



PROGETTO  
MAMBRINO

## HISTORIAS FINGIDAS



### El análisis lingüístico en la evolución mundial de los medios de comunicación

Padre Roberto Busa s.j.

(introducción de Stefano Bazzaco)

(traducción de Soledad Castaño Santos)

#### Abstract

El artículo «L'analisi linguistica nell'evoluzione mondiale dei mezzi di comunicazione» de Padre Roberto Busa, pionero de las Humanidades Digitales, fue publicado en origen en el volumen *Almanacco Bompiani: Le applicazioni dei calcolatori elettronici alle scienze morali e alla letteratura* de 1962. A pesar de que el texto fue escrito en una fase preliminar de tales estudios, en que la aplicación de la informática al análisis de las lenguas y literaturas parecía todavía un sueño inalcanzable, la lúcida mirada del jesuita italiano constituye a nuestra manera de ver aún hoy en día un punto de interés para valorar la evolución y los retos futuros en el campo de las Humanidades Digitales. Presentamos aquí la traducción española del artículo con el fin de rescatar las palabras de Padre Busa y ponerlas nuevamente al alcance de un amplio número de lectores.

Palabras clave: Humanidades Digitales; Padre Roberto Busa; *Index Thomisticus*; procesamiento del lenguaje natural; automatización

The article «L'analisi linguistica nell'evoluzione mondiale dei mezzi di comunicazione» by Fr. Roberto Busa, a pioneer of the Digital Humanities, was originally published in the volume *Almanacco Bompiani: Le applicazioni dei calcolatori elettronici alle scienze morali e alla letteratura* of 1962. Despite the fact that the text was written in a preliminary phase of such studies, in which the application of informatics to the analysis of languages and literatures still seemed an unattainable dream, the lucid gaze of the Italian Jesuit in our opinion constitutes even today an observation point of interest to assess the evolution and future challenges in the Digital Humanities field. We present here the Spanish translation of the article with the intention of rescuing the words of Fr. Busa and making them available to a broader number of readers. Keywords: Digital Humanities; Father Roberto Busa; *Index Thomisticus*; Natural Language Processing; automatization

## **Introducción: las Humanidades Digitales vistas desde la lente de Padre Roberto Busa**

El jesuita Padre Roberto Busa, considerado el fundador de las Humanidades Digitales, redactó en 1962 el texto del que aquí proponemos la traducción. El artículo se titula en original «L'analisi linguistica nell'evoluzione mondiale dei mezzi di comunicazione» y fue publicado por primera vez en el *Almanacco Bompiani (1962). Le applicazioni dei calcolatori elettronici alle scienze morali e alla letteratura*, un volumen pionero dedicado a trazar un espacio de convivencia entre los estudios humanísticos y computacionales.

El concepto central que desarrolla Busa en estas páginas concierne el ingreso de la automatización en los ámbitos de las manifestaciones dialógicas humanas. Al respecto, el jesuita registra las nuevas vías que puede abrir esta innovación, señalando los pasos que ya se estaban dando en los campos del procesamiento del lenguaje natural y de la traducción no supervisada. El núcleo de su argumentación es por lo tanto la idea de que los ordenadores, imponiendo la «simbolización del conocimiento», suponen también la formalización del discurso crítico.

En esa época, Padre Roberto Busa estaba trabajando en la indexación de la *opera omnia* de San Tomás de Aquino por medio de fichas perforadas. Este proyecto lo ocupaba ya desde 1949, cuando supo vislumbrar los posibles avances que habría podido traer el tratamiento automatizado del lenguaje gracias al empleo de herramientas informáticas. La intuición, que se fundamentaba en el uso de la tecnología para el análisis del discurso no estructurado, tuvo repercusiones en distintas áreas del conocimiento, desde la lingüística computacional hasta el estudio de los documentos escritos, impulsando en las décadas posteriores la aplicación de técnicas estadísticas y de *machine learning* para la investigación literaria. Y si no sabemos hasta que punto el jesuita pudo tener conciencia del impacto que sus ideas habrían tenido en la mismísima manera de vivir y comunicar por parte de los seres humanos, lo cierto es que en la actualidad parece impensable imaginar nuestra sociedad, y con ella los espacios académicos, sin la presencia de medios informáticos que configuren, fomenten y sostengan nuestras modalidades de expresión y difusión del conocimiento

(Fiormonte, 2017, 16). En los años en que Alan Turing teorizaba el método formal y la noción de algoritmo y Vannevar Bush asentaba las bases del hipertexto<sup>1</sup>, el lanzamiento del proyecto del *Index Thomisticus* por parte de Padre Busa constituyó un verdadero hito para las disciplinas que indagan la textualidad y sus múltiples manifestaciones, porque aseguraba que era posible formalizar, medir y clasificar el lenguaje a través de un ordenador.

