

Hacia la construcción de una teoría de datos geográficos de proximidad

Constancio de CASTRO AGUIRRE

1. INTRODUCCIÓN

Creo que para el universitario inmerso en la tradición de los estudios geográficos este enunciado resulta por un lado pretencioso y por otro lado quizás muy sesgado. Pretencioso ha de resultar sin duda a la vuelta de tantos años de manejar los datos geográficos que alguien pretenda construir una teoría sobre los mismos ¿Es que para el geógrafo no es obvia la presencia del dato geográfico? Pero además si a esta pretensión se añade la exigencia de circunscribirnos al carácter de proximidad parece desde un principio una intención muy parcializada y excesivamente fragmentaria en sus alcances. El dato geográfico, nos parece oír, está ahí para quien quiera captarlo y carece por el mero hecho de ser geográfico de la limitación de la distancia.

Yo que personalmente arrastro esa misma formación y esos mismos pensamientos acabo de llegar a una nueva plaza en donde se cruzan muchas y distintas avenidas del saber: la avenida de la Geografía, la avenida de la Sociología, la avenida de la Economía, la avenida de la Psicología, etc... Esta plaza que los nuevos diseñadores del orden social han bautizado con el nombre de Ciencias Sociales y algunas veces —las menos— de Ciencias Comportamentales reúne una algarabía de discusiones, de grupos en disputa. De hecho me encuentro en plena plaza en medio de un grupo copiosamente nutrido. En todas las avenidas tengo amigos y al parecer sobrados de conocimientos que han venido acumulando por sus propios caminos. Es interesante escucharlos a todos y hacer conforme a la vieja tradición griega un paseo peripatético por los soportales de la plaza.

Cuando se han pasado años en este estilo de vida algunas de esas cosas obvias de la Geografía no lo parecen tanto y comienza un replantea-

miento desde las raíces de los conceptos. A fuerza de oír a los psicólogos cuestionarse qué cosa sea la medición resulta que su cuestionamiento ha llegado a las puertas de los físicos y estos, que habían tomado por obvia la expresión numérica de las observaciones fundamentales de la física, aceptan al fin como bien fundada la teoría de la medición (L. Gonella, 1979). De un modo semejante asistimos a una preocupación persistente de los psicólogos en definir una Teoría para los Datos que cubren las observaciones del comportamiento (Coombs, 1964). Todas estas actividades teóricas son hoy inexcusables en el programa de compromisos a cubrir por parte de los científicos sociales. En algunas parcelas se siente con mayor intensidad el estímulo hacia la innovación, pero va siendo un denominador común de todas ellas.

En 1964 Berry propuso un esquema que ha tenido aceptación entre los geógrafos la llamada matriz de datos geográficos (Berry, 1964). Consistía esta matriz en una disposición de filas y columnas, en donde las filas servían para identificar los elementos «locacionales» y las columnas para identificar las características o atributos, dentro de un cuadro regional. La intención de Berry era, sin duda, sana: suministrar un procedimiento objetivo para los estudios de Geografía Regional, Sistemática e Histórica demostrando que todos ellos arrancan de la misma matriz de datos geográficos. El planteamiento de Berry no va más allá de una manipulación estadística de los datos. Hoy día tras haber pasado dos décadas distraídos en la asimilación de técnicas estadísticas cada vez más refinadas, los geógrafos comienzan a cuestionar el trabajo cuantitativo. Existe una conciencia de malestar originada en el uso indiscriminado y ciego de técnicas estadísticas. Situado en un frente común con las denominadas Ciencias Sociales, al geógrafo le llega con frecuencia la objeción clamorosa que sus «partners» hacen al demonio de la cuantificación. Ahora viene a estar de moda la duda epistemológica. En una reciente publicación inglesa (Bennett, 1981) el conocido R. J. Johnston dedica un capítulo al tema de «la ideología y la geografía humana cuantitativa en el mundo de habla inglesa». Allí se dice por ejemplo en la pág. 40.

La ideología de las ciencias naturales nunca ha alcanzado total supremacía en la geografía humana, si bien hacia la mitad de la década de los 60 fue suscrita por una mayoría de miembros en la disciplina y ciertamente por una gran mayoría de la generación más joven. La generación anterior permaneció escéptica, especialmente aquellos que se consideraban como trabajadores de campo y que tenían la responsabilidad de recoger la información requerida para la evaluación de los modelos y la verificación de hipótesis. Al final de los años 60 habían atraído a más gente para engrosar las filas del escepticismo y en la década siguiente la crítica a la geografía humana cuantitativa fue haciéndose más firme. Esta crítica se alimentó en dos fuentes ideológicas: la ideología humanística y la ideología radical.

La lectura de unos y otros, humanistas y radicales, en sus aceradas críticas nos lleva a la conclusión de que combaten un fantasma. El fantasma del modelo físico en la construcción de la ciencia. Todos ellos sin excepción están pensando en una identificación del movimiento cuantitativo en la Geografía con el modelo físico de hacer ciencia.

