función de la organización interna de las palabras; el *sintáctico*, que estudia el modo en que las palabras se articulan para formar oraciones; el *semántico*, que revisa el significado de cada palabra, así como de cláusulas y oraciones; y el *pragmático*, que aborda las distintas interpretaciones de una oración en diferentes situaciones o en función de la dimensión extralingüística en que se inscribe (Cortez Vásquez et al., 2009; Camacho-Álvarez & Navarro Álvarez, 2020).

Desde la perspectiva glotopolítica, se propuso la categoría *Tecnologías digitales del lenguaje* para referir los instrumentos ideados y desarrollados para la comunicación en

línea, productos del PLN que, paralelamente, crean nuevas formas de homogenización y regulación lingüística (Bürki, 2023). También inscripto en el enfoque glotopolítico, el objetivo general del presente artículo es analizar las representaciones del lenguaje que tienen lugar en una serie de documentos de divulgación, proporcionados por las grandes compañías tecnológicas internacionales que buscan comercializar herramientas del PLN en español. Los objetivos específicos son: (1) indagar cómo es caracterizado el lenguaje en estos enunciados y analizar qué representaciones sociolingüísticas allí se activan; (2) estudiar qué fenómenos y dimensiones del lenguaje se proponen para ser regulados, modificados u omitidos en pos de un diálogo eficaz con la máquina, (3) analizar el vínculo que estos documentos establecen entre las consideraciones acerca del lenguaje y los alcances y limitaciones del PLN. Complementariamente, y a través de los saberes provenientes de distintas líneas de las ciencias del lenguaje, se historizan las ideas en torno a las *irregularidades del lenguaje* que tienen lugar en el material analizado, tanto para reflexionar acerca del carácter de irregularidad, como para mostrar continuidades de interés sobre estos fenómenos en discursos previos.

## 2. El enfoque glotopolítico y sus aportes al estudio del PLN

### 2.1. Planificación lingüística y glotopolítica

Hacia fines de la década de 1960, luego de la publicación fundacional de Einar Haugen acerca de la estandarización gramatical en Noruega (Haugen, 1959), empieza a estabilizarse de manera gradual un campo de estudios sobre la *planificación lingüística*, es decir, sobre la puesta en práctica e implementación de políticas lingüísticas, acciones ejercidas mayormente por las autoridades gubernamentales de los estados nacionales (Calvet, 1997). Desde entonces, trabajos provenientes de la sociolingüística y la sociología del lenguaje, de la mano de figuras como Charles Ferguson o Joshua Fishman, entre otros, inician una serie de reflexiones sobre la planificación del *corpus*, correspondiente a las decisiones vinculadas con la forma de la lengua (la estandarización, la gramática, la neología, la ortografía) y respaldadas mediante la elaboración de diccionarios, gramáticas, tratados ortográficos; y sobre la planificación del *estatus*, correspondiente a las decisiones vinculadas con las funciones de una lengua, su estatuto social, su enseñanza y su relación con otras lenguas (Kloss, 1969; Calvet, 1997).

Hacia la década de 1980, los sociolingüistas franceses Louis Guespin y Jean-Baptiste Marcellesi introducen el término *glotopolítica* para definir una línea de estudios cuyo objetivo fuera analizar las diversas formas en que una sociedad actúa sobre el lenguaje, sea o no consciente de ello. Referir las acciones sobre el *lenguaje* les permite superar

la dicotomía saussureana *lengua/habla* y abarcar tanto el análisis de las intervenciones sobre la *lengua*, por ejemplo, cuando una sociedad legisla sobre los estatutos recíprocos de la lengua oficial y las lenguas minoritarias; como las intervenciones sobre el *habla*, por ejemplo, en los casos en que se reprimen determinados usos; así como las intervenciones sobre el *discurso*, por ejemplo, cuando la institución escolar decide convertir en objeto de estudio la producción de un determinado tipo de texto (Guespin & Marcellesi, 1986).

En las dos últimas décadas, principalmente a partir de la conformación del equipo coordinado por Elvira Arnoux en la Universidad de Buenos Aires y del equipo coordinado por José del Valle en la City University of New York, se produce un auge de estudios de carácter glotopolítico y, paralelamente, un proceso de reflexión acerca de sus alcances. Arnoux (2008, p. 18) la define como el análisis de «las ideologías lingüísticas y las intervenciones en el espacio público del lenguaje», indagación que paralelamente atiende al modo en que dichas intervenciones participan en la instauración, reproducción o transformación de entidades políticas, relaciones sociales y estructuras de poder, ya sea en el ámbito nacional, regional o planetario.

Son muy diversos los materiales que pueden constituirse en objeto de un estudio glotopolítico, tanto en lo que refiere a su dimensión semiótica (textos orales y escritos, productos audiovisuales y dispositivos multimodales); así como a la naturaleza más o menos institucional del ámbito en que se producen y circulan<sup>1</sup>. Estos materiales abarcan: (1) documentos normativos que reglamentan el espacio de las lenguas: resoluciones, decretos, leyes; (2) instrumentos que regulan la actividad lingüística y discursiva: gramáticas, diccionarios, ortografías, manuales de estilo; (3) dispositivos regulatorios asociados a las nuevas tecnologías de la palabra y la comunicación: correctores, anticipadores, traductores automáticos; (4) documentos variados que exponen o reflexionan sobre prácticas lingüísticas: ensayos, artículos periodísticos, biografías; (5) actividades epilingüísticas (espontáneas, intuitivas) que tienen y exponen los hablantes acerca de determinados fenómenos, objetos o usos lingüísticos y prácticas metalingüísticas. De este modo, a diferencia de los principios de la planificación lingüística, la glotopolítica se abre al estudio de intervenciones que exceden los casos de textos o instrumentos normativos de carácter oficial, así como al de los distintos agentes que pueden llevar adelante tales intervenciones, más allá del Estado o las entidades transestatales.

## 2.2. Antecedentes de la problemática

Desde el enfoque glotopolítico, se ha señalado el accionar de regulación discursiva que ejercen, en la interacción comunicativa escrita en soportes digitales, mecanismos

como los guionadores, los facilitadores y los correctores, en la medida en que predicen y dirigen la redacción de correos electrónicos, *chats* de redes sociales y mensajes instantáneos de ciertas aplicaciones para teléfonos inteligentes (Arnoux & Lauria, 2023). Asimismo, se ha advertido la tendencia hacia la automatización del discurso como efecto de los usos de la IA, instancia que se manifiesta en la propagación de normas globalizadas, así como en recomendaciones sobre la unificación de la prosa, el armado de oraciones o párrafos y la selección léxica. Se trata de regulaciones que, al tiempo que facilitan y promueven la velocidad del accionar de buscadores, traductores automáticos y *chatbots* o asistentes virtuales; reducen sus costos económicos y potencian su velocidad (Lauria, 2019, Arnoux & Lauria, 2023).