Para determinar el impacto de tal intuición, sobre todo en relación con los estudios filológicos y literarios, hay que volver a la época en que el propósito de Busa se presentaba aún como un esbozo, es decir los años finales de la década de los 40, cuando la tecnología estaba ensayando el paso de la forma electro-mécanica a la computación digital *tout court* (Rockwell y Passarotti, 2019, 21). En ese momento liminal, el ordenador todavía no existía en la forma que conocemos hoy en día, sino que se presentaba como una enorme maquinaria de cálculo, con la que se podía comunicar por medio de dispositivos mecánicos como las fichas perforadas (en inglés *punched cards*). Tales fichas, que aparecían como reducidas cédulas en papel rígido con una pequeña brecha en la parte superior derecha y unos agujeritos cuadrados en correspondencia de las palabras-token analizadas, eran las herramientas ideales para ingresar y extraer datos a través del ordenador de forma semi-automatizada. Su producción requería un ingente trabajo manual y representaban una sólida manera de transmitir informaciones entre la máquina y los seres humanos, implicando la integración de dos distintos niveles informativos aptos para la comprensión y manipulación por parte de ambos.

Sin embargo, si la idea de emplear fichas perforadas para el manejo automático de los textos fue presentimiento de Busa, los medios para cumplir con ese sueño visionario fueron proveídos por la IBM (Passarotti, 2018). Thomas J. Watson, director de la empresa por esa época, tras la presentación por parte del jesuita de un boceto del proyecto que habría permitido indexar por medio del ordenador las concordancias en la obra completa de San Tomás, fue perspicaz en aceptar la propuesta. Por un lado Busa ponía las bases de la codificación de caracteres en formato

---

<sup>1</sup> Al respecto se vean Lucía Megías (2012, 43 y ss.) y Tomasi (2022, 20-21).

electrónico, yendo más allá de las diferencias lingüísticas; por otro Watson volcaba su compañía a la experimentación en ese campo, lo cual le habría asegurado notables éxitos y ganancias con la eclosión del *World Wide Web*.

La financiación concedida por la IBM a Busa en 1949, pactada durante un célebre encuentro en Nueva York, fue la primera de una larga serie que duró más de 30 años y que finalizó con la publicación en 1980 de los 56 volúmenes del *Index Thomisticus* (Busa, 1980). En ese largo lapso temporal, en principio a través de fichas perforadas y en un segundo momento gracias a cintas magnéticas oportunamente concebidas para la clasificación de palabras, Busa sentaba las bases de la lingüística computacional como disciplina científica y área de investigación. Sin embargo, su apuesta fue significativa no solamente desde el punto de vista pragmático (agilizando a nivel cuantitativo el procesamiento de los textos), sino en lo que supuso para el horizonte hermenéutico de las disciplinas humanísticas, es decir instituyendo procedimientos y espacios teóricos que consintieran la integración cualitativa de los medios informáticos a las prácticas discursivas de las Humanidades (Testori, 2017).

Las reflexiones de Busa sobre el fenómeno lingüístico, en otras palabras, guiaron los humanistas hacia una comprensión aún más profunda de los mecanismos del lenguaje y de sus niveles informativos, porque los ordenadores requerían un alto grado de formalización de algo que por su naturaleza se escapaba del análisis cuantitativo, es decir la semántica del discurso. Apuntando a nuevas vías para la interpretación de los artefactos textuales, Busa hacía hincapié en las propiedades performativas del lenguaje y arrojaba nueva luz sobre los patrones mentales que guían la producción textual, insistiendo en la posibilidad de transformar enormes moles de datos en conocimiento.

Si tomamos el *Index Thomisticus*, los números del proyecto son impresionantes: 11.000.000 de fichas –una para cada palabra del corpus procesada–, más de 20.000.000 líneas de texto, 70.000 páginas, 56 volúmenes. Estos datos fueron recogidos por más de 70 colaboradores que ocupaban la antigua fábrica textil de Gallarate en provincia de Milán, los cuales gestionaban una cantidad interminable de fichas perforadas de forma manual, señalando concordancias y lemas, tanto que el proyecto se ha descrito recientemente como una de las iniciativas de las Humanidades

Digitales más grande de todos los tiempos, precursora de la lectura distante y del estudio de la textualidad digital como *Big Data* (Rockwell y Passarotti, 2019; Tomasi, 2022, 20).

En los años de máximo esplendor de este proyecto, es decir en la década de los 60, cuando el laboratorio informático de Gallarate estaba en su punto más álgido, Busa redactó el artículo que aquí traducimos. En él introduce unos conceptos de gran interés para las Humanidades Digitales actuales: la transmisión al silicio de los contenidos de los documentos en papel, su sistematización y descripción por medio de metadatos, y su clasificación y explotación con técnicas de *information retrieval*. A la base de todas estas disposiciones subsiste la posibilidad de detectar unas constantes, unos patrones en el lenguaje que cada ser humano emplea para comunicar un determinado asunto. Presenta pues unos temas directamente relacionados con la individuación en el flujo del discurso científico y literario de unos *tics* lingüísticos que consientan ejecutar un estudio estadístico sobre las palabras con las que nos expresamos.

El análisis del discurso por medio de la automatización permitiría, en opinión de Busa, registrar las tensiones del discurso. Y esto porque demostrando ciertas costumbres de la lengua y cuantificando el empleo constante de ciertos términos se pueden registrar los elementos de continuidad entre distintos textos, detectar diferencias entre autores, disponer en un gráfico las palabras y atribuirles cierto grado de afinidad con las pasiones humanas. En suma, se pueden «operacionalizar» (Moretti, 2013) los contenidos informativos para transformarlos en algo más sintético y conmensurable. Padre Busa, en otras palabras, repasando las posibles aplicaciones de la automatización al lenguaje natural, está abriendo nuevos caminos posibles para la investigación humanística, apuntando a los procedimientos de la estilometría, de la lingüística forense, de la atribución autorial y de la *sentiment analysis* soportadas por herramientas computacionales.

