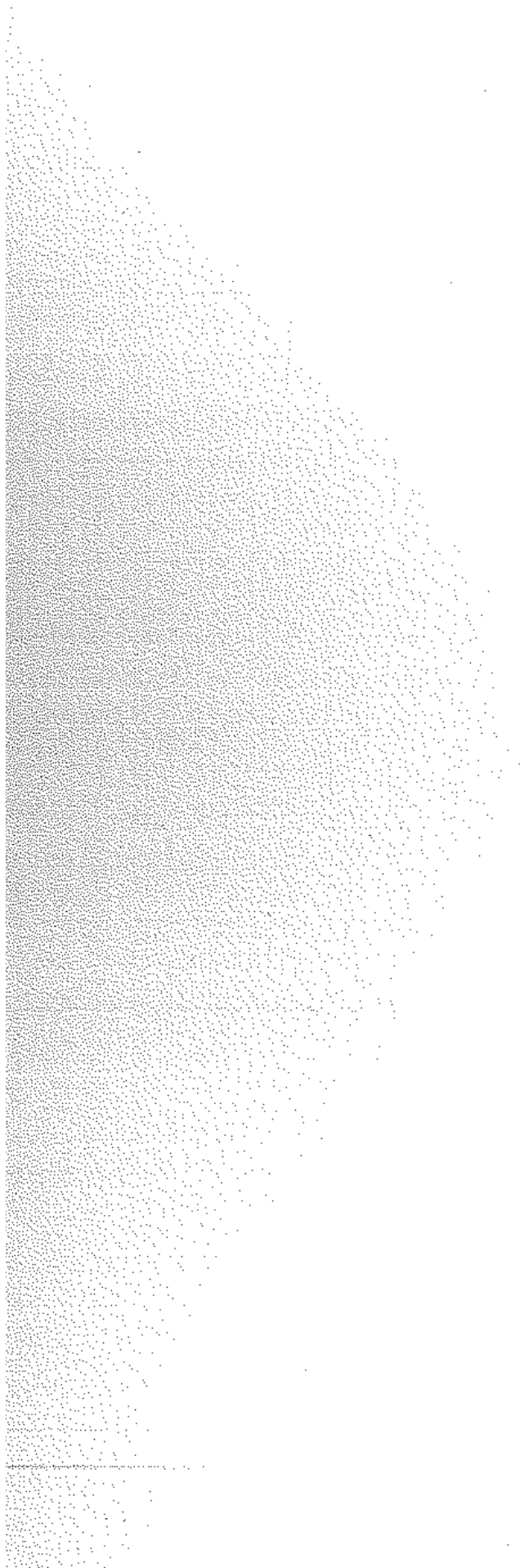


F. RODRIGUEZ ADRADOS

UTILIZACION DE ORDENADORES
EN
PROBLEMAS DE LINGÜISTICA

SEPARATA DE LA REVISTA DE LA UNIVERSIDAD COMPLUTENSE
Volumen XXV - Número 102 - Marzo-Abril 1976



UTILIZACION DE ORDENADORES EN PROBLEMAS DE LINGÜÍSTICA

F. RODRÍGUEZ ADRADOS

Facultad de Filología
Universidad Complutense

A la amable insistencia del profesor Torroja, Vicerrector de Investigación Científica de la Universidad Complutense, y de varios colegas se debe el hecho de que sea yo quien presente hoy este coloquio sobre la utilización de ordenadores en problemas de lingüística. Confieso que acepté este encargo no sin vacilación. Mi relación con el tema es más la propia del que teoriza desde puntos de vista de Lingüística General y querría aplicar el nuevo método a viejos problemas, que la del especialista propiamente dicho. Lo mismo en el conocimiento de los ordenadores que en el uso práctico de los mismos en problemas lingüísticos concretos hay en esta sala personas de la mayor competencia que habrían podido sustituirme con ventaja en esta presentación.