Noam Chomsky ha contrastado la asimilación de la lengua por parte del ChatGPT, en tanto motor estadístico de comparación de patrones alimentado por millones de datos; frente a la eficiencia, inconsciencia, automaticidad y rapidez con que la aprende una persona. Para Chomsky (2023), el defecto más profundo de la IA es la ausencia de capacidad crítica y de creación explicativa, falencias que la llevan a la predicción errónea de reglas sintácticas y a la reproducción literal de oraciones, determinadas por la incapacidad de interpretación que no se logra compensar desde el *big data*. En efecto, se detectó que más allá de la versatilidad del conocimiento lingüístico que maneja el PLN, obtenido a partir del entrenamiento de datos, hay dudas acerca del alcance de su gestión en casos como conversaciones informales o enunciados donde la incidencia de factores sociales, culturales o políticos es determinante; así como tampoco es clara su solidez en la reproducción multilingüe de datos (Papadimitriou & Manning, 2021).

En lo que refiere a la incidencia de la lengua en el desarrollo de los servicios de PLN, se observó que los modelos algorítmicos están entrenados con *corpus* en inglés, lengua en la que también está basada la mayoría de las herramientas del PLN, lo que tiende a (re)producir textos de acuerdo con los patrones estructurales y léxicos de la cultura textual anglosajona (Ananiadou, McNaught & Thompson, 2012; Bürki, 2023; Kannan, 2023). En las tareas de traducción automática del inglés al español y al francés, por caso, se ha detectado la pérdida de riqueza léxica y morfológica, producto de los efectos simplificadores de los algoritmos utilizados y, con ello, el modo en que este tipo de traducción conduce a una lengua artificialmente empobrecida (Vanmassenhove, Shterionov & Gwilliam, 2021).

La creación del programa *Lengua Española e Inteligencia Artificial* (LEIA) en 2019, iniciativa impulsada por la Real Academia Española (RAE) y la Asociación de Academias de la Lengua Española (ASALE), ha despertado el interés de estudios glotopolíticos que analizan cómo sus discursos de presentación instan tanto a tomar medidas frente a

la diversificación del español, como a cuidar el uso correcto de la lengua en los medios tecnológicos, dos propuestas que se retroalimentan y que llevan a desarmar el lema panhispánico de «unidad en la diversidad» para privilegiar el primero en desmedro del segundo (Alonso, 2021; Iglesias, 2024)<sup>2</sup>.

Finalmente, como antecedente clave para esta investigación, se ha anticipado cómo los servicios de las *Tecnologías Digitales del Lenguaje*, cuyo desarrollo precisa de una acción normativa homogeneizadora de las prácticas lingüísticas, encuentran dificultades al pretender captar fenómenos como las expresiones idiomáticas, las ambigüedades o las referencias contextuales (Bürki, 2023).

### 3. Metodología y delimitación del *corpus*

#### 3.1. Referencias metodológicas de la perspectiva glotopolítica

En términos metodológicos, bajo el propósito de indagar cómo operan las distintas intervenciones que se efectúan en el espacio público del lenguaje, el enfoque glotopolítico utilizado en este estudio desmonta y analiza las *representaciones sociolingüísticas*, o sea, aquellas representaciones sociales que refieren a objetos lingüísticos (lenguas, variedades, hablas, acentos, registros, géneros, modos de leer o de escribir, etc.), así como a los sujetos asociados con tales objetos (Arnoux & Del Valle, 2010)<sup>3</sup>. Este análisis implica atender al modo en que tales representaciones se inscriben en *ideologías lingüísticas*, esto es, en sistemas de ideas que articulan nociones lingüísticas (lengua, habla, comunicación) con formaciones culturales, políticas o sociales específicas (Del Valle, 2007, 2017). Para ello, detecta la ocurrencia de *ideologemas*, es decir, lugares comunes, postulados o máximas que, o bien materializados en la superficie discursiva o bien subyacentes al enunciado, funcionan como presupuestos del discurso (Amossy & Herschberg Pierrot, 2001; Angenot, 2010).

Otro lineamiento constitutivo del enfoque glotopolítico es el hecho de abordar los textos como discursos, lo que implica considerar las condiciones sociohistóricas en que han sido producidos y que han dejado su huella en el material textual, por lo que este enfoque articula el análisis del discurso como herramienta para abordar los materiales (Del Valle, 2017). Esta mirada discursiva a la que adscribe la glotopolítica posibilita reparar en representaciones e ideologías provenientes de otras temporalidades, pero que asumen en los textos metalingüísticos actuales un sentido histórico-político y una función social diferentes, atados a los tiempos en los que surgen (Lauria, 2020). En este sentido, el estudio del dispositivo enunciativo y polifónico, el análisis de las secuencias textuales y de las operaciones de reformulación, así como los modos de ejemplificar

han sido descriptos como procedimientos exploratorios prototípicos en las primeras entradas al estudio del *corpus* que propone este enfoque (Arnoux, 2008).

### 3.2. Criterios de conformación del *corpus*

Las distintas empresas internacionales que difunden y promocionan servicios del PLN contienen en sus respectivas páginas *web* una serie de documentos de divulgación (en algunos casos denominados *guías*) que, sin marcas específicas de autoría, se proponen la explicación de su funcionamiento. Tal como tiende a ocurrir con el discurso de divulgación, los documentos se despliegan bajo la dinámica interrogante y respuesta. «¿Qué es el PLN?», «¿Cómo funciona el PLN?», «¿Qué tareas desarrolla el PLN?», «¿Cuáles son las herramientas del PLN?» o «¿Cuáles son los desafíos del PLN?» son las preguntas prototípicas que orientan las explicaciones. El análisis se centra, entonces, en los enunciados que, con el fin de responder a estos interrogantes, definen o caracterizan el lenguaje humano.

Con el propósito de describir el modo en que se reunieron y delimitaron los materiales de análisis, se apela a la distinción entre *archivo discursivo* y *corpus*. Mientras que el primero responde al gran conjunto de materiales reunidos en una instancia previa al análisis, en función de los géneros discursivos delimitados y las temáticas identificadas; el *corpus* es el resultado de esos materiales luego de su proceso de selección, recorte y ordenamiento en función de las hipótesis de trabajo (Aguilar et al., 2014).

Se ejecutaron, de este modo, los siguientes pasos: (1) se relevaron las distintas páginas en español de las compañías tecnológicas internacionales que comercializan, dentro de los servicios de IA, los productos del PLN, proceso que se llevó a cabo entre los meses de enero y febrero de 2024 y que dio lugar a nuestro archivo discursivo inicial<sup>4</sup>; (2) se detectaron y focalizaron aquellos enunciados que, en pos de la comercialización de los productos y servicios, despliegan una reflexión metalingüística; (3) se efectuó un recorte de dichos enunciados para concentrar el análisis en las consideraciones que se proponen caracterizar el lenguaje, así como en los fenómenos lingüísticos y discursivos que deben ser regulados. De este modo, el *corpus* definitivo se conformó solo por fragmentos de los documentos correspondientes a las siguientes compañías: Amazon Web Service, BAOSS, Botpress, Elastic (The Search AI company), SAS (Software de Analítica y Soluciones) e IBM (International Business Machine Corporation), disponibles en sus respectivas páginas *web*<sup>5</sup>.