La inspiración del movimiento cuantitativo puede y debe nutrirse en otras fuentes distintas al modelo detallado del conocimiento natural. Un punto decisivo que fija la bifurcación de Ciencias Naturales y Ciencias Comportamentales llevando ambas la aspiración al rigor lógico y cuantitativo es la discusión en torno a la medición. Esta discusión que nace en los años 30 y es conducida en el seno de la Asociación Británica para el Avance de la Ciencia resulta desconocida para la mayoría de los científicos sociales de la actualidad. En otra oportunidad trataremos de ilustrar este punto con todo el detalle necesario. Por ahora diremos lo siguiente: la geografía humana de hoy con enfoque cuantitativo no tiene por qué adoptar ningún modelo físico; es más, las bases iniciales de apertura a los conceptos científicos, es decir las bases de la medición son distintas en la Geografía y en la Ciencia Natural. En este sentido nos inclinamos a reservar para la Geografía humana un nicho de la amplia clasificación de Ciencias Comportamentales. Esta idea será ampliamente expuesta y esperamos que con el rigor necesario en esa futura oportunidad que estamos anunciando. En este momento necesitamos entrar de lleno en la descripción siquiera breve y somera de una Teoría de Datos.

2. ENTIDAD-SOPORTE DE LA OBSERVACIÓN VS. VARIABLE-ATRIBUTO DE LA OBSERVACIÓN

Los estadísticos no han hecho quizá todo el hincapié necesario en este detalle. O más bien hemos sido nosotros, en este caso los geógrafos quienes hemos descuidado el tema. El hecho es que todo usuario de técnicas estadísticas se enfrenta desde un principio a Tablas de Datos en donde se especifican de un lado valores en la variable y de otro lado las frecuencias. Por supuesto es posible complicar estas Tablas determinando intervalos en la variable a modo de un continuo: en este caso las frecuencias señalan el número de observaciones habidas dentro de cada intervalo. Lo que nos importa destacar ahora es el hecho de que toda observación descansa en un ente que llamamos soporte de la observación. No hay observación posible de cuantía demográfica si no tenemos las localidades donde levantar dicha observación; no hay tipo de vegetación si no tenemos la superficie revestida de vegetación. Este hincapié parece una insistencia en lo trivial y así es, pero es importante hacerlo por las consecuencias que vamos a ver. Las Tablas de frecuencia han sido planteadas en realidad como un artefacto para facilitar el

cálculo manual de datos numerosos y para ofrecer una base a la representación gráfica. En esas Tablas queda muy patente el hecho que estamos comentando de una diferenciación entre la variable observada y los entes que soportan la observación. Pero donde aparece con más nitidez la diferenciación es cuando producimos una ordenación de la variable. Sucede sin embargo que los manuales de Estadística no le han otorgado gran importancia a este tema porque han estado más atentos a confeccionar parámetros de posición y dispersión basados en modelos físicos.

Podemos ver los conceptos y matizaciones que implica una ordenación de valores para introducirnos en la cuestión que estamos planteando. Por ejemplo, en el momento de ordenarse las observaciones conforme al criterio de la variable se producen unos lugares de ordenación. Si tenemos cinco localidades con sus respectivas cuantías de población, se ordenan estos valores

<i>BHT</i>	12500	→	POSICION CUARTA
<i>ZJV</i>	13800	→	POSICION QUINTA
<i>SLM</i>	7650	→	POSICION PRIMERA
<i>HRK</i>	11220	→	POSICION TERCERA
<i>CPN</i>	9740	→	POSICION SEGUNDA

y en consecuencia brotan las posiciones correspondientes. Las posiciones se definen como lugares de una ordenación y son cubiertas mediante los elementos que sustentan las observaciones. En nuestro caso las posiciones vienen a estar ocupadas por las localidades. Conforme a los hábitos estadísticos elementales si se produce una Tabla de valores ordenados da lugar a dos columnas: una columna con la variable ordenada y la columna correspondiente de frecuencias u observaciones acumuladas. Para el siguiente cuadro de quince temperaturas mínimas en un mes correspondientes a otras tantas localidades tenemos la siguiente tabla de frecuencias acumuladas:

8	10	7	12	8
9	11	10	9	13
10	12	11	10	11

X	7	8	9	10	11	12	13
f(x)	1	2	2	4	3	2	1
f _{ac}	1	3	5	9	12	14	15

FIG. 1.—Tabla de frecuencias acumuladas.

Surge un matiz problemático al pensar que se dan solamente siete valores distintos en temperatura correspondientes a quince observacio-

nes. Evidentemente tendrá que haber tantos lugares de ordenación como observaciones. Para estas ocasiones se suele arbitrar el siguiente procedimiento: 1.º Los lugares de ordenación que ponen de manifiesto los datos observados se corresponden con las frecuencias acumuladas. 2.º Se da entrada a la consideración de lugares latentes (en nuestro ejemplo el 2, el 4, el 6, el 7, etc.). 3.º Para esos lugares latentes se buscan valores mediante interpolación.

Queremos llevar a la mente del lector la idea de que el estadístico trabaja con elementos y variables, es decir con el elemento sustentante de la observación y con el valor observado como dos entidades conceptuales distintas. Hasta el punto de que en algunos casos esa distinción conceptual le sirve de base para generar nuevos valores. Estos valores por supuesto no son observados sino asumidos bajo ciertos supuestos o hipótesis de trabajo. Precizando las consideraciones hasta aquí hechas:

Los elementos de soporte son entes bien definidos que permiten una perfecta contabilización.