Concluye pues estableciendo un paralelo entre la automatización del análisis del lenguaje y la imprenta, lo cual le permite determinar lo que va a ser en el futuro el libro electrónico, es decir un pasaje de *medium* que implica no solamente un salto cuantitativo (con el rápido incremento y almacenamiento de datos), sino también la fijación de un nuevo modelo

cualitativo basado en la interpretación inductiva del lenguaje y en su atenta reproducción en el nuevo contexto. Volcar los textos al ordenador con el respeto de su tradición documental quiere decir actuar de forma sistemática, basarse en la observación empírica y la replicabilidad de las operaciones, demostrar una correcta atención a la calidad final de los datos que ingresamos, que tienen que ser fiables y certificados (Rockwell y Passarotti, 2019, 26-27).

Dispone y ordena lógicamente estos temas una voz espiritual y a la vez extremadamente lúcida en su argumentación, esa voz que en opinión de Passarotti y Nyhan se fundamenta en la necesidad de transmitir una entera visión del mundo frente al *digital turn* (2019, 3), que nos sigue hablando en la actualidad y a la cual pensamos que se debería oportunamente volver.

**«L'analisi linguistica nell'evoluzione mondiale dei mezzi di comunicazione», en *Almanacco Bompiani. Le applicazioni dei calcolatori elettronici alle scienze morali e alla letteratura* (1962), pp. 103-108, 117.**

## I.

1. El fenómeno lingüístico es más grande que nosotros: uno de los ingredientes de esta extraña fórmula de masa que es cada uno de nosotros. Los valores, de hecho, de los que somos un dispositivo tan frágil y maravilloso son, en sí mismos, mucho más amplios y mucho más grandes que nosotros mismos. Las manos, por ejemplo, nos sirven para muchas cosas sencillas o complicadas, pero están, como los camareros, por así decirlo, siempre a nuestras espaldas, las utilizamos sin prestar mucha atención. Pero si las pusiéramos delante de nuestros ojos y las escrutáramos y pensáramos un poco también en ellas, nos encontraríamos frente a todo un mundo por descubrir. Otro misterio es, por ejemplo, nuestra capacidad de gusto estético. ¿En virtud de qué «programa», cargado en ese robot que somos, sentimos la necesidad tan imperiosa, por ejemplo, de la simetría, la repugnancia a cualquier disonancia de color, de línea, de sonido? Pero nuestras verdaderas manos son nuestras capacidades expresivas: con los gestos, con el rostro, con las artes, con las palabras nosotros dañamos y medicamos, derrocamos y elevamos, mejoramos o estropeamos todo a nuestro alrededor. También estas son mundos por explorar en nuestro interior.

2. En nuestro hablar, el que tenemos en la boca y del que sabemos tan poco, hay tres estratos: lo que está presente en el campo de la conciencia, lo que es subconsciente y lo que es de ninguna manera inconsciente. Y en la misma zona de nuestro lenguaje que está iluminada por nuestra percepción y atención, una parte, pero no toda, es pasible de control en el sentido inglés de la palabra, es decir, una parte puede ser gobernada y, por tanto, también educada por nosotros. Es al menos teóricamente posible que un milanés decida acostumbrarse a decir «*vada*» en lugar del incorrecto «*vadi*», ¡al que quién sabe por qué atávica herencia

es tan aficionado! Hay otros sectores que escapan a un control organizativo, pero no del todo a la detección sistemática: no llegamos a cambiarlos, pero aún así nos damos cuenta, aunque sea en cierta medida. Otras zonas, por último, solo obedecen al subconsciente o incluso al inconsciente. Por ejemplo, solo con mucha sutileza llegaremos a darnos cuenta que si nosotros preferimos ciertas palabras a otras, lo hacemos porque nos rige una aspiración subconsciente de causar una supuesta buena impresión, y esto como consecuencia del mayor valor que nosotros adjudicamos a ciertas palabras, así como hacen las señoras con ciertas palabras del tipo «*genare*» o «*flattare*», mientras que otras escogerán las palabras en función de su propia capacidad definatoria y otras, de nuevo, en función de su estética, fonética o semántica o de su correlación y ritmo. Pero en un nivel más profundo, las estructuras gramaticales y sintácticas parecen brotar de las raíces inconscientes con las que la humanidad succiona su evolución vital de ese universo en que se agita, ¡por tan poco tiempo! Como una ameba en su caldo de cultivo. Los fundamentos del lenguaje se encuentran entre las zonas del comportamiento humano, que son inaccesibles a la educación y al autocontrol, porque están programados y comandados exclusivamente por lo que está en las raíces de nuestra fisiología y la desagradable e inevitable mezcla de fisiología y patología.

3. Así que no todo en nuestro hablar se deja conocer: de lo poco o de lo mucho que se permite conocer, no todo está influenciado por nuestra agresiva ambición de hacer también con nosotros y de nosotros «lo que queramos». No obstante, ¿por qué no podríamos dejar también sin atender ese parterre que podríamos, si quisiéramos, desenterrar? ¿Por qué querríamos tomar unos cuantos hilos de agua del río de nuestras palabras y enviarlos por conductos forzados? Exactamente por la misma razón por la que intentamos controlar el agua. El hablar es, de hecho, el principal potencial energético del que el hombre dispone y, por tanto, debe proporcionarse de forma económica. Las ideas son fuerza, solo cuando se pueden decir y escribir. Tampoco tienen ningún otro medio para llegar al individuo que las tiene.

