# Modelos deductivos para datos borrosos*

(Deductive models for fuzzy data)

Marín Martínez, Angel; González Sotos, León
Univ. Pública de Navarra.
Dpto. de Automática y Computación
Campus Arrosadía
31006 Iruñea

BIBLID [1137-4411 (1999), 5; 129-159]

*El presente trabajo tiene por objeto ofrecer una aproximación a los distintos modelos de datos que permiten el tratamiento de información incierta. Con este fin se introducen las ideas básicas del modelo relacional de bases de datos y la interpretación de la incertidumbre que ofrece la lógica borrosa. En el estudio de los distintos casos se hace especial énfasis en el problema de la redundancia de información así como en los métodos de deducción e interrogación disponibles.*

*Palabras Clave: Modelo relacional. Lógica borrosa. Prolog.*

*Lan honek informazio ezbaiezkoaren tratamendua baimentzen duten datu eredu ezberdinetara hurbilketa eskaintzea du helburu bezala. Honako xedearekin datu basearen eredu erlazionalari eta logika fuzzyak eskaintzen duen ezta baidaren interpretazioari buruzko oinarrizko ideiak erakusten dira. Kasu ezberdinen azterketan informazioaren erredundantziaren arazoa azpimarratzen da, hala nola dedukzio eta galdetzeko metodoak.*

*Giltz-Hitzak: Eredu erlazionala. Logika fuzzy. Prolog.*

*Ce travail cherche à offrir une approximation aux différents modèles de données qui permètent un traitement d'information incertaine. À cette fin on introduit d'abord quelques idées à propos du modèle relational de bases de données et de l'interprétation de l'incertitude avec la logique floue. Dans les différents cases on étude en particulier le problème de la redondance d'information et les méthodes de déduction et d'interrogation.*

*Mots Clés: Modèle relational. Logique floue. Prolog.*

* Trabajo desarrollado con una ayuda para investigación de Eusko Ikaskuntza.

## 1. INTRODUCCION

La automatización del tratamiento de la información ha exigido revisar tanto los métodos y procesos manuales en que se basaba como los conceptos y elementos de apoyo a dicho tratamiento. Los diversos procesos de almacenamiento y manipulación apuntan a un nuevo entramado teórico en que las ideas que dan relieve a la información contenida en un objeto tales como dato, registro o atributo adquieren un significado explícito. La descripción pormenorizada de los objetos y las operaciones, y el consiguiente análisis de sus relaciones, ofrecen una visión del problema en la que se unifican la organización del sistema y su *modus operandi.* Existen tantas perspectivas o enforques como formas relevantes de describir los objetos y sus operaciones. En cada uno de los casos el conjunto de características descriptivas escogido representa lo que llamaremos el *modelo de datos* con que se ha proyectado el sistema de información.

El modelo relacional, introducido por CODD, permite una organización de los datos basada en estructuras propiamente matemáticas y un posterior tratamiento de los datos de orientación algebraica. El estudio de las bases de datos, el análisis de la redundancia y de la dependencia entre los distintos atributos de un registro puede ser abordado mediante una formulación derivada de su organización lógica y ajena a la organización física o interna de los datos.

En el mundo real, sin embargo, es frecuente tener que tratar con información incompleta, difícilmente contrastable o inexistente. El intento de ampliar el modelo para poder tratar información imprecisa afecta de manera radical tanto a la organización de los datos como a las operaciones involucradas.

Para lograr un nuevo tratamiento se requiere previamente tipificar la información que será suministrada. Un valor tal como *alto* puede considerarse como una información precisa si es una de las cuatro etiquetas *bajo, mediano, normal, alto* establecidas en el dominio del atributo altura de los individuos, pero no resulta claro cómo compararlo con *no muy bajo* o *más bien alto,* ya que la ambigüedad de estos términos no permite operar con ellos. Si tales etiquetas se consideran predicados y se establece su validez como una función de la altura, podremos compararlos. La graduación de la validez de un predicado nos lleva a un tratamiento de la incertidumbre que es posible formalizar con ayuda de la teoría de subconjuntos borrosos.

Antes de estudiar los distintos modelos de datos borrosos, es conveniente recordar brevemente los conceptos elementales sobre el modelo relacional, así como los relativos a subconjuntos borrosos.

## 2. IDEAS PRELIMINARES

### 2.1. Bases de datos relacionales

Las ideas en que se basa el *modelo relacional* de bases de datos fueron introducidas por CODD en 1979 [13] y han contribuido desde entonces a dar un encuadre matemático y un enfoque riguroso al tratamiento de los sistemas de información. Para una revisión amplia del tema véase DATE 1990 [19].