Siguiendo los criterios bajtinianos que consideran el enunciado como la unidad real de la comunicación discursiva, fenómeno ejemplificado en la réplica de un diálogo (Bajtín, 1982), y en tanto los documentos mencionados reproducen la forma de un

cuestionario, consideramos cada respuesta como un enunciado. Se analizaron, entonces, seis enunciados (uno por cada documento de las seis compañías mencionadas), conformados por un total de 1094 palabras.

### 3.3. Especificaciones metodológicas y puntos de partida del análisis

En principio, para estudiar las consideraciones que estos documentos establecen sobre el lenguaje, se apeló a las herramientas del análisis discursivo-enunciativo (Kerbat-Orecchioni, 1986; Maingueneau, 2009), en particular al uso de deícticos, modalidades enunciativas y secuencias textuales, así como a la identificación de ideologemas<sup>6</sup>. Luego, para examinar los enunciados que refieren a la irregularidad del lenguaje, se rastrearon y estudiaron las representaciones sociolingüísticas que remiten a esta problemática y se identificaron sus antecedentes en discursividades previas. Paralelamente, se abordaron y analizaron los fenómenos específicamente descritos como «irregularidades», a la luz de los aportes que distintas líneas de las ciencias del lenguaje han referido al respecto.

A continuación, se presenta el análisis de los enunciados y sus resultados en el siguiente orden: primero, en la sección 4, se exponen las caracterizaciones que estos documentos establecen sobre el lenguaje; luego, en la sección 5, se expone el estudio de los fenómenos identificados como irregularidades, sistematizado en dos grandes dimensiones: metáfora, polifonía y polisemia; y léxico y variedades sociolingüísticas.

Son dos las ideas que busca demostrar este trabajo. Por un lado, se propone que la recurrente caracterización del lenguaje, a partir de su carácter complejo, opera discursivamente para justificar las limitaciones de efectividad de los productos y servicios del PLN. Por otro lado, se sostiene que bajo la etiqueta de la *irregularidad* se advierte sobre una heterogénea serie de fenómenos discursivos (metáforas, ironías, juegos de palabras) y sociolingüísticos (regionalismos, variedades, expresiones idiomáticas) que deben ser regulados u omitidos para el buen funcionamiento del PLN.

## 4. Entre la complejidad del lenguaje y las limitaciones del PLN. Ideologemas y consideraciones discursivas

Tal como había anticipado Ferdinand de Saussure (1945 [1916]), siempre recordado como el «padre de la lingüística moderna», emprender un estudio del lenguaje en su conjunto es una tarea infructuosa dadas las distintas dualidades que atraviesan el fenómeno lingüístico (lo acústico y lo vocal, lo social y lo individual, su historia y su estado actual) y dados los distintos dominios (físico, fisiológico, psíquico) que confluyen en su indagación. Siguiendo la célebre traducción al español del *Curso de lingüística general* efectuada por Amado Alonso, aprendimos que el lenguaje es «multiforme y

heteróclito» (De Saussure, 1945 [1916], p. 37) y que, en cambio, la lengua parecía presentarse de manera más acotada y sistematizable como para conformarse en el objeto de estudio de la lingüística.

El ideograma de la complejidad del lenguaje vuelve con fuerza en los documentos explicativos que instruyen sobre las aplicaciones del PLN. Si bien ya no se trata aquí de las dificultades que habría planteado el lingüista suizo consecuentes de un abordaje en simultáneo de sus distintas dimensiones, sino de los desafíos que implica su uso cotidiano para un buen funcionamiento del PLN, sí se puede decir que estos materiales refieren en pocas líneas a muy diferentes aspectos y niveles del lenguaje.

En este sentido, el documento de IBM plantea: «El lenguaje humano está lleno de ambigüedades que hacen increíblemente difícil escribir un *software* que determine con precisión el significado deseado de los datos de texto o voz» (IBM, 2024). Se trata de «irregularidades del lenguaje humano que los humanos tardan años en aprender, pero que los programadores deben enseñar a reconocer y entender con precisión desde el principio a las aplicaciones basadas en el lenguaje natural si quieren ser útiles» (IBM, 2024).

El fragmento expone una serie de elementos discursivos que potencian el rasgo complejo del lenguaje: por un lado, el modalizador apreciativo de enunciado *increíblemente*; por otro, los subjetivemas *difícil* y *lleno de ambigüedades*; finalmente, la referencia a la extensión temporal que implica su aprendizaje (*años*).

Vamos a observar que los otros instructivos recuperan distintos aspectos de esta caracterización. Amazon, por caso, retoma la problemática temporal al señalar que el lenguaje tiene «características [...] de uso que los humanos tardan años en aprender» (Amazon, 2024). SAS, en tanto, vuelve sobre la dificultad intrínseca del lenguaje: «El lenguaje humano es increíblemente complejo y diverso. Nos expresamos de maneras infinitas, verbalmente y por escrito» (SAS, 2024). Además de repetir el mismo adverbio (*increíblemente*) ahora en posición de subjetivema, así como de remitir al carácter inacabado que pueden tomar nuestras expresiones, nótese que se utiliza el término *verbalmente* por *oralmente*. Este detalle no es menor, ya que, como veremos, otro rasgo de estos documentos es cierta laxitud en la terminología lingüística utilizada, lo que puede redundar en la complejidad que se propone describir.

Según BAOSS, «el PLN es uno de los grandes retos de las IA, ya que es una de las tareas más complicadas y desafiantes: ¿cómo comprender realmente el significado de un texto? [...] Nuestro lenguaje está lleno de ambigüedades [...] Esto hace que el PLN sea una de las tareas más difíciles de dominar» (BAOSS, 2024). Como se puede observar, hay cierta estabilidad en las caracterizaciones, incluso se repite la misma terminología (*lleno de ambigüedades*, *difícil*). Aquí la complejidad del lenguaje se asocia con los

retos que debe atravesar el PLN; no obstante, la pregunta filosófica que se plantea en el enunciado parece llevar al extremo la mencionada dificultad. En otras palabras ¿se le puede solicitar al PLN una capacidad de comprensión que, en ciertos casos, excede incluso a las personas?