4. Por ello, Aristóteles echó un buen vistazo a su interior y descubrió la metafísica en los pliegues del lenguaje. Y para ese enciclopédico y positivo detector de hechos que ha demostrado ser, esta ha sido una de los más sensacionales: sentirse catapultado desde la pista que había recorrido, palmo a palmo, por la investigación positiva a un plano superior. Hasta el plácido y buen Santo Tomás de Aquino lo contempló, admirando con la nariz en el aire el puente, con el que un pagano, había logrado penetrar en el cielo desde esta tierra. Pero Filón, conocedor del Antiguo Testamento y la Teología Cristiana, partiendo del examen de la «palabra», había penetrado aún más los cielos, superando con creces la gran carrera de Aristóteles y Platón. Entre las palabras habían vislumbrado el destello del Logos, Verbum. Y no ha habido ningún idealista absoluto que haya logrado atreverse tanto como el filósofo cristiano a enuclear el valor de la expresión del «*verbum mentis*» y la consiguiente reverberación del mutuo amor, dentro de ese pensamiento absoluto que es el fuego de la consistencia, la vida y la fantasía: un director que es, al mismo tiempo, un arco voltaico que proyecta sobre la pantalla oscura de la nada esa sucesión de imágenes que somos nosotros, el mundo y la historia.

5. En la vida social, la gramática y el análisis lógico han educado durante muchos siglos ese algo indefinible que nosotros llamamos humanidad y humanismo, ese atisbo de gusto por la belleza, de sentido de la armonía, de apreciación de los valores formales, por el que incluso en el Politécnico subsiste la diferencia entre los que vienen de un liceo clásico y los que vienen de otras escuelas. La retórica había educado en el arte de la expresión. Bien decía el viejo Aristóteles que «*signum scientis est posse docere*», nuestros conocimientos están maduros cuando alcanzamos a transmitirlos. Todos hemos experimentado que cuando el profesor tarda dos horas en hacernos entender algo, es porque aún no lo posee perfectamente, y que para ser capaz de poseerlo perfectamente no debería haber hecho nada más que prepararse para decirlo primero. ¡Cuántas veces hemos visto en la vida la inagotable sabiduría del dicho de que entre el tener razón y el poder tenerla hay una enorme distancia! Hoy en día, a menudo nos preocupamos solo de hacer engullir nociones, como si el hombre fuera un almacén general, o como si no hubiera en él más que

memoria. En cambio, el hombre es, sobre todo, y al menos para su destino, una capacidad organizativa e inventiva, y no necesitaría educarlo como una mochila que se llena según la lista de lo que se necesitará para el campamento, sino que debe ser refinado en sus dispositivos, lubricado, rodado como una máquina-herramienta capaz de trabajar durante mucho tiempo en cualquier material. ¿Me equivoco al pensar que pagaríamos el precio de saber la mitad de lo que sabemos, si pudiéramos decir mejor lo poco que sabemos? Así que el cuidado de nuestros medios expresivos solía existir mucho más en el pasado que en la actualidad. Se enseñó a organizarse internamente para adecuar la combinación de palabras al propósito deseado: a pensar en cómo hablar antes de hacerlo. Pero el mismo Manzoni decía que esta única cosa, «pensar antes de hablar» es tan difícil en sí mismo que también a nosotros se nos debe disculpar un poco por esas tantas veces que nos abandonamos a hablar así como lo hacemos.

6. Poco a poco, las leyes universales del envejecimiento, las cuales afectan a las instituciones tanto como al hombre y a la naturaleza, han desgastado la mordacidad del análisis lógico y de la retórica. El poder de la decadencia ha ejercido su tiranía hasta tal punto que hoy en día solo se va a la escuela de recitación para el teatro, pero no para prepararnos todos para representar nuestro guion en las comedias y las tragedias de la vida. Y si ha escrito que la divinidad del Evangelio se ha demostrado, al menos, su pervivencia en las explicaciones dominicales, es decir, si a vosotros, burgueses, os parece que los curas somos frecuentemente tan descuidados en nuestras predicaciones, es porque nosotros, como vosotros, somos hijos de nuestro tiempo. Y el atardecer se presenta violáceo, también porque en Italia, se está tratando el latín como un viejo abuelo al cual se le desean otros cien años de vida, mientras que el subconsciente registra, piensa que dentro de pronto nos quedaremos sin él, no se siente en el fondo un horror por la verdad infinita.