En este modelo los datos se organizan en forma tabular, es decir están dispuestos en un conjunto finito de tablas de valores. A cada una de las tablas que forman la base de datos la

llamamos *relación,* en el sentido matemático usual del término. Las filas que componen cada relación de dicha base se llaman *tuplas.* Sus columnas aluden a cada uno de los *atributos* que componen el *esquema* de la relación.

A modo de ejemplo se muestra a continuación una pequeña base de datos formada por dos relaciones. Estas reflejan la situación del profesorado de un Centro, indicando primero su Filiación con un esquema que contiene los atributos *Profesor, Asignatura* y *Departamento* y seguidamente su Localización donde se indica el *Departamento,* el *Edificio* y el *Volumen* de profesorado correspondiente.

Figura 1

Filiación

| Profesor | Asignatura | Departamento |
|---|---|---|
| Ruiz | Geometría | Matemática |
| Pons | Historia | Economía |
| Ros | Mecánica | Física |

Localización

| Departamento | Edificio | Volumen |
|---|---|---|
| Economía | Azul | 30 |
| Química | Azul | 8 |
| Física | Rojo | 16 |
| Matemática | Amarillo | 33 |

En el modelo relacional las estructuras de datos resultan homogéneas, y las operaciones de recuperación y modificación de la información contenida pueden ser expresadas en forma algebraica.

Dejando a un lado las operaciones de actualización (inserción, borrado y modificación) de las relaciones, la obtención de información se realiza con un sistema de interrogación o consulta, que permite expresar formalmente preguntas. Las preguntas posibles y la capacidad expresiva del lenguaje de consulta dependen de la interpretación que se haga del modelo de datos construido.

Un sistema de consulta, bajo el enfoque conjuntista que estamos considerando, es interpretado como un repertorio de operaciones sobre relaciones destinado a obtener nuevos subconjuntos de tuplas, y en definitiva nuevas relaciones. Este conjunto de operadores que actúan sobre las relaciones forma lo que llamamos *álgebra relacional* [13, 14].

Estas operaciones son de dos tipos. Las del primero son las clásicas en la teoría de conjuntos y operan considerando las relaciones como conjuntos de tuplas. Entre ellas están la unión, intersección, diferencia y producto cartesiano. En el segundo grupo, con operaciones tales como selección, proyección y unión natural o cruce, el proceso de construcción depende ya del contenido de las tuplas y de los atributos presentes en las relaciones. Cada consulta es descrita como una composición de estas operaciones.

Esta diversidad de operaciones mejora la expresividad del lenguaje de consulta, aunque no todas ellas son independientes. Se ha demostrado en la teoría que $\{\sigma, \pi, U, -, x\}$ constituye un conjunto suficiente o completo para desarrollar el álgebra relacional.

### 2.2. Deducción en bases relacionales de datos

Aunque el álgebra relacional es un sistema de consulta completo, en el sentido de CODD, uno de los problemas que se plantea es su flexibilidad como lenguaje. Tal y como hemos dicho, la interrogación a la base se formula como composición de operadores algebraicos. Tiene pues un carácter finalmente funcional. Esto hace que en ciertos casos su propia complejidad no deje adivinar el significado intuitivo de la petición y que en otros se pasen por

algo las propias relaciones de dependencia entre los atributos, las cuales, de quedar establecidas mediante reglas clausales, pueden favorecer la capacidad del sistema para recurrir a la inferencia y contribuir a fijar ciertos hechos o resultados nuevos.

Basando la interrogación en un encadenamiento de inferencias, podemos hacer que sea un conjunto de reglas deductivas, incorporado a la base *factual* de datos, el que sustituya en dicho contexto a un conjunto de operadores que por universales resultan rígidos y de confusa interpretación. La formalización del sistema en tales términos supone un replanteamiento lógico de la situación. Partiendo de las bases de datos relacionales es aconsejable buscar algún enlace que permita aprovechar una metodología tan reconocida como la programación lógica. El propósito sería hacer del propio lenguaje PROLOG el instrumento de interrogación y consulta.

Recordemos que el lenguaje PROLOG está basado en una especificación lógica de aquellas situaciones o estados a representar en términos de cláusulas de Horn. Estas cláusulas representan combinaciones inferenciales de predicados de primer orden: Cuando la inferencia no requiere condiciones previas hablamos de *hechos* y en caso contrario de *reglas.* En realidad podemos hacer corresponder los predicados PROLOG a las relaciones de las bases de datos, mientras que sus instancias estarían representadas en la base de hechos. En este sentido existe un estrecho paralelismo entre los conceptos propios de las bases relacionales y los de la programación lógica. A su vez las relaciones entre estos predicados se establecen mediante inferencias condicionales recogidas en un conjunto de reglas.