El documento de Elastic, en tanto, ejemplifica el modo en que el PLN se enfrenta a los desafíos de la «irregularidad» y «ambigüedad» del lenguaje:

Los homónimos y la sintaxis pueden confundir los sets de datos. Incluso el mejor análisis de sentimiento no siempre puede identificar el sarcasmo y la ironía. A los humanos les lleva años aprender estos matices; incluso, entonces, es difícil interpretar el tono de un mensaje de texto o correo electrónico, por ejemplo. (Elastic, 2024)

De este modo, se puntualizan las limitaciones que presenta uno de los servicios específicos de estas tecnologías, el análisis de sentimiento, técnica que busca descifrar el tono emocional, la actitud o el sentimiento expresado en un texto. Finalmente, el documento de Botpress enuncia:

El lenguaje es complejo (más de lo que pensamos), por lo que crear un software que tenga en cuenta todos sus matices y determine con éxito la intención humana detrás de ese lenguaje también es complejo. Pero, al igual que ocurre con la inteligencia humana, un entrenamiento suficiente de la IA permite a una máquina superar estas complejidades. (Botpress, 2024)

En general, los distintos documentos repiten o reformulan estructuras similares. Inician los enunciados con una breve secuencia explicativa reducida a una definición descriptiva que caracteriza el lenguaje a través de su complejidad, para postular luego su consecuencia mediante una secuencia argumentativa que fundamenta entonces los desafíos del PLN. Sin mayor distancia de las guías anteriores, el documento de Botpress introduce la dimensión pragmática del lenguaje al referir la intencionalidad que orienta la formulación de enunciados. No obstante, tal vez menos escéptico que los procesadores mencionados previamente, aquí se pone el foco en el entrenamiento y la posibilidad de superación de tales complejidades antes que en sus limitaciones.

En este último documento, al igual que en los casos correspondientes a SAS y a BAOSS, la configuración de un enunciador articulado a través de un deíctico de primera persona del plural (un *nosotros* inclusivo [«Nos expresamos de maneras infinitas», «Nuestro lenguaje está lleno de ambigüedades», «El lenguaje es complejo —más de

lo que pensamos—») busca instalar que las consideraciones sobre la complejidad del lenguaje sean compartidas por el enunciatario y, con ello, reforzar la persuasión acerca de este aspecto como problema determinante.

Vale agregar, finalmente, que la erradicación de las irregularidades del lenguaje ha sido un objeto histórico de preocupación para los distintos proyectos utópicos que propusieron la creación de lenguas universales, ya sea filosóficas *a priori*, proyectadas como sistemas formales basados en estructuras matemáticas; ya sea *a posteriori*, creadas a partir de las lenguas naturales. Entre las primeras, mayormente de incidencia inglesa, el *Essay towards a Real Character and a Philosophical Language* (1668) de John Wilkins pretendió proyectar una lengua artificial de uso universal que buscaba reducir las redundancias, equívocos e irregularidades con el fin de hallar nomenclaturas adecuadas para los nuevos descubrimientos científicos en el campo de la física y el estudio de la naturaleza (Calero Vaquera, 2011). Wilkins, asimismo, era secretario de la *Royal Society of London for Improving Natural Knowledge* (1662), la sociedad científica más antigua del Reino Unido que no solo orientó el desarrollo de las ciencias naturales desde el siglo XVII, sino que también definió una forma de comunicación científica determinada por un estilo «lineal» y «transparente», y caracterizada por ciertos rasgos: la precisión, la objetividad, la claridad, la ausencia de subjetividad y de todo elemento emocional (Eco, 2005; Ciapuscio, 2011).

Entre los proyectos de lenguas *a posteriori*, tuvo una temprana relevancia el volumen *Del idioma universal, sus ventajas y posibilidad de obtenerlo* que, publicado anónimamente en 1852, y de manera previa al Volapük (1880) y al Esperanto (1887), propuso un sistema basado en la combinación de seis lenguas europeas (latín, italiano, español, francés, inglés y alemán), capaz de sortear toda irregularidad, anomalía o defecto de aquellas (Calero Vaquera, 2011). Forjados en la idea de la imperfección de las lenguas naturales, estos proyectos (entre muchos otros) buscaron ideal e infructuosamente ajustar y determinar cada forma de su uso. De manera aproximada, hasta que el entrenamiento del PLN sea lo suficientemente eficaz para que pueda dar cuenta de las formas que exceden la regularidad del lenguaje, este componente del IA precisa dialogar desde un lenguaje mayormente transparente y lineal.

## 5. Tipos de «irregularidades»

Al avanzar en la lectura de los documentos explicativos sobre el funcionamiento del PLN, los textos introducen una serie de ejemplos para ilustrar lo que previamente han denominado como «irregularidades», «ambigüedades» y «complejidades», términos utilizados de manera equivalente en estos enunciados. Algunos casos, sobre los

que no nos detendremos, refieren o bien a errores ortográficos u omisiones de signos de puntuación en la escritura, o bien el hecho de «tartamudear» o «mascullar» en la oralidad (SAS, 2024). Lo que nos proponemos en los siguientes apartados es el análisis de otro tipo de fenómenos que no dependen de un uso errático del lenguaje, ni de trastornos del habla por parte de los sujetos. En este sentido, organizamos en dos ejes los distintos fenómenos mencionados con el fin de sistematizar el análisis, aunque sin desconocer las posibles continuidades entre ambos: metáfora, polifonía y polisemia; y léxico y variedades sociolingüísticas.

### 5.1. Metáfora, polifonía y polisemia

Entre las «irregularidades» que entorpecen el funcionamiento del PLN, los documentos de IBM y de Amazon identifican las metáforas. A causa de sus efectos polisémicos, las metáforas resultaron proscriptas (al menos enunciativamente) del discurso científico desde el surgimiento de la *Royal Society* hasta bien avanzado el siglo XX y fueron asociadas al campo de la retórica y la literatura (Ciapuscio, 2009, 2011). Para los dispositivos de traducción automática, la metáfora implica un desafío mayor no solo por el riesgo que implica su traducción literal, sino sobre todo porque se trata de un recurso que también constituye nuestro sistema conceptual y estructura nuestra actividad cotidiana (recuérdese los casos prototípicos señalados por Lakoff y Johnson [1986], tales como «el tiempo es dinero» o «la discusión es un batalla»), por lo que su traducción excede la cuestión de las distintas lenguas para involucrar esquemas de comportamiento y sistemas culturales (Arduini, 2002).

Más allá de su identificación explícita, otros documentos alertan de manera indirecta sobre los desafíos que suponen, entre otros fenómenos, las metáforas. Por caso, Botpress señala: «los ordenadores destacan por responder a instrucciones de programación y órdenes predeterminadas en lenguaje llano». El movimiento del *lenguaje llano* (también llamado *lenguaje claro* o *ciudadano*), originado en los países de habla inglesa hacia 1960 bajo la denominación *Plain English* o *Plain Language*, surgió con la intención de simplificar discursivamente el lenguaje jurídico y administrativo, considerado tradicionalmente en términos de opacidad (Becker, 2020). Motivado política y económicamente por una funcionalización del lenguaje acorde con la racionalidad neoliberal, tecnocrática y hegemónica de la eficacia, la eficiencia, la rapidez y el lucro, el *lenguaje llano* proscribió el uso de metáforas y el lenguaje figurado en su búsqueda por la claridad, la precisión y la concisión absoluta (Lauria, 2019; Becker, 2020). De esta manera, en los documentos que describen y promueven el funcionamiento del PLN se activan, por un lado, consideraciones acerca de la linealidad y la transparencia

discursiva que forjaron durante siglos las representaciones del estilo propio del discurso científico. Por otro, se recuperan principios del *lenguaje llano* pensados en función de la simplificación del discurso jurídico y administrativo. Interesa notar, no obstante, que en el marco del PLN, la metáfora representa un problema que no se circunscribe a determinados géneros o discursos específicos, sino que involucra la comunicación cotidiana en términos generales.