7. En este punto, el monstruo de la noche, el tecnicismo triunfante, ha intervenido con su última criatura: la automatización. Algunos se estremecieron al pensar que era una excavadora cruda y dura que rugía, aplastando y arrancando las flores. Entre ellos, víctima delicada y gentil, el

humanismo. El mañana ya está aquí. El futuro ya ha comenzado: una colada de lava inunda y quema las verdes laderas de la montaña. En la torre de comando del monstruo, encapsulada entre manómetros, relojes, luces espías y diales, hay algunos hombres. Quizás al principio ni siquiera se dieron cuenta de los agudos gemidos y lamentos elegíacos de los «humanistas». Se contentan con... trabajar. Afirman que prestan un servicio de utilidad pública, porque creen que sin ellos su industria y el comercio no podrían satisfacer las necesidades humanas. Pero entonces—no han pasado todavía diez años— los hombres de la automatización han empezado a sacar la cabeza de la cabina de la torre electrónica, para preguntar a los filólogos y gramáticos, ocupados en los campos recogiendo la flor de las flores, cuestiones de esta naturaleza: ¿Si puede saberse, cuántos verbos transitivos activos hay en ruso, y cuántos intransitivos activos? ¿Cuántos hay en inglés? ¿Cuál es el mayor número de letras iniciales y finales en las que coincide el mayor número de palabras? ¿Qué palabras o situaciones lingüísticas se encuentran entre un radio de *n* palabras, solo cuando y siempre cuando «*faccia*» significa cara, cuáles otras solo y siempre cuando «*faccia*» proviene del verbo *fare*? Y de nuevo: ¿Si puede saberse, podrías agrupar todas las palabras del vocabulario según las distintas categorías morfológicas y gramaticales? Dígame todas las palabras que se pueden omitir, cuando, para acortar un texto sin perjuicio de su expresividad. ¿Pero sabe decirme en cada caso la configuración característica de ciertas categorías semánticas que no son ni morfológicas ni sintácticas ni estructurales? En otras palabras, ha ocurrido un hecho clamoroso, la máquina nos ha hecho conscientes de que ningún humanista posee un lenguaje propio de poder dar respuesta a tales preguntas. La máquina, criada del comercio banal y de la industria burda, ha documentado que todavía hay muy poco humanismo serio y sistemático. Los hechos económicos exigen, hoy en día, un incremento cualitativo de las ciencias gramaticales y léxicas: como una de las necesidades de su desarrollo vital. Pero también ofrecen una posibilidad. No es una pequeña venganza ni una pequeña satisfacción.

## II.

8. El centro de Gallarate es aún hoy el centro que el mundo ha transportado la mayor cantidad de palabras en fichas del mundo: ya hay casi cuatro millones y en continuo aumento. Se trata de 7 lenguas, (Aristóteles, Antiguos Italianos, Dante, Kant, Goethe, Textos hebreos del Mar Muerto, Fabbri, etc.) en tres alfabetos, latino, griego y hebreo. Pero cuando en el 1946 comencé a pensar seriamente en los índices verbales de los trece millones de palabras de Santo Tomás de Aquino, y cuando más tarde, en 1939 inicié los primeros experimentos con el IBM y , de nuevo, cuando en 1951 publiqué los primeros resultados, no solo fui el único y el primero en el mundo que se aventuró a inscribir la lexicología en el hipogrifo, sino que además desconocía el momento histórico en el que esto me sucedía. El haber tenido primero una idea no es un mérito, sino casualidad. Si no me hubiese llegado a mí, seguro que la idea le habría llegado a alguien más. Y quizás un día resulte que antes que a mí se le habría ocurrido a cualquier otra persona, a la que nadie había prestado atención en su momento. Y si se pudiese hablar de mérito, eso consistiría en la larga paciencia requerida para resolver, paso a paso, todas las dificultades e imprevistos que se encuentran al transformar una idea en una metodología: madura y práctica, aplicable por así decirlo, a la producción en serie. De la célebre frase «*genius is one per cent inspiración, ninenty-nine per cent perspiration*» (el genio se compone de uno por ciento de inspiración, un noventa y nueve por ciento de sudor). La única palabra cierta que yo no verifico es la primera. Pero, ¿quién hubiera imaginado entonces que las máquinas de tarjetas se considerarían hoy en día antiguas, y que veríamos la evolución, o más bien la metamorfosis de las computadoras electrónicas desde las memorias de superficies recubiertas de óxido de hierro, a las de entramado de anillos de ferrita, y, finalmente, a las criogénicas (películas muy finas superpuestas como un libro, utilizables a temperaturas cercanas al cero absoluto)? Desde luego, no imaginaba que el «*stretch*» construido para las investigaciones nucleares, tendría una memoria de algo menos de dos millones de posiciones, en la que toda la Enciclopedia Treccani podría nadar como un niño en la cuna, y otra memoria de un millón y medio de posiciones que tienen una

velocidad de conmutación de unas centésimas de milinésima de segundo. Pero, sobre todo, ignoraba que me estaba incluyendo en la sucesión de pasos, a través de los cuales la automatización de la contabilidad provocó la evolución mundial de los medios de comunicación.

**9.** Puedo condensar el movimiento que tomó la aceleración de una avalancha [en la evolución de los medios de comunicación] después de 1945 en cuatro fases. Primera etapa— El desarrollo de las comunicaciones y de las técnicas de organización permitió el crecimiento de las empresas hasta abarcar todo el mundo. Igualmente rápido fue el aumento de la influencia recíproca de los mercados y entre la política y el mercado. Por ello, se ha vuelto indispensable que el gestor pueda registrar un gran número de datos, para elaborar rápidamente resúmenes: a tiempo para controlar y si se desea modificar el curso de grandes masas de pequeños y extensos fenómenos periféricos. Los ordenadores respondieron a esta necesidad aportando a la vida económica la automatización de la contabilidad industrial y comercial. Pueden realizar hasta un millón de multiplicaciones y divisiones por segundo. Pueden imprimir los resultados propios a la velocidad de 60.000 líneas por hora para el alfabeto y 300.000 solo para los números.