La interrogación se basa en un objetivo que puede ser formulado mediante la combinación de predicados y cuyas posibles instanciaciones son revisadas a través de los hechos y reglas por el motor de inferencia del sistema PROLOG. En qué medida puede ser acoplado este mecanismo de inferencia a las bases convencionales es una de las cuestiones básicas que debe ser respondida si queremos dotar de cierta capacidad deductiva a los sistemas al tiempo que mantenemos en ellos una sólida estructura algebraica.

Un ejemplo sencillo podría darse a partir de las relaciones contenidas en la figura 1. Si las relaciones son expresadas mediante los predicados ternarios *filia* y *local* la lista de profesores que están en el edificio *azul* se obtendrá con ayuda de la regla

```
habita(Prof,Edif) :– local(Dpto, Edif, _), filia (Prof, _, Dpto).
```

donde el símbolo: – debe ser entendido como el “si” condicional y el predicado *habita* se verifica para aquellas instancias que hacen cierto cada uno de los que le siguen, al proponer el objetivo siguiente:

```
?– habita(P,azul).

P = Pons
```

Se comprende que el objetivo juega en este caso el papel de la consulta en las bases de datos, al tiempo que el juego de reglas y el motor de inferencia sustituyen a los mecanismos internos propios del álgebra relacional. Otro caso sencillo consiste en interrogar sobre el volumen total de los edificios en *azul.* Para ello introducimos las reglas

```
suma([],0).

suma([X| Y] ,S) :– suma(Y,S1), S is X+S1.

total(Edif,Tot):–bagof(Vol,Dpto^local(Dpto,Edif,Vol),L),suma(L,Tot).
```

Para obtener el resultado proponemos un objetivo sobre el predicado *total*

?– total(azul,Tot).

Tot = 38

No obstante, por atractivo que resulte el planteamiento, surgen en el acoplamiento de las bases de datos y el sistema de consultas PROLOG problemas que es preciso considerar. El primero de ellos responde a la distinta naturaleza de los sistemas: PROLOG en sus implementaciones es un sistema procedural que fija orden de consulta tanto en la base factual como en la de reglas, mientras que las bases de datos no son procedurales sino orientadas a conjuntos, ajenas por tanto al orden. Este hecho hace que el acoplamiento imponga a la base una mecánica característica de PROLOG.

Las características propias del intérprete PROLOG, con su búsqueda en profundidad de los predicados y la consiguiente inclusión del predicado de poda "cut" parecen encajar mal con un sistema de base conjuntista en que la determinación de las instancias es simultánea. Tras un primer período de propuestas de acoplamiento, la evolución de la programación lógica en este entorno de las bases de datos condujo a la definición de un lenguaje próximo a PROLOG llamado DATALOG. Este lenguaje se propone como una versión no procedural, basada en una búsqueda en amplitud y que conserva la estructura clausal característica de la programación lógica.

En cada una de las soluciones hablamos respectivamente de acoplamiento o de integración del lenguaje para la programación de las consultas, a sabiendas de que esta última alternativa es una meta inalcanzable en sistemas físicos basados en el procesamiento secuencial de operaciones. Parece más propio a la hora de considerar implementaciones concretas (cf. [12]) una distinción entre un acoplamiento fuerte y uno débil. En todo caso ninguna de estas cuestiones afecta a la naturaleza de la información considerada, que hasta ahora hemos considerado como precisa y conocida. La vaguedad, en lo que se ha comentado, afecta a la integración de los sistemas. Veremos ahora en qué modo la vaguedad del dato repercute sobre los sistemas de información hasta ahora estudiados.

Antes de entrar en el análisis de las propuestas de lenguajes o sistemas de información en que se ha incorporado de forma efectiva un tratamiento de la incertidumbre, debemos considerar un marco conceptual en el que admita una descripción clara y rigurosa. En la teoría de subconjuntos borrosos se fundamentan muchos de los sistemas que contemplan el tratamiento de información incierta o vaga.

### 2.3. Incertidumbre y subconjuntos borrosos

La teoría de subconjuntos borrosos de ZADEH [52] resulta fundamental a la hora de describir una extensión del modelo relacional clásico, la cual puede ser entendida en un doble sentido. En el enfoque más sencillo las relaciones borrosas se establecen entre datos precisos [55], expresando con un valor en [0,1] el grado de interrelación entre ellos. En otro sentido más amplio, se puede llevar la borrosidad a los propios datos, originándose modelos en los que las relaciones utilizan información imprecisa [38].