Al menos, al interés teórico por el tema desde ese punto de vista general y al haber intentado en ocasiones tratamientos de problemas lingüísticos por medio de despojos exhaustivos, siquiera en la intención, de materiales, se une en mí otra circunstancia que tal vez justifique el que esté hablándoles en este momento, y es el haber estado presente desde el comienzo en la organización del Simposio. Este procede de una idea del Doctor Torroja y forma serie con otros anteriores celebrados en la Universidad Complutense, de entre los cuales me es grato recordar el que tuvo lugar en diciembre de 1973 sobre «Teoría de Sistemas», en el cual la Lingüística ocupó un lugar. Tal vez de allí

provenga la idea de dedicar a la Lingüística un Simposio completo. El tema se concretó en una reunión que una serie de colegas celebramos en Octubre pasado y la organización fue tomando forma a lo largo de una serie de reuniones más.

Realmente, la idea de organizar un Simposio sobre Lingüística y ordenadores no puede ser más oportuna. La aplicación de los ordenadores a la Lingüística y a los estudios de Humanidades en general ha tomado fuera de España un incremento muy grande en muy pocos años. Es bien sabido cómo en Francia, en Italia, en Bélgica, en Estados Unidos, en otros países funcionan Centros dedicados especialmente a estas tareas; centros algunos de los cuales están hoy aquí representados y de cuyas actividades vamos a tener en estos días noticia detallada. Esto es muy grado para nosotros y cumplo con un deber de cortesía saludando aquí a sus representantes, de los que tanto esperamos aprender, y deseándoles una grata estancia entre nosotros.

Tampoco en España son ya hoy una novedad estos estudios. A los esfuerzos del Centro de Cálculo de la Universidad Complutense, al que todos los interesados en este campo debemos gratitud, se une el de otras instituciones; y se une también el empeño de una serie de lingüistas a algunos de los cuales tendremos estos días ocasión de escuchar y que nos informarán de sus trabajos. No hemos querido reducir este Simposio a la Universidad Complutense, sino acoger a todo el que, pertenezca a esta Universidad o a otra, tenga algo que decirnos, pues se trata, precisamente, de establecer un contacto entre los diversos cultivadores de este campo de estudios.

Pues la oportunidad de este Simposio consiste precisamente en dar, por así decirlo, estado oficial dentro de España a unos estudios que han ido cultivándose aquí y allá, en forma aislada e independiente. Pienso que los interesados en este tema aprenderemos mucho unos de otros y aprenderemos también de nuestros huéspedes extranjeros. Precisamente las diferencias entre los intereses —cultivo de la fonética y la lexicografía, la métrica y la dialectología, la semántica y la estilística, lenguas antiguas y modernas, etc.— hace más útil y necesario intercambiar experiencias sobre un campo en el fondo común. Y cambiarlas también con los especialistas en los propios ordenadores o en la Lógica. Así podrá mostrarse también a las personas que ahora se acercan con interés al tema de la aplicación de los ordenadores a la Lingüística, cómo la nueva técnica y las viejas Humanidades pueden colaborar en empresas comunes.

En efecto, ningún otro resultado del coloquio podría superar al de que despertase, sobre todo entre las nuevas generaciones de lingüistas, un interés que las hiciera adentrarse en este campo con una preparación y una seguridad mayor de las que nosotros, los de la generación anterior, que hemos conocido ya tantos modos y modas de la Ciencia lingüística, podemos desplegar.

Y, sin embargo, querría dejar constancia aquí, a riesgo de decir cosas que muchos de los presentes verán con razón como verdades perogrullescas, de que la novedad del nuevo método es relativa, de que en el fondo no hace más que hacer posible la continuación de investigaciones anteriores con instrumentos más eficaces. Pienso que esta comparación de la Lingüística anterior —escindida en tantas escuelas— y la que podemos esperar que nazca de los ordenadores es la que tiene mayor interés en este momento introductorio del coloquio, más que una relación de detalle sobre aspectos de la Lingüística computacional que encontrarán un lugar más adecuado en las distintas ponencias.