**10.** Segunda etapa— La industria, cuyo desarrollo se ve exacerbado por las exigencias de la «defensa», y la intensificación paralela de las relaciones entre la producción industrial y la investigación científica, han impuesto la automatización del cálculo científico. El Euratom, por ejemplo, se ha visto obligado a comprar el ordenador IBM 7090 para su centro de Ispra, que cuesta aproximadamente tres millones de dólares, es decir, casi dos mil millones de liras.

**11.** Tercera etapa— Las actividades de producción, intercambio y defensa requieren la automatización para «*l'information retrieval*» que yo traduciría como la disponibilidad oportuna de conocimientos útiles. La cantidad de publicaciones científicas, ya enorme, está en continuo aumento. Estados Unidos tiene ahora una media de 40.000 patentes al año. Por otra parte, la aceleración de la evolución científica es tal que las

publicaciones en física nuclear después de dos años ya solo sirven para la historia de la física. Pero en lo que respecta a las técnicas informáticas, la actualidad útil de las noticias es probablemente de poco más de medio año. Ahora imaginen que una industria armamentística necesita conocer el comportamiento de ciertos materiales en determinadas nuevas situaciones. ¿Cuánto tiempo tardarán en examinar todas las ciencias relevantes para encontrar lo que le conviene? No necesitarán los índices analíticos, porque, por definición, encontrarán algo que no es comúnmente conocido, ni tampoco le bastarán las indicaciones bibliográficas, porque estas solo contienen los títulos, mientras que usted, por la razón ya expuesta, necesita hurgar en el contenido mismo de lo impreso. Necesitará alguna vez los resúmenes. Pero intenta que se lean todos y ya me dirá si es demasiado tarde cuando termine. ¿Cómo hacer entonces para seguir el ritmo de todas las publicaciones de todo el mundo casi simultáneamente a su aparición? ¡Me parece que el DDT se descubrió por primera vez dos o tres veces seguidas! Por lo tanto, es necesario condensar un máximo de información científica para poder encontrar en un mínimo de tiempo todo lo que interesa en la búsqueda de lo nuevo. El objetivo de la automatización es conseguirlo.

12. Los ámbitos en los que se ha canalizado son: los nuevos tipos de simbolización del conocimiento, es decir, los alfabetos con impresiones magnéticas, cómo transcribir y copiar con estos nuevos alfabetos, que solo la máquina puede leer, el contenido de lo que se imprime con los alfabetos de tinta sobre papel (se esfuerzan para hacerlo mediante la fotolectura, fonografía, etc.); cómo condensarlo (resumirlo, reducirlo al estilo telegráfico, acortar las palabras); cómo clasificarlo; cómo investigarlo. Un capítulo de este esfuerzo se representa en la traducción automática. No me refiero a la ciencia ficción de traducir un texto literario o filosófico a máquina, sino a la técnica de traducir a máquina publicaciones contemporáneas sobre el mismo tema, en las ciencias unificadas tal y como son hoy, por tanto, concebidas y expresadas de la misma manera y con un vocabulario, cuyas únicas diferencias son las de las dos lenguas. Esta técnica añade a los problemas anteriores el de cómo, a partir de situaciones o factores lingüísticos característicos del contexto de una palabra, se puede

identificar automáticamente su función gramatical y lógica y, en los casos de polisemia, su significado en ese preciso lugar; y el de cómo la máquina puede traducir la sintaxis de una lengua a la de otra lengua. La Universidad de Georgetown, en Washington DC, ha abierto hace un año un centro en Fráncfort del Meno donde treinta personas capturan continuamente publicaciones científicas rusas, que son después traducidas al inglés por el ordenador 704.

**13.** Cuarta etapa— La automatización del tratamiento de la información requiere la automatización en la elaboración de índices, concordancias y todo tipo de estadísticas posibles de hechos lingüísticos. En el Euratom en Ispra visite el grupo Tetis. Id a Washington al Instituto de Lenguas y Lingüísticas de Georgetown. Se darán cuenta de que entre los investigadores de las técnicas de tratamiento de la información se está desarrollando una lexicología y una lingüística más sistemáticas, más exhaustivas, más ampliamente útiles, y me atrevo a decir, más humanísticas que las tradicionales hasta ahora. Y dentro de poco, desde los huertos del humanismo, las voces de tenor de los filólogos orquestrarán los méritos de la automatización, comentados baritonalmente por los matemáticos.