Las variables consisten en magnitudes o valores que se asignan a los elementos de acuerdo con alguna característica mensurable.

3. LAS DECISIONES INICIALES DEL GEÓGRAFO

Volvamos ahora a la vieja idea de la representación matricial de Berry pero profundizando en el concepto de manejar las observaciones geográficas:

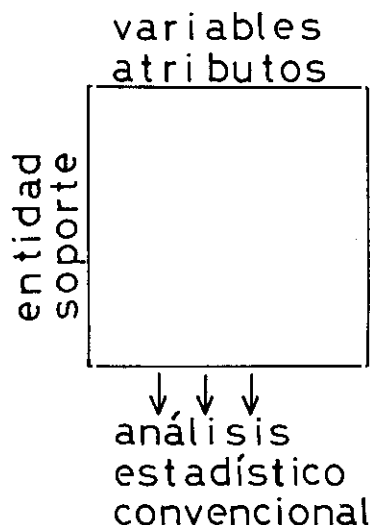


FIG. 2.—Representación matricial de Berry.

las filas de esta matriz aluden a una definición locacional, bien sea una localidad o un agregado de localidades, bien sea una superficie delimitada administrativamente, o bien un foco de servicio y suministro para un área circundante. Vemos pues que la entidad soporte de observaciones en en sí misma variada y no está dada en el sentido de un elemento físico. Es decir cuando el científico quiere estudiar la atmósfera se encuentra con unos elementos dados: los gases y partículas existentes. Una vez situado frente a los mismos estudia su composición química, sus reacciones, sus desplazamientos, etc... La tarea del geógrafo es algo distinta por cuanto tiene que decidir la definición del elemento sustentante de sus observaciones; no le basta con situarse frente al espacio terrestre como el científico se sitúa frente a la atmósfera. La decisión del geógrafo afecta a la unidad espacial que va a servirle de elemento. Esta unidad espacial de acuerdo con las exigencias anteriormente establecidas conviene a la perfección con una definición administrativa del territorio; cuando tal delimitación administrativa no existe surge la ambigüedad y la imposibilidad por tanto de registrar observaciones adecuadas. Indudablemente el geógrafo habla muchas veces de unidades espaciales que escapan a una definición administrativa; así sucede con la idea de comarca o región. Las Urdes, al Ampurdán, la Alcarria, La Maragatería son ámbitos espaciales de este tipo que carecen de la condición de estar bien definidos. La definición administrativa del territorio le provee al geógrafo los resultados de registro y control de los distintos sectores de la administración; esta gama de información cubre la asignación de magnitudes y valores a los elementos en cuestión. Sin embargo la administración define sus unidades con un sentido jerárquico y acumulativo; de ahí la existencia de diversos niveles de agregación que pueden ser escogidos por el investigador de acuerdo con los objetivos del estudio. Si se pretende estudiar la distribución de centros hospitalarios es posible que ello no tenga sentido a nivel de municipios; pero si lo que se busca es la distribución espacial de escuelas será muy plausible recoger las observaciones a nivel municipal. Por consiguiente una regla de oro que ayudará al geógrafo a definir los elementos locacionales es la siguiente:

El nivel de escala en la escogencia de unidad espacial debe ser el *mínimo nivel de agregación* que permiten los objetivos del estudio.

Pero hay otro factor adicional de semejante importancia que también depende enteramente de las decisiones del investigador. La escogencia de elementos debe ser tal que ofrezca suficiente base a la variabilidad de cada característica. Es decir cada característica debe estar representada en su mayor riqueza posible. Con pocos elementos de sustentación la información recogida tiende a ser pobre. Un estudio de las metrópolis superiores al medio millón de habitantes en España no ofre-

ce la riqueza que podría obtenerse a nivel de toda Europa. De ahí la importancia de definir un contexto de elementos. Generalmente los geógrafos adoptan en torno a este punto una actitud de reconocimiento tácito; pero es indudable que con vistas a un riguroso progreso en la exploración debe descartarse cualquier vaguedad. Todo esto significa que al iniciarse la toma de información y con miras a la construcción de los datos el geógrafo decide dos cosas previas: una se refiere a las unidades espaciales que servirán de soporte a las variables y otra es la reunión de estos elementos espaciales en un contexto significativo para las pretensiones del estudio.

La primera decisión —la de elegir los elementos espaciales y por tanto el nivel de agregación— tiene consecuencias graves en el trabajo posterior. Así ha sido reconocido en una estimulante publicación de CATMOG (ver Opensahw). En esta publicación se plantea con rotundidad el hecho de que los elementos espaciales pueden modificarse por agregación incluso a lo largo del estudio. Dice así en la pág. 3:

Para múltiples propósitos las zonas en un sistema de zonificación constituyen los objetos o individuos geográficos que son las unidades básicas para la observación y medición de los fenómenos espaciales. Habitualmente en un experimento científico la definición de los objetos de estudio debe anteponerse a cualquier intento de medir sus características. Sin embargo este no es el caso de los datos espaciales en donde los objetos espaciales solamente existen después que los datos recogidos para un conjunto de entidades se someten a una agregación arbitraria con miras a producir un conjunto de unidades espaciales. Considérese un ejemplo con las siembras de trigo y patata. Los datos para un conjunto de entidades (granjas) pueden agregarse para producir un conjunto de unidades espaciales (parroquias o condados).