Efectivamente, el uso del ordenador lo que hace, ante todo, es perfeccionar nuestras posibilidades de reunir inventarios de datos más completos y accesibles que los actuales; de obtener una base más sólida para la estadística lingüística, y de obtener, igualmente, una serie de aplicaciones prácticas. En los tres casos, continúa métodos de trabajo lingüístico ya conocidos, pero los lleva más allá. Veamos los tres puntos uno tras otro.

Los inventarios de datos lingüísticos o aplicables al estudio lingüístico que se hacen con el ordenador en nada difieren de los que se realizan en otros campos, incluidas las Humanidades: técnicas de almacenamiento de datos bibliográficos y otra información, de datos relativos a la forma de los vasos griegos o los cánones de la estatuaria de diversas épocas, de otros sobre instituciones jurídicas medievales, por ejemplo. Todo esto ya se hacía, mediante sistemas de fichas o de listas, sólo que ahora puede hacerse con mayor exhaustividad y pueden pedirse al ordenador todos esos datos por orden de frecuencia o en correlaciones diversas unos con otros o en clasificaciones y tabulaciones varias.

Tampoco es nueva la confección de índices de palabras en orden directo o inverso o de concordancias ni la de listas de elementos fonéticos o morfológicos o métricos, etc., etc., ordenados con diferentes criterios, ni de mapas de rasgos dialectales. Pero, una vez más, la

exhaustividad de estas listas y su manejabilidad es mucho mayor. Por ejemplo, los índices del italiano antiguo del SEIOD contienen en cada palabra toda clase de indicaciones de frecuencia y gramaticales que permiten un tratamiento para una larga serie de estudios lingüísticos diversos. Un índice del vocabulario micénico como el que está en proyecto, une a la totalidad de los datos lingüísticos otros de tipo arqueológico que resultan de utilidad para valorar los primeros. Se puede, también, intentar lo que antes no era posible mediante el establecimiento de las latitudes combinatorias de elementos de un texto extenso: esto es precisamente lo que proponemos en nuestra ponencia sobre ordenadores y semántica, en relación con las latitudes combinatorias tanto distribucionales como opositivas y transformacionales de los lexemas y de los morfemas gramaticales.

La aplicación de los ordenadores a la Estadística lingüística, que es el segundo punto a que nos referíamos, nos lleva una vez más a un campo ya trabajado antes: baste citar los nombres de Herdan, Guiraud, Muller y tantos más. Era bien conocida la relevancia del factor frecuencia a efectos de definición del estilo, de la autenticidad de las obras, de la datación de las mismas, y del estudio en sí del Léxico, la Sintaxis o la Métrica. En definitiva, lo que Saussure llamaba *lanque* se define por su frecuencia superior, mientras que las frecuencias inferiores son las propias de los usos marginales o desviados. Pues bien, el ordenador no hace otra cosa que ahorrar tiempo al matemático que calcula a partir de los datos; y, añadamos, que facilitar la recolección de los mismos y su clasificación. Puede decirse que en la Estadística lingüística está el primer modelo de lo que puede hacerse con un ordenador en Lingüística.

Otro caso en que los ordenadores se aplican a mejorar técnicas ya existentes, es el de las que llamaríamos aplicaciones prácticas, a las que nos hemos referido en tercer lugar. Aludimos con esto a campos tan variados como la cartografía lingüística, la seriación y filiación de manuscritos en Crítica Textual, el desciframiento de lenguas, la traducción automática. Por muchos que sean los problemas y los abusos que a veces se han cometido (luego diremos algo de ello), es evidente que los ingentes recursos de los ordenadores para memorizar, computar y clasificar datos, xplorando sus latitudes combinatorias, para representar sus combinaciones también, pueden contribuir a presentarnos una idea realista de los mismos, a clasificar los conjuntos en que se integran, a establecer invariantes y relaciones fijas: todo lo cual es una

ayuda inestimable en los terrenos de estudio mencionados y en otros muchos más.

El ordenador, como otra máquina cualquiera, es un servidor del hombre en tantos campos y también en el de la Lingüística. No sólo responde a las preguntas que se le hacen, si están bien hechas, sino que, gracias a sus capacidades, puede sugerir que se le hagan preguntas que, en otro caso, ni siquiera haríamos, pues al saber consciente o inconscientemente la dificultad de lograr una respuesta, no se nos ocurrirían.