Junto con las metáforas, Amazon, IBM y Elastic agregan la complejidad que supone la presencia del «sarcasmo», un subtipo de la ironía caracterizado por una actitud displicente, entendida y compartida por el hablante y el oyente (Cabedo Nebot, 2009). Sin embargo, no sería errado pensar que esta actitud excluya su carácter compartido y conocido en los casos en que uno de los interlocutores en cuestión sea una máquina. En línea similar, las guías de Elastic y BAOSS añaden irregularidades como «las ironías» y «los chistes». Veámoslas por separado.

¿Qué dimensión que excede a la linealidad del lenguaje implica la ironía? Se trata, tal como lo planteara Jacqueline Authier-Revuz (1984) en su investigación pionera sobre las heterogeneidades enunciativas, de la «heterogeneidad mostrada no marcada». Esta forma polifónica introduce otro discurso u otra voz que si bien no cuenta con marcas específicas de distanciamiento como lo serían las comillas, su marco contextual y contextual (así como ciertos tonos en la oralidad) deja de manifiesto que lo dicho no debe ser tomado literalmente. En este sentido, puede interpelar al PLN en la medida en que opera como una antífrasis.

En lo que refiere al chiste, ya en su célebre estudio *El chiste y su relación con el inconsciente*, Sigmund Freud destacaba, entre sus técnicas lingüísticas, «la múltiple acepción de una palabra» y «el doble sentido o juego de palabras» (Freud, 1991 [1905], pp. 31, 36), procedimientos luego repensados a través de la polisemia y la ambigüedad para abordar las distintas interpretaciones que dispara el uso de ciertos términos o expresiones con el fin de generar un efecto humorístico (Nerlich & Chamizo Domínguez, 1999; Palacios, 2018; Ghailani, 2019). En efecto, la limitación de los juegos de sentidos ha sido anticipada como otro de los mecanismos prescriptivos desplegados para la utilización de dispositivos de la IA (Arnoux & Lauria, 2023).

Si la referencia al chiste alude implícitamente a la problemática de la polisemia, esta será referida de manera explícita por el documento de IBM al mencionar el problema que supone para el PLN el caso de «los homónimos y los homófonos», y por el documento de Elastic al referir cómo «los homónimos [...] pueden confundir los sets de datos». Entre los *homónimos*, términos que se caracterizan por tener la misma forma, pero distintos significados, se distinguen: los *homófonos*, palabras que se pronuncian

de la misma manera (aunque puedan tener diferencias ortográficas), pero conllevan significados distintos (*tuvo* y *tubo*; *hojear* y *ojear*); y los *homógrafos*, palabras que se escriben de la misma manera, pero tienen significados diferentes (*cerca* como sinónimo de *cercano* y *cerca* como sinónimo de *alambrado*).

## 5.2. El léxico y las variedades sociolingüísticas

Las otras dimensiones del lenguaje que tienen lugar en los documentos que buscan comercializar el PLN y que son vistas como irregularidades apuntan a cuestiones lexicales vinculadas con la variedad lingüística, es decir, con la diversidad de usos de una misma lengua según la situación comunicativa, geográfica o histórica en que se emplea, pero también según el nivel de conocimiento lingüístico de quien la utiliza.

En este sentido, SAS remite la dificultad que implica el uso de «términos y palabras coloquiales», marcas de variedades funcionales o difásicas, determinadas por la situación de comunicación<sup>7</sup>. Son tres los rasgos que caracterizan el registro coloquial: el estilo informal, la escasa planificación y la relación de igualdad (social o funcional) entre los interlocutores que constituye el marco de familiaridad vivencial y emotiva de la comunicación (Briz, 1996, 2002). Si bien los tres factores pueden conspirar contra una buena respuesta por parte del PLN, es justamente el último elemento el que pone en jaque su funcionamiento, dada la desigualdad que plantea el uso del lenguaje para el hablante nativo y para la máquina.

Asimismo, se destaca el hecho de que «cuando hablamos, tenemos acentos regionales» (SAS, 2024), o sea, exponemos marcas de variedades geográficas o diatópicas, que tampoco favorecen el funcionamiento de las aplicaciones automáticas del PLN. Al respecto se ha observado cómo la política lingüística panhispánica ha sido operativa para garantizar, a través de normas globalizadas y criterios de homogenización, el accionar veloz y económico (en costos y tiempos) de traductores automáticos, buscadores y asistentes virtuales (Arnoux & Lauria, 2023). En este sentido, SAS insiste sobre la problemática de las variedades al referir que «existen cientos de lenguajes y dialectos» (SAS, 2024), otro enunciado que también puede sumar a la confusión, al utilizar la noción de «lenguajes» por la de «lenguas» y, con ello, contribuir (tal vez por desconocimiento de precisiones terminológicas o tal vez, incluso, por deberse a una traducción inexacta) a la representación de la complejidad del lenguaje que estos documentos pregonan.

A ello, IBM agrega «las expresiones idiomáticas» como escollo para los traductores automáticos del PLN. Pensemos en casos prototípicos del español que, debido a su uso extendido y a su memorización se han estandarizado, convencionalizado y fijado en la lengua, tales como «poner el grito en el cielo» o «llevar las de perder». Su dificultad se

debe tanto a su rasgo de «idiomaticidad», ese componente abstracto que hace que su significado global se distancie de la suma de los significados parciales de sus componentes; como a su morfosintaxis peculiar que suele diferenciarse de la rigurosidad de las normas gramaticales (Zuluaga, 1980; Sevilla Muñoz, 1999; Baptista, 2006).

Finalmente, las últimas irregularidades vinculadas a este plano del lenguaje tienen que ver, por un lado, con las «palabras prestadas de otros idiomas» (SAS, 2024); y, por otro, con los «neologismos» (BAOSS, 2024). Si bien ambos fenómenos son tomados de manera diferenciada, nos interesa mencionarlos articuladamente dado que se trata de las dos opciones con que cuentan los hablantes para nombrar los cambios y las nuevas realidades que diariamente se generan. La incorporación de un elemento léxico al sistema para dar cuenta de un concepto se conforma o bien mediante la adopción o adaptación de una forma extranjera (*préstamo*) o bien mediante la creación de un nuevo término (*neologismo formal*), o bien mediante la aplicación significativa de dicho concepto a una forma ya existente (*neologismo semántico*), es decir, una nueva acepción de una palabra ya antigua (Guerrero Ramos, 2013). Estos fenómenos nos reenvían a la problemática del tiempo de entrenamiento necesario del PLN, mencionada por sus documentos explicativos. Si el aprendizaje de los programadores es lento en comparación con el de los humanos, entonces los neologismos, por su propio carácter dinámico y siempre novedoso, agudizan la dificultad en el veloz adiestramiento que precisa este componente de la IA.