De la misma manera este tema de las unidades espaciales plantea problemas al concepto de correlación espacial, ya que en principio no es el mismo objeto o individuo espacial al que le asignan caracteres distintos. El objeto espacial se fragmenta en áreas que ofrecen esos distintos caracteres con lo que nos enfrentamos al problema de la agregación latente siempre en todo momento. La raíz de este problema está en el carácter extensivo de los individuos espaciales (estamos llamando individuo al elemento-soporte de la observación) los cuales llevan en sí dos aspectos que hacen difícil su manejo: uno es el aspecto de continuo infinitamente divisible y otro es el aspecto de su imposibilidad de solapamiento.

4. LA NOCIÓN DE VECTOR O PERFIL

Tras haber establecido el «individuo espacial» estadísticamente hablando llueven las observaciones sobre el mismo. Un aspecto que puso

de relieve la matriz de Berry es el de la multiplicidad de valores que pueden asignarse a estos individuos. Ahora bien la estadística se ha ocupado de examinar la comparecencia de valores en estas variables, determinando para cada variable parámetros de posición y dispersión; en el caso de estudiar conjuntamente las variables se han hecho análisis factoriales cuya utilización por los geógrafos desborda las posibilidades y limitaciones de la técnica. Esto es lo que hemos señalado en el gráfico antes reproducido; es decir, se toman los valores para cada variable a lo largo de las unidades espaciales consideradas como soporte; esa es la significación de las flechas. Por ejemplo en el caso de que tuviéramos como unidades de soporte los municipios de un contexto (tómese por tal la cornisa cantábrica) y tomáremos en consideración en ellos la cuantía de población, la flota pesquera, el parque de automóviles, los índices de enfermedades respiratorias, etc..., el análisis estadístico operaría sobre estos valores obteniendo un promedio de cuantía de población, de flota pesquera, etc..., para toda la cornisa cantábrica.

El geógrafo echa en falta hoy otro enfoque para la manipulación de datos. Lo que sucede en realidad es que necesitamos una técnica que nos permita manejar *conjuntamente* todos los valores asignados a cada unidad de soporte; en el ejemplo anterior necesitamos una técnica que nos permita manejar en bloque todos los valores y magnitudes asignadas a cada municipio. Este bloque no es otro que el perfil de cada municipio y si se quiere una expresión de cuño matemático el vector. Ahora bien, es imprescindible que el geógrafo asimile las limitaciones y posibilidades que se ofrecen en este capítulo del desarrollo matemático.

Primero es importante entender que la noción de vector al igual que la noción de número entero es una invención matemática, es decir es una *forma* de pensar. No es un hecho de la naturaleza, como la luz, el peso de los cuerpos, el desplazamiento de los cuerpos, la temperatura, etcétera. Los hechos de la naturaleza y de la sociedad el hombre los encuentra y choca con ellos; los conceptos matemáticos en cambio son inventados por la razón. El hombre intenta con ellos dar una forma inteligible a la naturaleza y al comportamiento social. Como tales formas presentan una arquitectura y se atienen a unas leyes de coherencia que no es posible olvidar.

Matemáticamente hablando el vector admite una expresión geométrica y una expresión más abstracta sin ningún apoyo intuitivo que llamaremos lógica. Geométricamente el vector puede entenderse como un punto definido mediante coordenadas en «*m*» dimensiones. Apoyándonos en nuestra intuición es fácil representar un punto en dos y tres dimensiones; se trata de dar un salto imaginativo y llegar a «*m*» dimensiones. Prescindiendo del apoyo geométrico el vector puede ser considerado también como una disposición *ordenada* de «*m*» valores numéricos. Ambas definiciones valen para nuestro propósito. Según esto y siguiendo con el ejemplo de los municipios cantábricos tendríamos re-

presentado cada municipio mediante una sucesión de valores: cuantía demográfica, un índice de capacidad de flota pesquera, número de automóviles matriculados, números de casos de enfermedades clasificadas en el período de un año etc... Queremos destacar con el ejemplo la idea de que pueden reunirse observaciones muy diversas con tal de que aparezcan cuantificadas de acuerdo con un criterio de medición. Es decisivo que haya un orden o sucesión determinada en el vector, el cual debe preservarse en todos los municipios. Se tiene así una correspondencia estricta en los valores numéricos que definen cada municipio; es decir el valor numérico que aparece en tercer lugar será el parque automovilístico. El vector así definido es susceptible de operaciones que se conforman a ciertas leyes y restricciones. Por ejemplo, se pueden llevar a cabo sumas y diferencias entre vectores, con tal de que respondan al mismo orden pero la multiplicación no es conmutativa. Lo que bajo ningún concepto tiene sentido es sumar internamente los valores del vector. Una relación binaria que va a proporcionarnos gran utilidad para el análisis multidimensional es la función geométrica de distancia; de este modo se podrá obtener una matriz de distancias entre vectores.