Como punto de partida consideremos el concepto de subconjunto borroso. Dado $X$ conjunto referencial no vacío, un subconjunto $A$ ordinario queda determinado por su función característica $\chi_A : X \to \{0, 1\}$, que se define

$$\chi_A(x) = \begin{cases} 1 & si\ x \quad A \\ 0 & si\ x\ /\ A \end{cases}$$

Al ampliar el recorrido de la función al intervalo real [0,1], queda caracterizado otro tipo más amplio de subconjuntos de $X$ a los que se denomina *conjuntos borrosos* [52], en los que el grado de pertenencia a $A$ de un elemento $x \quad X$ viene dado por $A(x) \quad [0,1]$.

Dados dos subconjuntos borrosos $A$ y $B$ en el referencial $X$, se entiende que $A = B$ siempre y cuando se tenga $A(x) = B(x)$ para todo $x \quad X$, y que $A \subseteq B$ *cuando* $A(x) \leq B(x)$ para todo $x \quad X$.

ZADEH propuso interpretar las operaciones unión, intersección y complementación de subconjuntos borrosos mediante

$$(A \quad B)(x) = A(x) \vee B(x)$$

$$(A \cap B)(x) = A(x) \wedge B(x)$$

$$A^c(x) = 1 - A(x)$$

donde $\vee, \wedge$ representan el supremo e ínfimo en [0,1].

Los subconjuntos borrosos pueden ser representados mediante familias de ordinarios. Dado $\alpha \quad [0,1]$ y $A$ borroso en $X$, llamaremos a $A_\alpha = \{x \quad X \mid A(x) \geq \alpha\}$ su corte de nivel $\alpha$. La llamada *identidad de resolución* establece que $A = \quad_{\alpha \ (0,1]} \alpha \cdot A_\alpha$, es decir

$$A(x) = \sup_{\alpha \ (0,1]} \min(\alpha, A_\alpha(x)) \qquad x \quad X$$

También es de importancia el *principio de extensión* [55], que permite extender funciones $f : X \rightarrow Y$ de modo que admitan como argumentos y valores subconjuntos borrosos de $X$ e $Y$: Para $A$ borroso en $X$ se define uno en $Y$ con

$$f(A)(y) = \begin{cases} \sup_{y=f(x)} A(x) & si\ \ f^{-1}(y) \neq 0 \\ 0 & si\ \ no \end{cases}$$

Otra expresión análoga se obtendría para varias variables. Para más detalles puede consultarse KLIR [28].

Las relaciones borrosas se definen como subconjuntos borrosos de productos cartesianos. En el caso de una binaria $R$ se dice que es

| | | |
|---|---|---|
| *reflexiva* | *si* | $R(x,x) = 1 \quad x \quad X$ |
| *simétrica* | *si* | $R(x,y) = R(y,x) \quad x,y \quad X$ |
| *transitiva* | *si* | $R(x,z) \geq \sup \min (R(x,y), R(Y,z)) \quad x,z \quad X$ |

$R$ es una *proximidad* si es reflexiva y simétrica, y *similaridad* si además es transitiva.

El propio ZADEH [53] muestra las condiciones en que las similaridades, como versión borrosa de las relaciones de equivalencia, determinan particiones o recubrimientos borrosos de $X$. Uno de los resultados básicos apunta que, cuando $S$ es similaridad, cada corte de nivel $S_\alpha$, para $\alpha \quad [0,1]$, constituye una relación de equivalencia y determina por tanto una partición de $X$.

La última de las propiedades puede también ser expresada usando la composición de relaciones, que se define para $R,S \quad F(X \times X)$ como

$$(R \text{ o } S)(x,z) = \sup_{y} \min(R(x,y),R(y,z))$$

en cuyo caso la transitividad de *R* queda caracterizada mediante la condición *R* o *R* $\subseteq$ *R.* Cuando no se cumple dicha propiedad son relevantes las relaciones $R^2$, $R^3$,... y se demuestra que la menor relación transitiva que incluye a R es la llamada *clausura transitiva*

$$\hat{R} = \bigcup_{n \geq 1} R^n$$

De la clausura transitiva de relaciones de proximidad resultan similaridades.

### 2.4. Sistemas para el tratamiento de la imprecisión

Aunque el modelo relacional constituirá nuestra referencia en cuanto a organización y tratamiento de los datos, han existido con anterioridad entornos de gestión de información, rudimentarios desde el punto de vista de la integridad e independencia de los datos, que han ido dando solución a la incompletitud, la imprecisión y la vaguedad de los datos.

Calificar la información es una manera de describir la complejidad de la situación de partida. El catálogo *incompleto, impreciso, vago* es una forma convencional de diferenciar las soluciones que han ido siendo aportadas.

