

COMENTARIO:

LA COMPUTADORA AL SERVICIO DE LA FILOSOFIA

Profesor Daniel Herrera R.

Entre los años 1973 y 1976 se han publicado los primeros 31 volúmenes del Index Thomisticus (Fromann Verlag-Gunther, Stuttgart) procesado mediante computación electrónica. La obra comprende todas y cada una de las palabras y frases hechas contenidas en las obras completas de Santo Tomás y con la indicación de la frecuencia con que aparecen y los diferentes contextos en que ocurren.

Para los investigadores no se trata de una obra más que se deba añadir a la ya extensa bibliografía tomista, sino de un instrumento indispensable para el análisis filosófico y lexical de los textos que deben fundamentar toda interpretación de la doctrina del Angélico. Por otra parte, como trabajo pionero en la utilización de las computadoras dentro del campo de la filosofía vale la pena conocer en detalle el proceso de realización de esta obra.

Los autores del proyecto.

La primera idea del Index se debe al Padre Roberto Busa, S.J. profesor de filosofía en el Instituto Aloisiano de Gallarate, quien en 1946 entró en contacto con el centro IBM de aquella localidad. Durante tres años el proyecto fue consultado con expertos de diversos países. En 1949 el análisis de los textos comenzó a procesarse en el centro IBM de Gallarate ininterrumpidamente hasta 1966. Se continuó durante dos años en el Centro Nacional Universitario de Cálculo de Pisa y durante quince meses en el laboratorio IBM de Boulder (USA). A partir de 1971 el procesamiento se continúa haciendo en el Centro de Investigación de la IBM en Venecia.

En la elaboración del Index han intervenido expertos de diferentes especialidades: especialistas en la automatización del análisis lingüístico, filósofos y teólogos, filólogos y lingüistas, los autores de la edición crítica de las obras de Santo Tomás, los editores del Thesaurus Linguae Latinae de Munchen, la Mediaeval Academy of America y la American Philological Association.

Contenido de la obra.

La obra contiene el vocabulario empleado en 179 obras latinas escritas entre el siglo IX y el siglo XVI lo que ocupa 1.700.000 líneas de texto, con un total de 10.600.085 palabras.

Los escritos están distribuidos en tres grupos. El primero comprende 49 obras del Santo. El segundo está integrado por los comentarios del Angélico a las obras de Aristóteles (13), a la Biblia (10) y a otros autores (4) como Boecio, Dionisio y al autor del "Liber de Causis". El tercer grupo está compuesto por las "Reportaciones", es decir, obras copiadas por alumnos mientras el Santo explicaba la clase o predicaba en el templo. Se incluyen además 61 obras de otros autores desde Aimón de Auxerre (+ 850) hasta Cayetano (+ 1553) y otras 27 obras de autores desconocidos. La razón de incorporar en el Index las palabras contenidas en estas obras, obedece al hecho de que, al menos en épocas pasadas, formaban parte del Corpus Thomisticum, ya porque erróneamente fueron atribuidas a Santo Tomás o porque completaban las obras dejadas incompletas por el Maestro.

Finalidad de la obra.

La finalidad inmediata de la obra es el establecimiento de una base léxica como instrumento previo para la interpretación del pensamiento del Angélico. Sin embargo, dada la extensión del campo que abarca la obra y la riqueza de sus análisis el Index llena otras finalidades que quisiéramos destacar.

Los diccionarios del latín medieval existentes sólo contienen el significado de los términos usados hasta el siglo XII. El Index es el primero que presenta aquella latinidad que constituyó la lengua científica internacional desde el siglo XII hasta el siglo XIX. Por consiguiente, constituye un extraordinario auxiliar para el estudio de obras como las de Copérnico, Newton, Descartes, Leibniz, Spinoza, Kant que fueron escritas en latín. El Index puede ser considerado como una base para un diccionario histórico del Latín de los sabios. Debido a que todavía no existe un diccionario de estas características, los diccionarios históricos de las lenguas modernas no llegan a superar una terminología específicamente literaria y no pueden abarcar

las terminologías filosóficas, científica y tecnológica, las cuales no pueden ser identificadas si no se identifica antes el léxico del que se derivan y con el cual están interrelacionadas.

Una parte importante del Index es el "diccionario electrónico latino" instrumento original que contribuirá notablemente a dar al estudio científico del latín una profundidad y una totalidad de sistematización estructural que hasta ahora hubiera sido imposible imaginar. Servirá también como instrumento de comprobada validez para realizar estudios acerca de cualquier otro autor, incluidos los clásicos latinos.