El ordenador puede sentar precisiones concretas en vez de las vagas afirmaciones sobre la baja o alta frecuencia de determinados elementos de la lengua. Puede contribuir a nuestra imagen del funcionamiento de la lengua, haciendo ver la frecuencia variable, según las épocas o los estilos, de unos mismos elementos o sus combinaciones. Puede precisar cómo son manejados sistemas estilísticos o literarios abiertos de la literatura popular, sobre todo de aquella que utiliza fórmulas o expresiones fijas. Puede incluso descubrir unidades que un análisis intuitivo y deficiente había descuidado recoger o bien describirlas mejor; y contribuir al descubrimiento de las más variadas estructuras de lengua y estilo.

Pero, evidentemente, no puede —y aquí está el límite— hacer inducciones por su cuenta. De una parte, carece de la *faculté de langage* y una serie de fracasos en el terreno de la traducción mecánica han evidenciado, si ello era preciso, que es incapaz de resolver problemas cuyos datos y cuyas reglas no le han sido dados previamente. De otro, las colecciones de datos clasificados y computados que nos presenta exigen la atención del lingüista, que es quien ha de inducir de ellos las características de lengua o estilo o de significado a las que responden. El lingüista operará ahora con una riqueza de datos indudablemente mayor, pero tanto más grande será su responsabilidad, en cuanto que no podrá echar la culpa de sus fracasos al material de base.

Aunque, claro está, otra responsabilidad suya es lograr que ese material de base sea homogéneo y adecuado. Hay el peligro de un mecanicismo indiscriminado, de un aprovechamiento mediocre de las nuevas técnicas para lograr una especie de sustituto del trabajo científico que la Administración exige a los miembros de los cuerpos docentes e investigadores y que luego evalúa por número de páginas o número de trabajos. Y hay ciertos ídolos que conviene mantener lejos de nuestra adoración. Por ejemplo, el puro cuantitativismo, que evalúa homogé-

neamente rasgos de significado muy diferente, así en el caso de ciertas utilidades del computador para la filiación de los manuscritos de un autor o para establecer el parentesco entre las lenguas de una misma rama lingüística. Y, sobre todo, la idolatría del símbolo, más peligrosa, sin duda, que la de la palabra. Pues el ordenador es perfectamente neutral y computa aquellos elementos que se le dan, transformados en símbolos. Ahora bien, si les damos un valor absoluto cuando, a veces, son el resultado de análisis o inexactos o parciales, entonces nos exponemos a caer en la trampa que nos hemos preparado nosotros mismos.

Se trata, pues, de hacer los análisis y las preguntas correctas y de utilizar las salidas para lograr inducciones correctas. Si se toman estas precauciones —nada fáciles, por supuesto— qué duda cabe de que la máquina será nuestro servidor y, concretamente en nuestro caso, llevará más allá las posibilidades del lingüista.

Y, añadamos, no habrá peligro de esa deshumanización por la que, instintivamente, nos sentimos amenazados cuando se habla de la introducción de la máquina en el estudio de nuestras viejas ciencias, que tienen una tradición de más de dos mil años. Al contrario, yo diría que el ordenador está llamado a contribuir a humanizarlas, usado dentro de sus justos límites y sin pretensión de imperialismos.

De un lado, es claro que evita un trabajo engorroso y mecánico que, por otra parte, hacíamos con las viejas técnicas manuales de un modo insuficiente. La máquina, también aquí, ha llegado para aliviar al hombre de tareas subordinadas que encargaba a alguien a su servicio o que le quitaban un tiempo y una holgura mental preciosos.

Pero, sobre todo, ya que me estoy refiriendo a la Lingüística, el ordenador está llamado a introducir una riqueza humana que se perdía en el esquematismo y la abstracción de los movimientos lingüísticos más en boga en el momento de su llegada. Me estoy refiriendo al Estructuralismo y a la Lingüística generativa.