## 6. Consideraciones finales

En los enunciados analizados se activan representaciones sociolingüísticas que, históricamente, han balizado y determinado tanto los rasgos de estilo del discurso científico, en términos de precisión y transparencia; como los proyectos de lenguas y lenguajes universales, en su búsqueda por una regularidad absoluta. Asimismo, más cercanos en el tiempo, resuenan recomendaciones del *lenguaje llano*, en términos de claridad y sencillez. De este modo, se proponen directivas que, con cierta aceptación y consenso en distintos ámbitos, advierten, por un lado, sobre el uso de metáforas, ironías, juegos de palabras y dobles sentidos. Por otro lado, y dada la intención de homogenizar la lengua (en particular, una tan extendida como lo es el español), se recuperan lineamientos del discurso panhispánico en su búsqueda por limitar marcas de diversidad lingüística, tales como acentos regionales, términos locales y coloquiales, expresiones idiomáticas; simplificación que facilita el uso de herramientas informáticas y permite alcanzar, ahorrando costos de producción, a un auditorio amplio y heterogéneo. Paralelamente, el énfasis puesto en el ideograma sobre la complejidad del lenguaje,

manifestado a través de una serie de descripciones estables y recurrentes a lo largo de los documentos analizados, busca justificar las limitaciones y los prolongados tiempos de entrenamiento de los productos del PLN.

Por lo dicho hasta aquí, sostenemos la importancia de atender a los próximos avances en el desarrollo de las tecnologías digitales del lenguaje y, en particular, a su posibilidad de crear o identificar nuevas irregularidades lingüísticas o discursivas. Queda pendiente para futuros estudios analizar cómo interpretan o qué respuestas de desambiguación otorgan estas tecnologías a la utilización de un lenguaje irónico o metafórico, o bien frente al uso de terminología local y neologismos.

Concluimos preguntándonos si, más allá de incidir en la automatización del uso del lenguaje para el diálogo con máquinas, estas consideraciones podrán tener consecuencias en el propio discurso cotidiano de los destinatarios de estos productos. En concreto ¿Qué efectos puede conllevar, por caso, el hecho de concebir un recurso tan enriquecedor en términos cognitivos y conceptuales, como lo es la metáfora, como un obstáculo para la claridad del mensaje? En caso de que el intercambio con estas tecnologías se vuelva continuo ¿Se diferenciarán los usos y formas que impone este diálogo frente a la conversación con otras personas o a la escritura en otros ámbitos? ¿O estas simplificaciones y recomendaciones permearán otras prácticas discursivas? ¿Nos dirigimos hacia la búsqueda de la literalidad absoluta? No podemos anticiparlo ni tampoco será un fenómeno que se puede mensurar o examinar a corto plazo. Lo que sí podemos señalar es que la dimensión prescriptiva que orienta el uso de estas tecnologías se alinea con una serie de discursos y prácticas que apuntan a una homogeneización que involucra cada vez más fenómenos y dimensiones de nuestro discurso, nuestras lenguas y variedades.

## 7. Referencias

- Aguilar, P., Glozman, M., Grondona, A., & Haidar, V. (2014). ¿Qué es un *corpus*? *Entramados y perspectivas*, 4(4), 35-64.
- Alias G., & Cassanelli, R. (2019). *NLP aplicado al análisis del texto*. Universidad de Mar del Plata. [https://www3.fi.mdp.edu.ar/informatica/Trabajos\\_Finales/GAlias-RCassanelli-TFG-II-2019.pdf](https://www3.fi.mdp.edu.ar/informatica/Trabajos_Finales/GAlias-RCassanelli-TFG-II-2019.pdf)
- Alonso, G. (2021). Un legado que proteger y un futuro por escribir: la RAE y el proyecto Lengua Española e Inteligencia Artificial. *Anáfora*, 4, 1-21. <https://revistas.uncu.edu.ar/ojs/index.php/traslaciones/article/view/7396>
- Amazon. (2024). ¿Qué es el Procesamiento del Lenguaje Natural? <https://aws.amazon.com/es/what-is/nlp/>

- Amossy, R., & Herschberg Pierrot, A. (2001). *Estereotipos y Clichés*. Eudeba
- Ananiadou, S., McNaught, J., & Thompson, P. (2012). *The English Language in the Digital Age*. Springer Nature. <http://www.meta-net.eu/whitepapers/e-book/english.pdf>
- Angenot, M. (2010). *El discurso social. Los límites históricos de lo pensable y lo decible*. Editorial Siglo XXI.
- Arduini, S. (2002). Metáfora y Cultura de la traducción. *Revista Electrónica de Estudios Filológicos*, VI, 1-8. <https://www.um.es/tonosdigital/znum4/estudios/metaforacultura.htm>
- Arnoux, E. (2008). *Los discursos sobre la nación y el lenguaje en la formación del Estado (Chile, 1842-1862)*. Estudio glotopolítico. Santiago Arcos.
- Arnoux, E., & Del Valle, J. (2010). Las representaciones ideológicas del lenguaje. Discurso glotopolítico y panhispanismo. *Spanish in Context*, 7(1), 1-24. <https://doi.org/10.1075/sic.7.1.01nar>
- Arnoux, E., & Lauria, D. (2023). La prescripción en los discursos sobre la lengua. En López Ferrero, C., Carranza, I., & van Dijk, T. (eds.), *Handbook of Spanish Discourse Studies* (pp. 129-142). Routledge. <https://doi.org/10.4324/9780367810214-12>
- Authier-Revuz, J. (1984). Hétérogénéité(s) énonciative(s). *Langages*, 73, 98-111. <https://doi.org/10.3406/lgge.1984.1167>
- Báez, P., Arancibia, A., Chaparro, M., Bucarey, T., Núñez, F., & Dunstan, J. (2022). Procesamiento de lenguaje natural para texto clínico en español: el caso de las listas de espera en Chile. *Revista Médica Clínica Las Condes*, 33(6), 576-582. <https://doi.org/10.1016/j.rmcl.2022.10.002>
- Bajtín, M. (1982). *Estética de la creación verbal*. Siglo XXI.
- BAOSS. (2024). *Procesamiento del Lenguaje Natural con Phytón*. <https://www.baoss.es/procesamiento-del-lenguaje-natural-pln-con-python/>
- Baptista, L. (2006). Tratándose de expresiones idiomáticas, ¡no te rompas la cabeza ni busques cinco pies al gato! *Revista Electrónica de Didáctica*, 6, 1-7.
- Becker, L. (2020). Lenguaje claro/llano/ciudadano y lectura fácil: ¿nuevas variedades de comunicación digital de masas más allá del español general/común/total o internacional/ neutro? En Greußlich, S., & Lebsanft, F. (eds.), *El español, lengua pluricéntrica*. (pp. 223-250). Bonn University Press. <https://doi.org/10.14220/9783737009997.223>
- Botpress. (2024). *Procesamiento del Lenguaje Natural y Comprensión del Lenguaje Natural: Guía en profundidad en 2024*. <https://botpress.com/es/blog/guide-to-nlp-nlu>
- Briz, A. (1996). *El español coloquial: situación y uso*. Arco.
- Briz, A. (2002). *El español coloquial en la clase de E/LE un recorrido a través de los textos*. SGEL.