Por razón de su contenido y de su metodología el Index puede servir tanto para los estudios de lingüística pura como de lingüística aplicada. Los datos estadísticos que proporciona son extremadamente valiosos como ayuda comparativa para traductores o para quienes se dedican a la ordenación automática. Aunque las categorías y proporciones lingüísticas propuestas difieren de las que se encuentran v.g. en la lengua de los textos científicos actuales, sin embargo, en cuanto a la parte en que se corresponden y que es bastante considerable, constituyen una confirmación válida de que son las fórmulas y las partes comunes a toda lengua como tal. Además, si se compara el léxico tomista con los de otros autores, respecto a materias idénticas el Index aporta una contribución importante a los estudios acerca de la crítica textual, en cuanto autenticidad, cronología y estilística y proporciona abundantes datos estadísticos sumamente valiosos.

Para comprobar los métodos y estructuras del Index y para permitir que sea utilizado más ampliamente en todas las ramas de la lingüística aplicada, se han realizado investigaciones paralelas en el Centro de Gallarate sobre otras ocho lenguas: hebreo, arameo, moabiteo, griego, inglés, italiano, ruso, alemán y otros cuatro alfabetos, sobre textos que contienen más de cuatro millones de palabras que abarcan textos variados como los manuscritos de Qumram, los prolegómenos de Kant, textos italianos de diferentes siglos, resúmenes de física nuclear y artículos rusos de física y química.

Las máquinas y los hombres.

El uso de las técnicas electrónicas para procesar datos ha sido un factor determinante para realizar el Index. Sin embargo no ha podido anular la contribución de los sabios.

El trabajo de las máquinas se ha utilizado para codificar cada palabra según el tipo de frase para la lematización de cada palabra, la distinción de los homógrafos, la codificación de las palabras de acuerdo a su "peso" en los diversos tipos de concordancia, el léxico electrónico, la definición del sistema de homografía y de polimorfía, el inventario de todo el vocabulario,

primero a nivel de formas y luego a nivel de lemas, la adaptación de los índices estadísticos a las concordancias, la subdivisión de los índices, la estructuración de las diversas maneras de concordar las palabras de un texto, la contextualización "razonada" automáticamente y la eliminación de redundancia contenida "ad modum unius", las frases compuestas de varias palabras.

Es interesante saber que las páginas para la imprenta pasan directamente de la computadora a una fotocomponedora experimental que imprime directamente los films necesarios para la impresión en offset. La fotocomponedora está regida por la banda magnética sobre la que está registrado el texto. Como el film es una matriz perfectamente compuesta, se suprime el largo y penoso trabajo de revisión y de la corrección de pruebas y además se logra una excelente presentación de formato.

La ayuda casi mágica de la máquina todavía tiene que completarse con el trabajo de los sabios y los compiladores en una parte esencial que comprende cuatro fases: a) aplicar al texto, frase por frase, el signo de "tipo de la frase" junto con otros signos de código; b) perforar el texto sobre las tarjetas, registrando sobre la banda magnética y corregir los errores cometidos en la transcripción; c) compilar el diccionario electrónico latino; y d) organizar la programación y ejecución de las operaciones de la máquina.

Para tener una idea del trabajo paciente que supone esta obra gigantesca, debemos considerar que el 1.700.000 líneas del texto han sido examinadas una a una, palabra por palabra, letra por letra, por lo menos ocho veces: dos veces para la redacción, dos veces para la perforación y la verificación, dos veces para control integral mediante la lectura, una vez, por lo menos, para la lematización y selección de homógrafos, y otra vez, por lo menos, para las correcciones finales.

La composición del Léxico electrónico latino exigió registrar en banda magnética unos 90.000 lemas del Lexicon Totius Latinitatis de Forcellini. Más de 10.000.000 de códigos fueron escritos a mano, perforados, transcritos en la banda magnética y finalmente examinados y corregidos nuevamente. Sólo esta fase requirió más de 100.000 horas de trabajo.

Estructura de la obra.

La obra se compone de dos partes principales: los Índices (recapitulativos de distribución, de vocabulario, de frecuencia, índices individuales) y las Concordancias de todas las obras analizadas.