### III.

**14.** Pero entonces, incluso dentro de esa máxima expresión de nuestra libertad, personalidad, capricho, que es nuestro hablar, hay formulas matemáticas. Y es cierto: no se puede hablar «como uno quiera», sin obedecer ninguna ley. Si os abandonaseis a la voluntad de sacar, más allá de ciertos límites, del gran mar de combinaciones que son aritméticamente posibles entre los elementos de vuestro vocabulario, ciertas secuencias de palabras que son inusuales a este lado de esos límites, estad seguros que os encerrarían en algún lugar para someteos a la cura de sueño. Pero no solo en este sentido hay leyes en el habla. El número —quién sabe qué alegría sentiría el bueno de Pitágoras si estuviese vivo— ha aparecido como la estructura de soporte del lenguaje, al igual que las

proporciones de las medidas y las relaciones de las proporciones son el esqueleto de la forma y de la belleza. Y la estadística lingüística, que nuestro Davanzati utilizó hace siglos, es tanto más alentadora cuanto que el número sigue reinando entre los fundamentos de las ideas y de la lógica, como lo demuestran la lógica simbólica y la álgebra de las preposiciones, al igual que pertenece a la sustancia del suelo y fuente del ser, como destaca la teología trinitaria católica. Dado que el lenguaje puede traducirse en términos que combinan una gran masa de pequeños elementos, ya que es un entretreído de repeticiones y frecuencias, su matemática no es solo una matemática determinista, sino aún más la matemática de la probabilidad y el azar, una matemática maravillosa que está más cerca del misterio de Dios, del espíritu y del arte. Juan Joaquín Becher, fallecido en 1682, polígrafo, implicado en las hazañas de la teoría flogística, mereció la gratitud imperecedera de su madre Alemania por haberle enseñado a extraer el alcohol de las patatas. Pues bien, un hombre con tan vastos y empíricos intereses podría ser llamado el precursor de la codificación numérica de las palabras.

En su *Character pro notitia linguarum universali*, Francofurti, 1661, proclamó que no hay una sola lengua que pueda ser entendida por todos, salvo que cada concepto se exprese con un numeral o un jeroglífico correspondiente. Esto es todo lo que hace falta –y aún falta mucho y se está trabajando intensamente en ello– para que cualquier ordenador, digital o analógico, pueda servir de traductor fiel y confidencial: ese ordenador que los alemanes llaman Hochgeschwindigkeitstrottel: ¡un idiota de altísima velocidad!

**15.** Para lo que puede servir la estadística de los factores lingüísticos, ampliada tan largamente como lo permitan las increíbles posibilidades de la automatización, pueden ilustrarse los siguientes ejemplos.

En Gallarate, por encargo de los profesores Tagliavini y Croatto de la Universidad de Padua, se realizó la transcripción fonética de un texto de Fabbri de aproximadamente 20.000 palabras de forma automática. Este fue el punto de partida de un censo de los fonemas y los trifenemas del habla italiana. La tesis con la que A. Zampolli presentó las conclusiones, tuvo mucha repercusión, porque finalmente se conocieron los trifenemas

más frecuentes, es decir, los que se combinan para formar el mayor número de palabras. A partir de ahora, la reeducación de los sordomudos se concentrará en estos, para evitar las desgracias que tuvimos nosotros, cuando como de niños nos atiborraron de expresiones francesas (¿Las recordáis? *Hibou, genou, caillou, émail, épouvantail...*) con el resultado de que hoy nosotros poseemos palabras que nunca utilizamos correctamente y cometemos errores en las más comunes.

La proporción del uso de los sustantivos, verbos, adjetivos, preposiciones, etc. oscila en torno a unas bisagras fijas, que varían, sin embargo, según la edad, el sexo, el temperamento, etc. Un censo de estos porcentajes, ampliado en los discursos y a la composición de miles de alumnos de diferentes entornos –extensión que solo la automatización hace posible– permitiría identificar curvas de normalidad, que servirían como ayuda para el diagnóstico de la psique del hombre en la edad en que es más susceptible al influjo educativo.

Este verano los periódicos dieron cobertura mundial a las conclusiones del inventario de la métrica de la *Iliada* realizado en Nueva York por James Mc Donough. La *Iliada* surgió como obra de un solo autor. Fui yo quien, hace años, le dio a este joven estudioso americano su primera oportunidad. Él comenzó entonces a marcar en fichas solo el número de sílabas de todos los versos. Cuando se transcribió así toda la *Iliada*, un ordenador reveló los ritmos y las proporciones del uso de los distintos metros. Para hacerlo a mano, a parte del tiempo que habría llevado, no habría habido más posibilidad de control que volver a hacerlo de la misma manera. Pero hoy en día, si no queréis creer en las conclusiones, podréis volver a comprobar todos los cálculos en unos minutos desde las fichas iniciales.

La cronología de las obras platónicas ha sido reconstruida en el pasado, y hoy en día sigue fuera de discusión, precisamente mediante la estadística de los estilemas, aunque realizada sin ayuda de máquinas automáticas.

Un procedimiento similar podría utilizarse para abordar la controversia sobre la autenticidad de ciertos escritos, por ejemplo, Shakespeare o Marlowe: precisamente y siempre porque en el estilo de cualquiera hay situaciones que son sus características personales y

permanentes, no menos de sus huellas dactilares. Lo cual es bastante obvio, si se reflexiona que todo lo que expresamos, está siempre con nosotros mismos que lo expresamos. Existen estudios sobre la agudeza de los acentos tónicos. Las palabras con acento tónico en la *i* y la *e* dan lugar a sentimientos elevados y agudos; las que tienen acento tónico en la *o* o la *u*, expresan sentimientos depresivos; y las que tienen el acento en la *a*, sentimientos neutros. Tomamos un paisaje de *Los novios* y su traducción al francés. El flujo de la agudeza de los respectivos acentos tónicos se informó con curvas en papel milimetrado. La sucesión de los distintos niveles emocionales expresados por las palabras del texto se codificó en curvas en otro papel milimetrado. El resultado fue que el curso de la agudeza de los acentos en el texto original coincide con la curva descrita por el sentimiento. No es el caso de la traducción francesa, en la que el ritmo fonético de los acentos no fluyó inconscientemente de la inspiración interior ni se tuvo en cuenta como elemento a «traducir».