Las primeras soluciones a la información *incompleta* fueron consideradas en el marco de las bases de datos relacionales por el propio CODD [13]. La introducción del valor nulo como valor posible en un atributo, pretendía evitar inconsistencias, pero con su independencia del dominio del atributo se empieza a cuestionar la homogeneidad del modelo relacional. Los trabajos de LIPSKI [32] y GRANT [22] sobre la semántica de estos valores multiplican el número de casos, que van desde el *desconocido* al *inexistente.*

En ocasiones se dispone de datos cuya información no se ajusta al tipo de los elementos del dominio. Por ejemplo, en vez de un valor concreto se conoce un intervalo de acotación, un conjunto de valores posibles o una distribución de probabilidad. Aquí cabe ya hablar de información *imprecisa.* Su integración en el modelo relacional no siempre respeta la homogeneidad de los elementos del dominio. Como única vía para salvaguardar el modelo se han propuesto dominios y modelos de datos complejos que incluyen los casos elementales. En este punto se produce la incorporación de los subconjuntos borrosos y de las distribuciones de posibilidad como elementos que integran en su complejidad a los valores elementales, los subconjuntos ordinarios, los intervalos y las distribuciones de probabilidad, e incluso la gama de valores nulos anteriormente considerada.

El término *vaguedad* afecta a descripciones del valor de atributo mediante etiquetas lingüísticas. Palabras como `próximo', `lejano', `joven' o `alto' representan información *vaga.* Su interés depende de la posibilidad de combinarla con información precisa como 173 cm. o 21 años. La propuesta de ZADEH consiste en entender la etiqueta como una restricción borrosa sobre los valores del dominio y representarla como una distribución de posibilidad, y en última instancia como un subconjunto borroso cuyo referencial sea el dominio. Dado que apenas entraremos a considerar el uso del lenguaje natural, nuestro marco de interés se centra en el tratamiento mediante subconjuntos borrosos de información imprecisa.

En ocasiones la complejidad de las operaciones que se pretenden realizar va más allá de la complejidad de las estructuras de la información. Puede, por ejemplo, suceder que los datos sean precisos y la consulta sea imprecisa o vaga. No existiendo en un principio modelos de datos imprecisos, los primeros sistemas de consulta imprecisa o de información impre-

cisa desarrollados, ajenos al modelo relacional, carecen de un álgebra que respalde las operaciones. Podemos pues distinguir dos etapas en la asunción de la imprecisión: sistemas ponderados de recuperación de información y modelos de datos borrosos.

## 3. SISTEMAS DE RECUPERACIÓN PONDERADOS

En el sistema de recuperación borroso de TAHANI [44] el contraste entre la relación de registros $r$ y la consulta $q$ genera un índice de encaje $\gamma(q,t)$ para cada registro $t$ $r$. Cuando la consulta es del tipo `$A_i$ es $m$' donde $A_i$ representa uno de los atributos y $m$ es un término cuyo significado se describe mediante el subconjunto borroso $M$ el índice se define como $\gamma(q,t) = M(t[A_i])$, donde $t[A_i]$ representa la componente en $t$ correspondiente al atributo $A_i$. Mediante conectivas lógicas *and, or* y *not* se describen consultas más complejas. Para ellas se definen los correspondientes índices de encaje.

$$\gamma(p \text{ and } q,t) = \min(\gamma(p,t), \gamma(q,t))$$
$$\gamma(p \text{ or } q,t) = \max(\gamma(p,t), \gamma(q,t))$$
$$\gamma(\text{not } q,t) = 1 - \gamma(q,t))$$

Finalmente la respuesta del sistema a la consulta $q$ sería un subconjunto borroso $S(q)$ definido por el índice de encaje de cada registro: $S(q)(t) = \gamma(q,t)$ para cada t r.

El sistema así concebido tiene un doble apoyo: una relación de registros con información precisa y una formulación de consultas imprecisas mediante proposiciones multivaluadas. En todo caso no existe propiamente información imprecisa en la base de datos.

La imprecisión en la información de una base de datos se incorporó al ampliar las relaciones con un nuevo atributo correspondiente al grado de pertenencia de la tupla a la relación considerada. De este modo en un formato relacional convencional se representaban relacionas borrosas. Cada relación $r$ vendría dada por una aplicación $r : D_1 \times ... \times D_n \rightarrow [0,1]$, y representada por el esquema $[A_1,...,A_n,G]$ donde $G$ [0,1] representa el grado de pertenencia como atributo. La posibilidad de instalar la imprecisión en el marco de un sistema relacional convencional no debe hacernos ver más de lo que hay: Los datos siguen siendo precisos, es el valor de verdad, antes asociado a {*Cierto, Falso*}, el que se evalúa en el dominio más general $D_U = [0,1]$.

El modelo ha tenido especial incidencia teórica en el desarrollo de lenguajes para el razonamiento aproximado. Concretamente, se encuentra en el origen de los lenguajes PRUF y FRIL, y por extensión en las diversas variantes borrosas de PROLOG que posteriormente han ido apareciendo.