Por reacción, quizá, contra la atención a lo individual de la lengua en los años del idealismo lingüístico de Croce y Vossler y de la Estilística a la manera de Bally, las nuevas escuelas, al menos en sus momentos iniciales, han exagerado tremendamente lo que es común y general, lo que es esquema y regla en la lengua. Por hablar primero del Estructuralismo, ya Hjelmslev hablaba del álgebra del lenguaje, que pretendía reducir a un juego simple de invariantes, unidades y funciones. Estas unidades y funciones se ha tendido, sobre el modelo de la Fonología

por demás simplificada de la escuela de Praga, a concebirlas de la manera más simple posible. Sobre la base de escasísimos ejemplos, ejemplos de simplicidad privilegiada, se han presentado por varios lingüistas modelos de oposiciones semánticas y de análisis semánticos de lo más simple. No se ha hecho un esfuerzo para llegar a descripciones exhaustivas de los elementos semánticos de una lengua, ni en su sistema ni en su funcionamiento. Hay, ciertamente, excepciones que procuraban arrancar del análisis exhaustivo de datos y de inducciones a partir de ellos. Pero estaban en minoría.

Esto es cuanto a la Semántica. Pero ¿qué decir de descripciones de Morfología o de Sintaxis o de Métrica que no tienen en cuenta las frecuencias, que dan casi igual valor a los hechos raros y anómalos y a los más centrales de la lengua y que dan, por tanto, una imagen desdibujada de ésta? Aquí estamos en el caso contrario: la pérdida de los perfiles del núcleo sistemático de la lengua, frente a su acentuación exagerada en el anterior. Y siempre por la misma causa: que se parte de inventarios incompletos.

Si el ordenador se usa correctamente y no para juegos o curiosidades lingüísticas, puede ser la base de un mejor conocimiento de lo que es sistemático en la lengua y lo que es marginal o pertenece a diversos niveles de la misma; y siempre con una severa computación, que hará ver cómo las transiciones son siempre graduales y varían en el detalle. No será el final de la concepción estructural de la lengua, que existe desde toda antigüedad y seguirá existiendo siempre por el hecho simple de que la lengua tiene un núcleo estrictamente estructurado. Pero la humanizará y cubrirá su escueta geometría de la carne y la sangre de los tejidos vivos, que se manifiestan en la actividad lingüística. Y ésta será mejor conocida, pues no hay antítesis, por mucho que se haya afirmado, entre su estudio y la descripción de textos: si éstos se analizan correctamente y se hacen las inducciones y luego las deducciones pertinentes, es claro que el conocimiento de la actividad lingüística y del aspecto creador del lenguaje, que se manifiesta sobre todo en los dominios de baja frecuencia, saldrá beneficiado.

Cosas parecidas podríamos decir acerca de la Lingüística generativa, cuyas diferencias respecto a la estructural, existentes, han sido exageradas. En realidad, han sido las unidades y funciones de la misma, casi todas tradicionales, las que están en su base, en unión de un punto de vista lógico-matemático nuevo. Aunque en el curso del desarrollo

de esta escuela ha llegado, como era previsible, a preocuparse por los problemas concretos de los textos de las lenguas naturales, tales como las desviaciones, los grados de gramaticalidad, etc., es bien claro que su punto de partida implicaba un interés predominante por un modelo de lengua muy simplificado, más afín a los lenguajes simbólicos que a las lenguas naturales.

Intentaban e intentan los generativistas, antes que cualquier otra cosa, crear un modelo que explique la actividad lingüística, tanto en su aspecto de creación como en el de competencia y reconocimiento. Y ello en términos generales y abstractos. Para ello han desplegado un aparato metodológico y terminológico de origen lógico-matemático: términos como conjunto, concatenación, generación, modelo, matriz, ordenado y desordenado, etc., etc., lo testimonian bien claramente. Su atención a la Semántica ha sido fundamentalmente de tipo lógico, salvo en la medida en que han tomado prestada de las escuelas anteriores una Semántica un tanto primitiva para hacer de ella un componente de la estructura profunda cuya relación con el sintáctico les ha proporcionado y les proporciona graves problemas. Pero siempre dentro de una consideración abstracta y general.