No pretendemos aquí hacer una exposición de todas las posibilidades que nos ofrece el cálculo vectorial. Nos basta con indicar que se trata de una representación matemática y formal que permite recoger una multiplicidad de información; debido a la expresión geométrica que hemos adoptado para definir el vector podemos decir correctamente que se trata de un instrumento multidimensional. Nuestro objetivo en el presente trabajo consiste en construir una Teoría de Datos para lo cual procedemos a establecer algunos criterios de clasificación.

5. OBSERVACIÓN COMPACTA Y DESDOBLADA

Si nos retrotraemos a la distinción entre los elementos soporte y las variables de observación hay un hecho fundamental que persiste en la diferenciación: se trata de dos decisiones distintas del investigador. Ya hemos comentado en párrafos anteriores la complejidad de la decisión que atañe a los elementos soporte o individuos espaciales. Con respecto a las variables que deben ser cubiertas en la observación hay evidentemente una segunda decisión del investigador completamente diferente de la anterior. Importa señalar que en principio las variables pueden ser éstas o aquellas, muchas o pocas, pero no hay un repertorio establecido fijo e inalterable. A este conjunto de variables decidido por el investigador le corresponde en cada individuo espacial un perfil de valores que en términos matemáticos hemos denominado vector. Tenemos pues una terminología varia que conviene precisar y delimitar. La Fig. 3 nos ayudará a esta tarea.

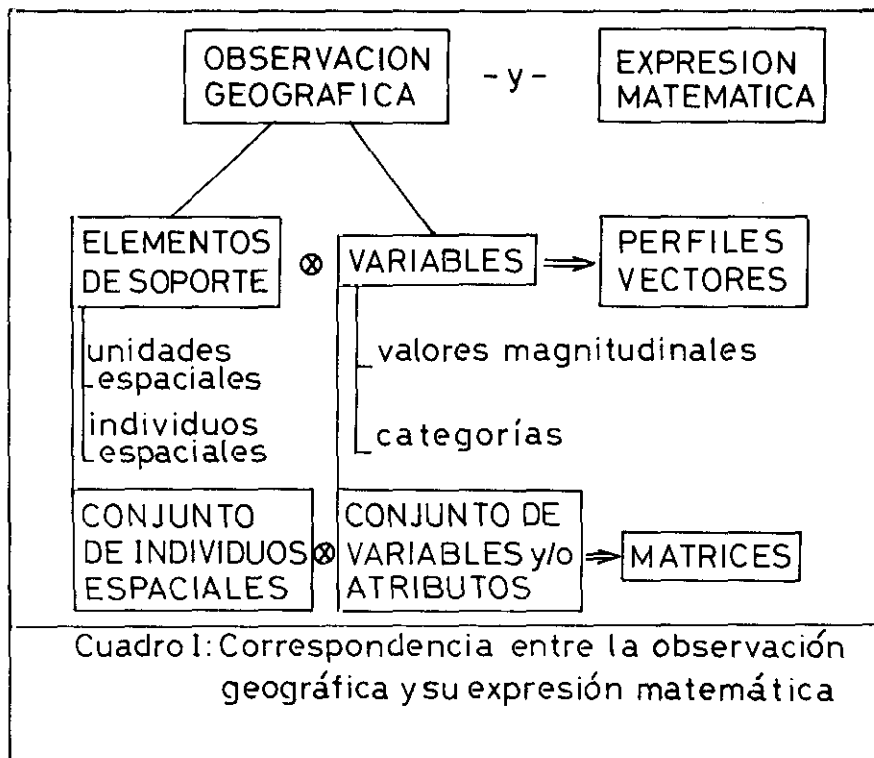


FIG. 3.—Correspondencia entre la observación geográfica y su expresión matemática.

Queremos presentar ahora una operación de desdoblamiento en la entraña misma de la observación. Tras lo que venimos diciendo queda claro que existe un deslinde conceptual el cual no tiene nada de artificioso y responde a decisiones distintas en las tomas de observación que realiza el investigador. A consecuencia de este desdoblamiento pueden caracterizarse dos conjuntos tal y como queda señalado en la Fig. 3. Pero queda igualmente claro que puede tomarse la observación en forma compacta: el individuo espacial fundido en las características a modo de un perfil de valores magnitudinales. Ambos casos pueden contemplarse como posibles dando lugar a la primera dicotomía que servirá de base a la taxonomía de datos que vamos a elaborar. Conviene decir algo más incisivo acerca del desdoblamiento. La selección de rasgos o variables es como queda dicho una responsabilidad del investigador. En este sentido cabe que los rasgos elegidos sean exclusivamente comportamentales, vg. la preferencia que muestran los habitantes de cada unidad espacial con respecto al lugar de residencia. Se tiene entonces una especie muy peculiar de desdoblamiento, puesto que pueden coincidir en ser los mismos los elementos soporte y las variables. Lo que

ocurre que es bajo un aspecto (elemento-soporte) son considerados sujetos activos de la preferencia y bajo otro aspecto (variables) son considerados objetos pasivos de la preferencia.