Manteniendo estos modelos de datos borrosos sencillos, que se basan en la ponderación de la imprecisión en las relaciones, la estructura relacional no quedaba seriamente resentida y cabía progresar hacia lenguajes de consulta que se desarrollaran de forma paralela a los convencionales.

En una línea marcada por el pragmatismo, BOSC [3] propone enlazar la teoría de subconjuntos borrosos con un lenguaje de consulta convencional en las bases de datos como SQL. De este modo intenta crear un formato para las consultas imprecisas en el marco de los sistemas disponibles comercialmente. Al igual que en los dos casos anteriores el modelo de datos es una representación de las relaciones borrosas con un dominio $G$ en que se refleja el grado de pertenencia de la tupla a la relación.

En SQL la selección de de tuplas de relaciones que cumplen determinadas condiciones se logra mediante el bloque

SELECT < atributos > FROM < relaciones > WHERE < condicion >

La extensión de este bloque propuesta por BOSC para el tratamiento de consultas imprecisas, además de la definición de condición borrosa, incorpora un parámetro que indica el número *n* de respuestas solicitadas o bien un umbral *t* del grado de pertenencia que restrinja las respuestas mostradas. El bloque pasa a convertirse entonces en

SELECT *n/t* < atributos > FROM < relaciones > WHERE < condicion borrosa >

Las condiciones borrosas son expresiones que se componen de cuatro tipos de términos: booleanos, transformaciones, alteraciones e interacciones. Las transformaciones expresan subconjuntos y relaciones borrosos asociados a una palabra como *alto* o *aproximadamente igual*. Las alteraciones vienen a expresar variantes de la cuantificación asociando por ejemplo la palabra *muy* a la función de [0,1] en [0,1] dada por $f(x)=x^2$. Por último las interacciones expresan concexiones lógicas entre subconjuntos borrosos.

En las operaciones conjuntistas sobre tablas, como la unión, la intersección y la diferencia, se aplican a las relaciones borrosas las definiciones obtenidas al usar sobre conjuntos borrosos los operadores [*max,min,'*].

La redundancia puede darse en este modelo cuando se tienen tuplas con los mismos valores para los atributos y grados de pertenencia diferentes. La resolución en una sola tupla admite, en la extensión de SQL propuesta por BOSC, tres variantes. Al incorporar a *select* la palabra clave *unique* se le asocia el mayor grado de entre las tuplas redundantes, si la clave es *unimin* el menor y con *uniavg* el grado medio.

Teniendo en cuenta que las consultas se expresan mediante fórmulas del cálculo de predicados de primer orden parecía razonable progresar a una representación en lógica borrosa de los cuantificadores generalizados. En un trabajo posterior [5] BOSC se propone representar expresiones cuantificadas con vistas a la evaluación de consultas. Para su interpretación considera tres métodos: en el primero la agregación de las evaluaciones de las fórmulas elementales se basa en la suma de grados considerado por ZADEH; en la segunda interpretación, dada por PRADE, las agregaciones resultantes derivan en integrales de SUGENO: y en la última los cuantificadores se interpretan como operadores OWA (ordered weighted averaging) con arreglo a la teoría desarrollada por YAGER [51].

Su último trabajo en un contexto más cercano a las bases de datos relacionales y a la extensión de SQL tiene que ver con una interpretación en términos borrosos del operador relacional división[6].

Los sistemas anteriores vienen a mostrar modelos dirigidos a integrar la imprecisión como uno más de los atributos en un registro. Estos modelos preservan la naturaleza y la precisión de los valores de los atributos con dominios bien tipificados y datos hogéneos. Pero cuentan también con varios inconvenientes, algunos propios de los sistemas con datos homogéneos: dificultades de interpretación del nulo, imposibilidad de incorporar en el sistema conocimiento semántico sobre los datos, etc.; algunas de sus dificultades derivan de la semántica aplicada al atributo de pertenencia. En ocasiones este atributo puede ser interpretado como grado de pertenencia de la tupla a la relación y en otras como dependencia entre los atributos y la clave.

Una forma de integrar conocimiento semántico de los datos es incorporar formas de

imprecisión específicas en cada uno de los dominios de la relación, en vez de hacer atribuciones globales a la tupla. Las soluciones planteadas, que ya dan lugar propiamente a modelos de datos borrosos, se caracterizan en un caso por el mantenimiento de la identidad de los valores conservando la tipificación homogénea del dominio y en otro por la incorporación de valores complejos en forma de distribuciones de posibilidad. Ambas alternativas consituyen lo que BUCKLES [11] denomina modelos de datos *homogéneos* y *heterogéneos.*

## 4. MODELOS DE DATOS HOMOGÉNEOS

En este apartado tratamos modelos cuyos datos son elementos de tipo numérico o escalar en un dominio finito. Sobre los dominios de valores existentes o dominios actuales la imprecisión se puede expresar al establecer particiones de dichos dominios en clases de equivalencia formadas por elementos intercambiables o indistinguibles. El primero de los modelos fue introducido por BUCKLES y PETRY [7] en 1982, y fue posteriormente ampliado por SHENOI y MELTON [41] en 1989.