#### IV.

16. En la época de Gutenberg, junto a los manuscritos que permanecían en forma de cuadernos y registros, surgió el libro impreso. Hoy en día, junto a estos y a este, que permanecerá, se encuentra el «libro magnético». Y para el depósito de los conocimientos humanos, esto representa un verdadero cambio de dimensión. Pero no solo es cuantitativo, ni solo en términos de velocidad. También es cualitativo. Si bien es cierto que el lenguaje de las calculadoras electrónicas marcará con toda probabilidad el fin de los intentos de lenguas universales artificiales, también es cierto que la interpretación inductiva del fenómeno lingüístico a través de las fórmulas de la probabilidad (evolución del lenguaje hacia nuevas especificaciones y al mismo tiempo su involución o entropía hacia pérdidas graduales de semanticidad: aquí también leyes de mezcla de la vida y la muerte), esta inducción, digo, en la medida en la que la automatización lo hace posible, promete reiniciar el ciclo de la conciencia

lingüística y gramatical con mayor profundidad sistemática y documentación.

17. Uno de los signos que hoy se encuentra en un punto de inflexión es el hecho de que hay en el mundo aproximadamente 200 centros ocupados en este derrocamiento de la torre de Babel, esto es, en salvaguardar en el lenguaje la fisiología unificadora de la comunicación y en frenar la patología de las barreras y la separación. Entre ellos, una docena, después del de Gallarate, se ocupan en estos momentos del campo del análisis léxico puro. Los demás lo aplican a las técnicas de información y traducción automática. De hecho, el desarrollo de la automatización lingüística es triangular. Otra señal es que instituciones como los ministerios de comercio o de defensa y otros –Estados Unidos, URSS, Nato, Euratom, etc.– lo financian desde hace algunos años. En Francia, Holanda, Israel, Checoslovaquia, están en marcha proyectos de gigantescas elaboraciones electrónicas –se habla de 120 millones de fichas para el *Trésor de la langue Française*– con el fin de disponer de materiales para la compilación de diccionarios históricos de la lengua nacional. Por otra parte, incluso el expediente del *Index Thomisticus*, que se está elaborando en Gallarate, podría definirse como el primer Tesoro del lenguaje científico de nuestra Edad Media.

18. Domenico De Domenichi, un veneciano, «*de ordine plebeio*», se convirtió en el vicario del Papa Sixto IV. En el prefacio de un incunable estampado en Venecia en 1480, comenta a la muy reciente invención de la imprenta: «*Placuit autem clementissimo Deo his nostris temporibus novam artem docere homines*». A continuación, informa de la asombrosa noticia de que tres hombres, en solo tres meses de trabajo, consiguieron imprimir 300 ejemplares del volumen: «*ad quae nota eorum vita hand quaquam sufficeret si cum digitis et cum calamo aut penna scribenda forent*» y luego concluye: «*si quid in me auctoritatis etiam admoneo: ne tanta Dei beneficencia abutantur*» ¿Qué debemos decir hoy?

### Bibliografía citada

- Busa, Roberto, «The annals of humanities computing: The Index Thomisticus», *Computers and the Humanities*, 14/2 (1980), 83-90.
- Fiormonte, Domenico, *Per una critica del testo digitale. Letteratura, filologia e rete*, Roma, Bulzoni, 2017.
- Lucía Megías, José Manuel, *Elogio del texto digital: Claves para interpretar el nuevo paradigma*, Madrid, Fórcola, 2012.
- Moretti, Franco, «Operationalizing: Or, the Function of Measurement in Literary Theory», *New Left Review*, II, 84 (2013), 103-19.
- Passarotti, Marco, «Padre Busa, il gesuita che inventò l'ipertesto grazie ai computer IBM», *IBM thinkMagazine*, 2017 <<https://www.ibm.com/easytools/runtime/hspx/prod/public/X0027/PortalX/page/pageTemplate?s=78c374df5c884363b46454a5ffefb5d9&c=6623351d59604a11b2c845760f87280f>> (cons. 03/06/2022).
- Passarotti, Marco y Julianne Nyhan, «Introduction, or Why Busa still matters», en *One origin of Digital Humanities. Fr Roberto Busa in his own words*, Springer, 2019, 1-17. DOI: <<https://doi.org/10.1007/978-3-030-18313-4>> (cons. 03/06/2022).
- Rockwell, Geoffrey y Marco Passarotti, «The Index Thomisticus as a Digital Humanities Big Data Project», *Umanistica Digitale*, 5 (2019), 13-34.
- Testori, Marinella, «Methods of quality, quality of methods. What does Roberto Busa have to communicate to digital humanists in the 21st century? From hermeneutics to performativity», *Digital Humanities Quarterly*, 11/3 (2017) <<http://www.digitalhumanities.org/dhq/vol/11/3/000329/000329.html>> (cons. 03/06/2022).
- Tomasi, Francesca, *Organizzare la conoscenza: Digital Humanities e Web semantico*, Editrice Bibliografica, Milano, 2022.