6. ESTRUCTURA RELACIONAL DE LA OBSERVACIÓN Y TAXONOMIA DE DATOS

Hay un segundo criterio que ponemos en marcha completamente independiente y ajeno al que acabamos de exponer. El geógrafo como todo científico busca la comparación de sus objetos. Los objetos del geógrafo tienen la peculiaridad de ser individuos espaciales revestidos de características variables. El geógrafo obtiene al comparar sus objetos el resultado que ha de guiarle en su exploración progresiva. Pero he aquí que la comparación tiene fundamentalmente dos enfoques: o bien se trata de una comparación con miras a establecer la proximidad entre los objetos o bien se busca establecer la relación dominativa entre los objetos. Bajo estos dos orígenes cabe realmente la gama de todas las comparaciones posibles. Se trata de una dicotomía de carácter fundacional la cual cruzada con la dicotomía anterior nos presenta un cuadro taxonómico que exponemos en la Fig. 4.

	COMPARACION DOMINATIVA	COMPARACION APROXIMATIVA
OBSERVACION COMPACTA (UN CONJUNTO)	II	I MATRICES DE PROXIMIDAD
OBSERVACION DESDOBLADA (DOS CONJUNTOS)	III MATRICES DE INCIDENCIA (CENTRALIDAD)	IV MATRICES DE PROXIMIDAD CONDICIONAL

CUADRO II : TAXONOMIA DE DATOS GEOGRAFICOS

FIG. 4.—Taxonomía de datos geográficos.

En este cuadro se presenta el resultado final de nuestra propuesta. Se trata efectivamente de una Teoría construida mediante conceptos que brotan en un análisis paciente de la observación geográfica. Cualquiera puede rastrear las huellas de esta construcción teórica en la obra de Clyde H. Coombs profesor recientemente jubilado en la Univer-

sidad de Michigan (Ann Arbor). Nuestra propuesta se refiere únicamente a datos de proximidad y en este sentido el geógrafo puede percatarse que establecemos dos cuadrantes el I y el IV referentes a Matrices de proximidad. Más adelante detallaremos y ejemplificaremos en lo posible el tipo de observaciones geográficas que pueden ser conducidas a estas Matrices. De momento vamos a intentar alguna aclaración rápida sobre los cuadrantes II y III.

Comencemos por el cuadrante III que tiene un ejemplo de análisis familiar para los geógrafos: la centralidad. Quizás venga bien describir en qué consiste operacionalmente hablando la centralidad para así reflejar en ella los rasgos sobresalientes de esta Teoría de Datos. Todos los geógrafos han coincidido unánimemente en definir la centralidad como un atributo de los asentamientos humanos mediante el cual es posible decidir que un lugar es más central que otro; he aquí una comparación dominativa. Pero ese atributo el geógrafo lo fija en una dotación de actividades o servicios que acompaña a las localidades; se insinúa con esto la observación desdoblada. El geógrafo admite que es posible establecer un rango de las actividades de acuerdo a su dimensión ubicativa y al mismo tiempo establecer una jerarquía de los lugares conforme con el rango de las actividades que ostentan. La matriz de incidencia es el instrumento de base para el análisis de centralidad; consiste en una disposición de los lugares en filas y las actividades en columnas. El cruce de fila con columna señala la incidencia positiva o negativa; es decir, positiva si el lugar considerado ostenta la actividad, negativa cuando no la ostenta. La centralidad puede definirse según esto formalmente como una estructura o patrón matricial en el que las columnas y filas aparecen ordenadas siguiendo la lectura de sus incidencias respectivas. Véase una ilustración en la Fig. 5; se señalan los lugares mediante las primeras letras del alfabeto y los servicios mediante la letra *S* afectada de subíndice. Hay dos servicios, el S_1 y S_4 , que se sitúan en el mismo rango por medirse con resultados iguales en la dimensión ubicativa.

El cuadrante II lo presentamos sin ninguna alusión concreta. Se trata aquí de comparar vectores tomados en bloque sin el desdoblamiento al que antes hemos aludido. Es decir descendiendo a un lenguaje más ilustrativo se trata de comparar el perfil de varias ciudades; cada ciudad está definida por cuantía de población, parque automovilístico, recaudación de impuestos por espectáculos, número de teléfonos, etc. Si las ciudades se comparasen por un solo atributo, vg. cuantía de población, no habría problema alguno en decidir la posición dominante de unas con respecto a otras. El problema por tanto estriba en establecer esas posiciones dominantes utilizando no uno, sino varios atributos simultáneamente. Formalmente todo esto significa que la estructura multidimensional del vector habría que llevarla sin escamotear ni forzar las observaciones a una estructura unidimensional.

		S ₁	S ₂	S ₃	S ₄	S ₅	S ₆		
A		1	1	0	1	0	1	B	
B		1	1	1	1	1	1	C	
C		1	1	0	1	1	1	A	
D		1	0	0	1	0	1	D	
E		1	0	0	1	0	0	E	

		S ₃	S ₅	S ₂	S ₆	S ₁	S ₄
B		1	1	1	1	1	1
C		0	1	1	1	1	1
A		0	0	1	1	1	1
D		0	0	0	1	1	1
E		0	0	0	0	1	1

CUADRO III: DE LA MATRIZ DE INCIDENCIA AL PATRON MATRICIAL DE CENTRALIDAD									
--	--	--	--	--	--	--	--	--	--

FIG. 5.—De la matriz de incidencia al patrón matricial de centralidad.