### 4.1. Modelo de Buckles y Petry

En la ampliación del modelo relacional propuesta por BUCKLES-PETRY [7] las bases de datos borrosas se componen de relaciones cuyas tuplas, a diferencia de las clásicas, admiten como valores en los distintos atributos subconjuntos no vacíos del dominio, y no necesariamente átomos o subconjuntos unitarios. Las tuplas, en este caso, serían de la forma

$$t = (t_1,...,t_n) \quad \text{con} \quad t_k \subseteq D_k \ \text{ y } \ t_k \neq \emptyset$$

Cuando el valor de una de las componentes $t_k$ de la tupla es un subconjunto, si sus elementos se interpretan como intercambiables o indistinguibles, el subconjunto es una expresión de la imprecisión de los datos, y su propia cardinalidad denota el grado de imprecisión en esa componente. Si se trata de un átomo la información es sumamente precisa y si es el dominio completo $D_k$, completamente imprecisa. Se aprecia entonces el sentido de la exigencia $t_k \neq \emptyset$.

Esta idea obligaba a dar una significación al concepto de relación en las bases de datos relacionales, menos estricta que la clásica

Definición 1 ([7]) *Llamaremos relación en base de datos borrosa a un subconjunto del producto cartesiano* $\bar{P}(D_1) \times ... \times \bar{P}(D_m)$ *donde* $\bar{P}(D_i) = \bar{P}(D_i) - \emptyset$

Esta definición acarrea la de *tupla borrosa*, que debe ser entendida como un elemento de una relación en base de datos borrosa. Por otro lado, la definición impone un modelo de datos totalmente distinto al construido a partir de las relaciones borrosas, y que ha servido de base a las soluciones mostradas en el apartado anterior. En ambos modelos aparece el modelo relacional clásico como un caso particular: allí dando grado de pertenencia 1 a todas las tuplas y aquí tomando como subconjuntos únicamente átomos.

En las definiciones planteadas no parece existir restricción sobre los subconjuntos que aparecen como valores. De hecho, el reagrupamiento en subconjuntos de elementos indiferenciables de $D_k$ se hace depender de una relación binaria que cuantifique la indistinguibilidad de los valores $t_{ki}$ de $t_k$.

En el modelo BP se postula como relación a través de la cual se introduce la imprecisión, una similaridad, es decir una relación binaria borrosa reflexiva, simétrica y *max-min*-transiti-

va. Se ha demostrado [53] que este tipo de relaciones permite establecer una partición del dominio. En efecto, si $S$ es una similaridad en $D$ para cada $\alpha$ [0,1] el corte de nivel $S_\alpha$ es una relación de equivalencia y por tanto se obtiene una partición $\Pi_\alpha$ del dominio $D$. La distinción de las clases de equivalencia correspondientes a este nivel $\alpha$ puede servir para determinar los subconjuntos válidos.

En aras de la eficiencia, el proceso de partición en clases de equivalencia puede restringirse a los datos del dominio presentes en las tuplas de la base, es decir al dominio actual o temporal, i.e.

$$\hat{D}_k = \underset{t \in r(R)}{} t_k \subseteq D_k$$

y por otro lado extenderse al producto cartesiano de los dominios $\hat{D}_1$ x...x $\hat{D}_n$.

Si el proceso de reagrupamiento en tuplas borrosas depende de las clases de equivalencia obtenidas cuando los respectivos dominios están dotados de similaridades $S_k$ y el grado de indistinguibilidad se fija mediante un nivel $\alpha_k$, podemos redefinir el concepto de tupla válida en una relación de una base de datos borrosa como

Definición 2 ([41]) Dada la relación $r \subseteq \tilde{P}(D_1)$ x,..., $\tilde{P}(D_n)$, *donde* $\tilde{P}(D_k) = P(D_k) - \emptyset$, *y las similaridades* $S_1,...,S_n$, *se dice que una tupla t r es válida respecto a particiones de niveles* $\alpha_1,...,\alpha_m$ *en los dominios actuales* $\hat{D}_1,...,\hat{D}_m$ *respectivamente, si cada componente* $t_k$ *de t es un subconjunto no nulo de una clase de equivalencia de la partición* $\Pi_{\alpha_k}$ *de* $\hat{D}_k$.