- Bürki, Y. (2023). Retos frente a las tecnologías digitales del lenguaje. Una perspectiva glotopolítica. *Traslaciones*, 10(20), 70-99. <https://revistas.uncu.edu.ar/ojs3/index.php/traslaciones/article/view/7409>
- Cabedo Nebot, A. (2009). Análisis y revisión del sarcasmo y la litote: propuesta desde la Teoría de la Relevancia. *Boletín de Filología*, 44(2), 11-38. <https://doi.org/10.4067/S0718-93032009000100001>
- Calero Vaquera, M. L. (2011). Las irregularidades lingüísticas desde la perspectiva de los inventores de lenguas universales. En Sinner, C., & Zamorano, A. (eds.), *La excepción en la gramática española. Perspectivas de análisis* (pp. 17-36). Iberoamericana/Vervuert. <https://doi.org/10.31819/9783865278722-002>
- Calvet, L. J. (1997). *Las políticas lingüísticas*. Edicial.
- Camacho-Álvarez, M. A., & Navarro Álvarez, E. (2020). Procesamiento del lenguaje natural con Python. *Revista de Cómputo Aplicado*, 4(13), 24-28. <https://doi.org/10.35429/JCA.2020.13.4.24.28>
- Chomsky, N. (2023). The False Promise of ChatGPT. *The New York Times*. <https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html>
- Ciapuscio, G. (2009). Lenguaje y medicina: actividades metalingüísticas en artículos de opinión de mitad del siglo XX. En Eckkrammer, E. (ed.), *La comparación en los lenguajes de especialidad* (pp. 231-241). Frank & Timme.
- Ciapuscio, G. (2011). De metáforas durmientes, endurecidas y nómades: un enfoque lingüístico de las metáforas en la comunicación de la ciencia. *ARBOR*, 187, 89-98. <https://doi.org/10.3989/arbor.2011.747n1010>
- Cortez Vásquez, A., Vega Huerta, H., Pariona Quispe, J., & Huayna, A. (2009). Procesamiento de lenguaje natural. *Revista de Ingeniería de Sistemas e Informática* 6(2), 45-54. <https://doi.org/10.15381/risi.v6i1.3330>
- De Saussure, F. (1945 [1916]). *Curso de lingüística general*. Losada.
- Del Valle, J. (2007). Glotopolítica, ideologías y discurso: categorías para el estudio del estatus simbólico del español. En Del Valle, J. (ed.), *La lengua, ¿patria común?: ideas e ideologías del español* (pp. 13-29). Editorial Iberoamericana. <https://doi.org/10.31819/9783954878741-003>
- Del Valle, J. (2017). La perspectiva glotopolítica y la normatividad. *AGLO*, 1, 17-40. <https://glotopolitica.com/wp-content/uploads/2018/04/aglo20selection.pdf>
- Elastic. (2024). ¿Qué es el procesamiento de lenguaje natural? <https://www.elastic.co/es/what-is/natural-language-processing>
- Eco, U. (2005). *La búsqueda de la lengua perfecta*. Crítica.

- Esparza Urzúa, G. (2021). Alan Turing, bases, forma y críticas a la inteligencia artificial. *Cuadernos salmantinos de filosofía*, 48, 49-74. <https://doi.org/10.36576/summa.144493>
- Freud, S. (1991 [1905]). El chiste y su relación con el inconsciente. *Obras completas VIII*. Amorrortu.
- Gelbukh, A. (2010). Procesamiento de Lenguaje Natural y sus Aplicaciones. *Komputer Sapiens*, II(1), 6-11. <https://www.gelbukh.com/CV/Publicaciones/2010/Procesamiento%20de%20lenguaje%20natural%20y%20sus%20aplicaciones.pdf>
- Ghailani, A. (2019). Aspectos lingüísticos y pragmáticas del chiste. *Hispanista*, XX(77), 1-10. <http://www.hispanista.com.br/artigos%20autores%20e%20pdfs/620.pdf>
- Guerrero-Ramos, G. (2013). El préstamo lingüístico. Uno de los principales procedimientos de creación neológica. *Quaderns de Filologia. Estudis lingüistics*, XVIII, 115-130.
- Guespin, L., & Marcellesi, J. (1986). Pour la Glottopolitique. *Langage*, 83, 5-34. <https://doi.org/10.3406/lgce.1986.2493>
- Haton, J., & Haton, M. (1991). *La inteligencia artificial, una aproximación*. Paidós.
- Haugen, E. (1959). Planning in Modern Norway. *Anthropological Linguistics*, 1(3), 8-21.
- IBM. (2024). International Business Machine Corporation. *¿Qué es el procesamiento del lenguaje natural?* <https://www.ibm.com/es-es/topics/natural-language-processing>
- Iglesias, A. (2024). *Lengua española e inteligencia artificial. Un nuevo horizonte en la regulación de la lengua* [ponencia]. I Congreso Internacional de Lingüística del NOA, Universidad Nacional de Salta, Argentina.
- Kannan, P. (2023). *Addressing Equity in Natural Language Processing of English Dialects*. Stanford University. <https://hai.stanford.edu/news/addressing-equity-natural-language-processing-english-dialects>
- Kerbrat-Orecchioni, C. (1986). *La enunciación. De la subjetividad en el lenguaje*. Editorial Hachette.
- Kloss, K. (1969). *Research Possibilities on Group Bilingualism: A Report*. CIRB.
- Lakoff, G., & Johnson, M. (1986). *Metáforas de la vida cotidiana*. Catedra.
- Lauria, D. (2019). Sobre el programa “Justicia en lenguaje claro” del Ministerio de Justicia y Derechos Humanos de la Nación (Argentina). *Entremeios: Revista de Estudos do Discurso*, 18, 43-61. <https://doi.org/10.20337/ISSN2179-3514revistaENTREMEIOSvol18pagina43a61>
- Lauria, D. (2020). De “el español da batalla” a “la batalla por el español”. Ideologías lingüísticas en la prensa cultural argentina contemporánea: el caso de la Revista Ñ. *La Rivada*, 14, 37-57. <http://www.larivada.com.ar/index.php/ediciones-anteriores/132-numero-14-julio-2020/2-dossier/250-de-el-espanol-da-batalla-a-la-batalla-por-el-espanol>