En los primeros 31 volúmenes publicados encontramos lo siguiente:

Los ocho volúmenes introductorios contiene los cuatro primeros Indices, o sea, las Tablas que indican la distribución de todas las palabras (147.080 formas y 25.018 lemas) en el conjunto de las obras. En detalle estos primeros ocho volúmenes nos ofrecen:

1) Indices de distribución de los lemas y de las formas

2) Sumarios del léxico:

- a) paradigma de los lemas (contiene un repertorio de lemas codificados morfológicamente y articulados para la flexión automática, desarrollados según su polimorfía, con la lista de especificativos, la frecuencia total y el porcentaje en que aparecen, en primer lugar en las tres subdivisiones de obras del autor -obras, comentarios y referencias luego en su conjuntos y finalmente en las obras de otros autores-),
- b) racionario de verbos: repertorio de formas agrupadas según su lema, con los códigos de su morfología, lematización, tipo de concordancia, homografía y polimorfía,
- c) repertorio de formas dispuesto por orden alfabético, directo, es decir, de izquierda a derecha,
- d) repertorio alineado a la derecha y el orden alfabético inverso, es decir, de derecha a izquierda de la primera palabra, con los códigos morfológicos del lema y su número de código, con los dos totales de conjunto.
- e) índice contrario de formas, como el repertorio, sólo que las formas están colocadas a la derecha y en orden alfabético inverso.

3) Indices de frecuencia: lemas y formas según frecuencia decreciente, prospecto comparativo de los lemas que aparecen más frecuentemente en las obras mayores, lemas y formas según códigos morfológicos, homografías y desinencias latinas.

Los otros 23 volúmenes (26.60 páginas) contienen la primera concordancia de las obras de Santo Tomás; 7.575.841 líneas representan los contextos de 2.575.972 palabras clave (los sustantivos y verbos "comunes"). Esta concordancia es la primera de la automación; es también la más voluminosa y la más rica de las publicadas. En ella se encuentran codificadas palabra por palabra la tipología del razonamiento, las categorías morfológicas y léxicas y las expresiones llamadas "sintagmas".

Están proyectados otros 25 volúmenes que contendrán: 1) la Concordancia segunda de las obras del Santo que comprenderá, en 15 volúmenes, cerca de 6.000.000 de palabras tomadas de dichas obras, a saber: las palabras invariables como preposiciones y conjunciones; los pronombres y adjetivos pronominales, palabras especiales como numerales, transliteraciones, abreviaturas, etc. los verbos auxiliares y serviles; todas las palabras que ocurren en citas literales, las palabras usadas en las citas. 2) Las Concordancias Primera y Segunda de las obras de otros autores, las cuales tienen el mismo formato de las dos Concordancias tomasistas. 3) Siguen otras 16 Tablas de resumen y estadística: cuál es la palabra más frecuente, cuál es la que se usa una sola vez, número de palabras por categoría morfológica, etc.

La obra se presenta en edición bilingüe latino inglesa. La edición, dada su naturaleza, es reducida: se publican sólo 500 ejemplares, pero se conservan las matrices para repetir la edición en el futuro. Hasta ahora se han suscritos a la obra 200 instituciones en donde existe verdadera investigación.

Si se atiende a la estructura de la obra se puede observar que cada palabra está acompañada no sólo por signos de puntuación y por una referencia, sino también por indicaciones de código "especificantes" que permiten una investigación y una apreciación más exactas. Estos códigos indican si la palabra se encuentra en citas literales o en citas que dan solamente el sentido, o si es el autor mismo el que expresa su propia idea: en este caso los códigos precisan si esta palabra se encuentra en las referencias a obras de otros autores, o a referencias o otras partes de su propia obra, o a niveles secundarios de exposición como son los títulos, las conclusiones, conexiones de frases o subdivisiones de la materia o, finalmente, si se encuentra en una frase que representa lo esencial de lo que el autor quiere comunicar.

Cada palabra está "lematizada", es decir, codificada según su morfología y designada por su lema.

No obstante su complejidad la obra cuenta con índices recapitulativos que hacen bastante fácil su manejo.

Conclusión:

Los datos anteriores son suficientes para que tengamos una idea de la empresa gigantesca acometida por los editores de este Index Thomisticus, de su valor intrínseco, de su significado para un trabajo investigativo en diversos campos.

¿Podría un día la Universidad del Valle poner al servicio del Departamento de Filosofía, del programa de Magister en Lingüística y de nuestro Centro de Computación este incomparable instrumento de investigación? Recordemos que la edición es bilingüe: latino-inglesa.