7. LOS DATOS GEOGRÁFICOS DE PROXIMIDAD

Estos datos se sitúan en los cuadrantes I y IV de la Fig. 4. Los datos de proximidad se presentan en forma matricial como allí se dice porque consisten en un valor que relaciona las unidades espaciales de dos en dos. Después diremos en que consisten esos valores. Ahora debe quedar claro cómo y por qué surgen las proximidades. Primero que nada hay que advertir que la proximidad es un concepto comportamental; en este sentido se conciben próximos dos objetos geográficos tanto por la semejanza de su perfil como por una determinada frecuencia de contactos. En un caso, en la semejanza del perfil, no tiene por qué contemplarse la interacción. Dos ciudades son semejantes porque son muy semejantes su cuantía de población y su infraestructura de servicios aún cuando no haya interacción entre ambas. En cambio las ciudades que se dan dentro de un contexto regional pueden dar lugar a un contacto que se expresa por diferentes vías. Imagínese el contacto telefónico. Según eso podría definirse una proximidad en función de la frecuencia de contactos. En segundo lugar, una vez establecido esto, hay que pensar que el concepto de proximidad geográfica no se expresa en unidades de distancia física, ni es de ningún modo asimilable a la distancia física. Hace algún tiempo se definió la Geografía como una disciplina de la distancia (J. W. Watson, 1955). En ese trabajo se insiste en la necesidad de darle a la distancia un contenido más comportamental que físico; véase por ejemplo el siguiente párrafo (pág. 41).

Cuando la distancia se trata a manera de una cuantía con la que los objetos se adaptan o dominan el medio ambiente, entonces ofrece claves valiosas para desentrañar la relación con el medio, una relación que conduce la geografía de la descripción a la explicación.

Pero estas consideraciones nos llevan a plantear el tema de la distancia como una función formal dentro de un modelo matemático que llamamos espacio métrico. Es decir la distancia tiene en el lenguaje ordinario una faceta de concepto físico. Si se quiere algún rigor sobre el tema hay que diferenciar la distancia experimentable en sentido físico y la distancia formal de un espacio métrico. En otra parte hemos hecho una amplia discusión sobre este punto (C. de Castro, 1978). Por consiguiente queda abierto el camino para llegar a proximidades a partir de los perfiles o vectores que anteriormente se definieron. Sin ir más lejos entre dos vectores de orden, « m » puede hallarse una función de distancia formal acudiendo a la famosa fórmula:

$$D(a, b) = \left[\sum_{i=1}^m (a_i - b_i)^2 \right]^{1/2}$$

Los dos vectores a , b están constituidos por « m » valores los cuales previamente están justificados como una medida, que no tiene por qué ser una medida extensa. La función $D(a, b)$ relaciona ambos puntos a y b en virtud de la distancia que los separa en un espacio m -dimensional. Esta función se expresa en valores reales (números reales) pero no sabemos a qué corresponden físicamente hablando las unidades en que están dadas esos valores. Volvamos otra vez al ejemplo de la cuantía de población, número de automóviles, recaudación en impuestos y números de teléfonos. Se trata por tanto de dos ciudades definidas por un vector de cuatro valores. Esas dos ciudades son representables en un espacio abstracto de cuatro dimensiones. Cuando hallamos la distancia de esos dos puntos hallamos un valor real (de números reales) pero sería un disparate traducirlo a kilómetros de distancia física. En este caso ¿qué uso puede tener esa función de distancia para el geógrafo? Aquí es donde juega su baza el concepto de proximidad. Si se tiene un conjunto de individuos espaciales, diez ciudades por ejemplo, definidas por los cuatro valores antedichos y se obtienen distancia para todos los pares de ciudades, que serán en este caso 45 pares, podemos usar esos valores de distancia como una *función monotónica de proximidad*. Vamos a designar por φ esta función. La función monotónica quiere decir lo siguiente:

$$\varphi[D(a, b)] > \varphi[D(x, y)] \Leftrightarrow D(a, b) < D(x, y)$$

Esto es, a mayor valor de distancia la proximidad se asume menor y viceversa. Por consiguiente ya no se habla de cuantías sino de una relación de proximidad, la cual como puede verse es una relación cuaternaria. Es decir la proximidad se establece entre cuatro elementos, a saber a , b por un lado y x , y por otro. Para el conjunto de diez unidades

se halla una matriz de proximidades que será cuadrada y simétrica y con valores cero en la diagonal. Este valor cero no es sino distancia cero indicativa de máxima proximidad. Para una matriz de proximidades en los términos que se acaban de describir disponemos de un algoritmo elaborado inicialmente por Shepard (ver Shepard, 1962) que puede darnos una representación en dos dimensiones. Llegan así a desembocar las relaciones de proximidad en un mapa tan plenamente justificado como los mapas que representan extensiones físicas.