- Lauria, D., & López García, M. (2009). Instrumentos lingüísticos académicos y norma estándar del español: ¿una nueva política lingüística panhispánica? *Lexis*, 33(1), 49-89. <https://doi.org/10.18800/lexis.200901.003>
- Maingueneau, D. (2009). *Análisis de los textos de comunicación*. Ediciones Nueva Visión.
- Morales Ascencio, B. (1997). La lingüística en el contexto de la inteligencia artificial. *Forma y Función*, 10, 25-50. <https://revistas.unal.edu.co/index.php/formayfuncion/article/view/17072/17928>
- Moreira, D., Cruz, I., González, K., Quirumbay, A., Magallan, C., Guarda, T., Coronel, M., Orozco, J., & Quirumbay, D. (2021). Análisis del estado actual del procesamiento del lenguaje natural. *Revista Ibérica de Sistemas e Tecnologias de Informação*, 42(2), 126-136. <https://www.risti.xyz/issues/ristie42.pdf>
- Moscovici, S. (1961). *La psychanalyse, son image et son public*. Presses Universitaires de France.
- Nerlich, B., & Chamizo Domínguez, P. (1999). Cómo hacer cosas con palabras polisémicas: El uso de la ambigüedad en el lenguaje ordinario. *Contrastes*, IV, 77-96.
- Nogueira, S. (2019). La brevedad en la escritura científica: representaciones de la concisión y la rapidez en la retórica clásica y la de la Ciencia Abierta. *Ciencia e Investigación*, 69(1), 21-32. <https://aargentinapciencias.org/wp-content/uploads/2019/02/3-Nogueira-cei69-1-4.pdf>
- Palacios, C. (2018). ¿De qué hablamos cuando hablamos de humor? Elementos para una teoría general de lo Irrisório. *Luthor*, 35, 46-60. <https://revistaluthor.com.ar/ojs/index.php/luthor/article/view/196>
- Papadimitriou, I., & Manning C. (2021). Language, En *On the Opportunities and Risks of Foundation Models*. Center for Research on Foundation Models (pp. 22-27). Stanford University.
- Rizzo, M. F. (2014). Los discursos de los Congresos Internacionales de la Lengua Española (1997-2007): perspectiva glotopolítica. [tesis de doctorado, Universidad de Buenos Aires]. <http://repositorio.filo.uba.ar/handle/filodigital/1521>
- Rolston, D. (1990). *Principios de inteligencia artificial y sistemas expertos*. McGraw-Hill.
- SAS. (2024). Software de Analítica & Soluciones. *Procesamiento del Lenguaje Natural*. ¿Qué es y por qué es importante? [https://www.sas.com/es\\_ar/insights/analytics/what-is-natural-language-processing-nlp.html](https://www.sas.com/es_ar/insights/analytics/what-is-natural-language-processing-nlp.html)
- Sevilla Muñoz, J. (1999). Divergencias en la traducción de expresiones idiomáticas y refranes (francés-español). [www.deproverbio.com/Dpjournaldp5199sevilla/](http://www.deproverbio.com/Dpjournaldp5199sevilla/)
- Turing, A. (1950). Computing Machinery and Intelligence. *Mind*, LIX(236), 433-460. <https://doi.org/10.1093/mind/LIX.236.433>

- Vanmassenhove, E., Shterionov, D., & Gwilliam, M. (2021). Machine Translationese: Effects of Algorithmic Bias on Linguistic Complexity in Machine Translation. En *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 2203-2213). Cornell University. <https://doi.org/10.18653/v1/2021.eacl-main.188>
- Zuluaga, A. (1980). *Introducción al estudio de las expresiones fijas*. Peter D. Lang.

## NOTAS

---

- 1 Vale aclarar que Guespin y Marcellesi (1986) indican que el término *glotopolítica* puede ser utilizado con dos fines. Por un lado, se dice que la sanción de una ley sobre el lenguaje o la realización de un diccionario es un acto glotopolítico, es decir, articulan el término para caracterizar una práctica. Por otro lado, se lo utiliza para describir o designar un análisis, una lectura o una interpretación, o sea, se puede analizar ese diccionario desde una perspectiva glotopolítica.
- 2 La política lingüística *panhispánica*, ejecutada desde fines del siglo pasado por la RAE y la ASALE (y sostenida financieramente por grandes empresas de capital español —Telefónica, Santander, Repsol—) ha privilegiado el uso de un español general, despojado de localismos (Arnoux & Lauria, 2023). Distintos estudios glotopolíticos han abordado el ideologema del *panhispanismo*, categoría que supone la adopción de una estrategia de unificación entre las diferentes naciones hispanohablantes en pos de un imaginario colectivo común y pluricéntrico, pero que en realidad ha colaborado a desdibujar las cualidades identitarias que la lengua representa (Del Valle, 2007; Arnoux, 2008; Lauria & López García, 2009; Rizzo, 2014; Iglesias, 2024).
- 3 La noción de representación social con la que trabaja la glotopolítica proviene de la psicología social. Se trata de conjuntos sociocognitivos de ideas, creencias, valores, introducidos en situaciones de acción social. Cada individuo o grupo social se apropia de esa representación y la integra a su sistema de valores, dependiente a su vez de su historia y del contexto social e ideológico que lo circunda (Moscovici, 1961; Nogueira, 2019).
- 4 Los documentos explicativos de las siguientes compañías constituyeron el archivo discursivo: Amazon Web Service, AUNOA IA, BAOSS, SAS (Software de Analítica y Soluciones), Botpress, Elastic (The Search AI company), Equisoft, Google Cloud, Hewlett Packard Latinoamérica, IBM (International Business Machine Corporation) y NetApp.
- 5 Debe decirse que algunas de estas páginas (Amazon, SAS) ofrecen múltiples opciones idiomáticas de lectura (alemán, español, francés, inglés, italiano, portugués, turco, etc.), lo que habilita a conjeturar que los enunciados pueden ser producto de un proceso de traducción. Se volverá sobre este punto.
- 6 Además de haber sido identificado como un procedimiento exploratorio prototípico del enfoque glotopolítico en el primer abordaje del *corpus* (Arnoux, 2008), la opción por el análisis enunciativo responde a su profuso dispositivo de categorías analíticas (modalidades, deícticos, subjetivemas, entre otras) que permite no sólo dar cuenta de manera exhaustiva del objeto discursivo *lenguaje*, sino también de la construcción del vínculo enunciator/enunciatario y su incidencia argumentativa en la promoción de los productos del PLN.
- 7 En el estudio de las variedades lingüísticas, se distinguen cuatro tipos, en función de la variable que interviene: las variedades funcionales o diafásicas (los registros de lengua), las variedades socioculturales o diastráticas (los niveles de lengua), las variedades geográficas o diatópicas (los dialectos) y las variedades históricas o diacrónicas.