Fundamentalmente, se trata de una orientación que a partir de unos principios axiomáticos (que a veces, en el fondo, son simplemente tradicionales: por eso se está reaccionando ahora contra las primeras concepciones de la estructura profunda) aplica un método deductivo. Realmente, este método ha sido un eficaz revulsivo de muchos vicios y limitaciones y tiene, sin duda, mucho que aportar todavía: aunque los que procedemos del estudio de las lenguas naturales y de las escuelas historicistas o estructuralistas nos resistamos a limitar a él el estudio de la lengua. Ciertamente, como arriba decía, el Generativismo va siendo aplicado a la descripción de las distintas lenguas y a problemas de estilo y de poesía. Ciertos desengaños, como los sucedidos en el terreno de la traducción mecánica por aplicar modelos demasiado simples, mientras que los más complejos de las lenguas naturales continúan siendo, en buena medida, inaccesibles, han llevado a algunos de los representantes de esta tendencia a un replanteamiento de problemas y posiciones.

Pues bien, resulta significativo y, diríamos, irónico, que sean precisamente los principios lógico-matemáticos cuya vigencia en la Lingüística generativa hemos mencionado, los mismos que actúan en el tratamiento de los símbolos por los ordenadores y que, por lo tanto, nos

facilitan el acceso a una Lingüística esencialmente diferente: aquella a que hemos aludido, la que trabaja a partir de inventarios exhaustivos de datos y hace inducciones referentes a los sistemas lingüísticos y a sus márgenes. Es decir, a una Lingüística más próxima al contexto humano de la lengua. Pues si es bien claro que también la Matemática y la Lógica son cosas humanas que operan en la lengua y que este punto de vista, así como aportaciones cual el punto de vista transformacional, los métodos de simbolización, etc., quedarán como ganancias definitivas, toda la escuela comportaba el riesgo de que el estudio de los textos de las lenguas naturales quedara abandonado y de que, con ello, perdiéramos una vía de acceso, vía difícil pero segura, al conocimiento de lo que es el lenguaje en todos sus aspectos.

El ordenador ha vuelto a enlazar, pues, con las antiguas escuelas lingüísticas de base tradicional, literaria y analítica. Les presta la ayuda de la más avanzada tecnología al servicio de la recogida, clasificación y computación de datos. Elimina y deja anticuadas sea las descripciones de tipo intuitivo sea las que son pura deducción a partir de axiomas o puras generalizaciones sin fundamento suficiente. Acoge, por supuesto, tanto las tesis estructuralistas como las generativistas, que no son incompatibles, pues se trata más bien de dos métodos para describir una misma realidad o aspectos de una misma realidad. Facilita su estudio de los datos concretos de las lenguas naturales, lima las asperezas de sus formulaciones demasiado abstractas, limitadas y esquemáticas.

Sólo hace falta desear que este mismo ordenador no nos ahogue con el exceso de datos, a nosotros que nos quejábamos de su escasez, y que nos deje libres la imaginación y el sentido crítico para interrogarle e interpretarle con acierto.

Y con esto sólo me queda, expresando, estoy seguro, el sentir de todos los presentes, dar las gracias a quienes han hecho posible este Simposio. Al Rector de nuestra Universidad, profesor González Alvarez, con cuya presencia aquí nos honramos. A su Vicerrector de Investigación, el Doctor Torroja, que es quien, como dije al principio, ha tenido la iniciativa y llevado la organización. Al Director General del Ministerio de Educación y Ciencia, señor Cambor, gracias a cuyo interés hemos podido formular la invitación que hemos hecho a nuestros huéspedes extranjeros. Y a todos los que han venido de otras Universidades a compartir nuestros trabajos y a los que han aceptado encargarse de las diversas comunicaciones.

Doy paso, pues, con esto a quienes han de intervenir esta mañana

con sus comunicaciones, que irán seguidas luego de discusión por parte de todos los presentes. De todos ellos y del Simposio en su conjunto esperamos todos aprender mucho estos días.