Nos queda por decir algo de las matrices de proximidad condicional. Cuando se habló en párrafos anteriores del desdoblamiento de la observación aludimos a estas matrices. Si comparamos el desdoblamiento en el cuadrante III (matrices de incidencia) con el desdoblamiento en el cuadrante IV sucede que allí existían dos conjuntos realmente distintos: el conjunto de individuos espaciales (lugares) y el conjunto de variables (actividades). Aquí en cambio se trata de un mismo conjunto de objetos geográficos que reciben dos tratamientos distintos, el de ser elemento soporte de la observación y el de servir de variable caracterizadora de la observación. Concretamente si tenemos el conjunto de las 17 comunidades autónomas y tomamos las preferencias residenciales de sus habitantes como observación, resulta que cada comunidad interviene como soporte a la observación en ella tomada mientras que lo que caracteriza su perfil preferencial está constituido por el resto de las comunidades. De este modo resulta una matriz de 17×17 en donde cada fila se considera la comunidad que emite sus preferencias y las columnas todas actúan de receptáculo. En esta matriz los datos de proximidad se leen en función de cada fila y por tanto la matriz deja de ser simétrica. En este caso los valores numéricos de la matriz (vg. porcentaje de preferencias expresadas) se toman como una función de proximidad en los siguientes términos:

$$\varphi[P(b/a)] > \varphi[P(c/a)] \Leftrightarrow P(b/a) > P(c/a)$$

La fórmula $P(b/a)$ se lee «preferencia de b a partir de a ». Desde el punto de vista formal es interesante notar la siguiente diferencia entre las matrices de proximidad del cuadrante I y IV. En ambas se trata de un espacio común para los objetos geográficos cuya representación se asume a manera de puntos en dicho espacio. Pero en el cuadrante I las distancias entre los puntos son *absolutas* siendo mínima la distancia de un punto consigo mismo; en cambio en el cuadrante IV las distancias son *relativas* de cada punto con respecto a un *punto ideal* asumido en ese espacio común. Por eso puede ocurrir por ejemplo que las preferencias expresadas por los habitantes de $\langle a \rangle$ con respecto al lugar $\langle a \rangle$ no den la preferencia máxima. El techo de máxima preferencia y por tanto de máxima profundidad comportamental se representa en estos casos por un punto ideal; las valoraciones expresadas en cada fila de la ma-

triz de proximidad condicional hacen referencia a ese punto ideal. Por estas razones cabe expresar la siguiente relación de proximidad:

$$\varphi[P(a/a)] < \varphi[P(b/z)] \Leftrightarrow P(a/a) < P(b/a)$$

Es decir la preferencia del lugar $\langle a \rangle$ según expresan sus mismos habitantes puede ser menor que la preferencia del lugar $\langle b \rangle$ expresada por ellos mismos. En el modelo formal damos cabida a estas expresiones asumiendo que los habitantes de $\langle a \rangle$ tienen un punto ideal de referencia y el lugar $\langle b \rangle$ está más próximo a dicho ideal que el propio lugar $\langle a \rangle$.

Damos por terminada esta exposición de una Teoría de Datos Geográficos. El hecho de que los llamamos datos de proximidad no tiene nada que ver con un acotamiento de distancias físicas. Esperamos que haya quedado claro que el concepto de proximidad se refiere a un modelo de manipular la información. Queda pendiente el tema de la representación en mapas de estos datos de proximidad. Somos conscientes de que abordamos un punto de vista y por tanto no tenemos pretensiones exhaustivas. Caben indudablemente otras perspectivas de trabajo en el quehacer geográfico y podemos guardar hacia las mismas un mayor o menor grado de simpatía. Nuestro propósito es ofrecer una opción con alicientes encuadrada en una concepción moderna de las Ciencias Comportamentales. Hay en esta opción una pretensión que debe ser subrayada. En todo conocimiento geográfico existe una tendencia irreprimible a desembocar en mapas. Esta Teoría de Datos tiene como aspiración máxima la de ofrecer una plataforma de análisis que termina en Mapas Comportamentales. Estos Mapas no son una metáfora sino una expresión que se adecúa con entera justeza a cualquier Mapa de los que se conocen en el lenguaje cartográfico más estricto. Con estas consideraciones nos situamos de cara a un nuevo trabajo: la confección y justificación matemática de los Mapas Comportamentales.

Mayo, 1984.

BIBLIOGRAFIA

- BENNETT, R. J. (1981): *European Progress in Spatial Analysis*. Pion, London.
- BERRY, B. J. (1964): «Approaches to regional analysis: a synthesis». *Annals Ass. Am. Geogr.*
- CASTRO, C. (1978): «Una indagación acerca del espacio métrico, espacio físico y espacio psicológico en Geografía». *Lurralde*, Ingeba, págs. 23-35.
- COOMBS, C. H. (1964): *A Theory of Data*. J. Wiley.
- GONELLA, L. (1979): *Problems in theory of Measurement today*. Acta IMEKO, Moscou.

- OPENSAHW, S. (sin fecha): *The Modifiable Area Unit Problem*. CATMOG, n.º 38.
- SHEPARD, R. N. (1962): «The analysis of proximities». *Psychometrika*, vol. 27, págs. 125-140.
- WATSON, J. W. (1955): «Geography, a discipline in distance». *Scottish Geographical Magazine*, vol. 71, n.º 1, págs. 1-13. (Tomado de HURST, *Transportation in Geography*, McGraw Hill, 1974, págs. 40-50).