Esta restricción impone que los elementos $t_{ik}$ de los subconjuntos $t_k$ que como valores forman tuplas válidas sean efectivamente indistinguibles según las similaridades $S_k$. Por tanto no puede cualquier subconjunto ser un valor en una tupla válida.

Según esta definición, todas las tuplas de una relación clásica son válidas para cualquier nivel cuando se toma como similaridad entre elementos de los dominios la igualdad, ya que entonces las clases de equivalencia son átomos. Por tanto el modelo relacional clásico sería un caso particular.

La identidad de las tuplas válidas no está completamente determinada por la definición anterior; es preciso disponer de criterios para distinguir tuplas válidas y asegurar una relación consistente. De este modo entramos en el problema de la redundancia.

En el modelo BP la consistencia de la relación no se afronta directamente mediante el concepto de redundancia sino a través de la idea de interpretación de una tupla

Definición 3 ([7]) *Dada una tupla borrosa* $t = (t_1,...,t_n)$ *llamaremos interpretación de t a una tupla ordinaria* $(\alpha_1,...,\alpha_n)$ *donde* $\alpha_i$ $t_i$ *para cada i.*

En el caso ordinario cada tupla tiene una única interpretación y la redundancia se expresa mediante la constatación de que existe una interpretación idéntica para dos tuplas redundantes. Buckles extrapola esta idea a las tuplas borrosas, y de este modo la redundancia queda expresada en términos de las tuplas ordinarias. En línea con lo anterior, una condición suficiente para la redundancia de dos tuplas borrosas $t$ y $t'$ podría ser que posean una interpretación idéntica.

Si $T$ representa el conjunto de las interpretaciones de la tupla borrosa $t$, esta condición sobre $t$ y $t'$ podría expresarse diciendo que

$$T \cap T' \neq \emptyset$$

Cuando se dispone de una partición de $D_1$ x...x $D_n$ en clases de equivalencia la condición anterior viene a expresar que para dos tuplas $t$ y $t'$ las interpretaciones $T$ y $T'$ corresponden a la misma clase de equivalencia. Por eso la redundancia entre tuplas válidas para una partición se puede expresar

Definición 4 ([42]) *Dos tuplas válidas t y t' son redundantes cuando para cada componente los subconjuntos $t_k$ y $t'_k$ están en la misma clase de equivalencia.*

Este planteamiento justifica el siguiente resultado

Proposición 1 (Principio de interpretación única [7]) *Dada una relación borrosa cuyas tuplas son válidas respecto a una partición y no redundantes, entonces cualquier interpretación pertenece a una sola tupla.*

Demostración: En efecto si una interpretación perteneciera a dos tuplas $t$ y $t'$ de $r$ entonces $T \cap T' \neq \emptyset$. En este caso $T$ y $T'$ estarían en la misma clase de equivalencia y por tanto serían redundantes.

En todo caso la condición de doble interpretación no es necesaria para la redundancia, ya que existen otros casos con interpretación única y redundancia. Puede ser el caso en que las interpretaciones cumplen $T \cap T' \neq \emptyset$, pero ambas están en la misma clase de equivalencia.

La formulación de la redundancia en términos de clases de equivalencia permite enlazar con las hipótesis formuladas sobre la partición de $D_1$ x...x $D_n$. Al contar con la existencia de similaridades $S_k$ asociadas a los dominios y la fijación de niveles $\alpha_k$ para cada uno de ellos, podemos expresar la redundancia de tuplas borrosas del siguiente modo

Proposición 2 *Dado un conjunto de niveles $\alpha = (\alpha_1,...,\alpha_n)$ y similaridades $S_1,...,S_n$, dos tuplas válidas t y t' son redundantes si y solo si $\min_{x,y \quad t_k \quad t'_k} S_k(x,y) \geq \alpha_k$ para k=1,...,n*

Demostración: Consideremos el caso en que la relación tiene un solo atributo con dominio $D$ asociado a una similaridad $S$ donde se ha fijado un nivel $\alpha$, ya que por extensión a través del producto cartesiano podríamos obtener el resultado global.

Si los subconjuntos corespondientes a $t$ y $t'$ expresan elementos indistinguibles hasta cierto nivel $\alpha$, entonces $S(x,y) \geq \alpha$ para cualquier $x,y \quad t$ y lo mismo para $x,y \quad t'$. Si en este caso $t$ y $t'$ son redundantes entonces los subsonjuntos $t$ y $t'$ pertenenecen a la misma clase de equivalencia, es decir para cualquier par $x,y \quad t \quad t'$ los elementos son indistinguibles hasta el nivel $\alpha$. Por tanto

$$S(x,y) \geq \alpha \quad \text{para} \quad x,y \quad t \quad t' \quad \text{o bien} \quad \min_{x,y \quad t \quad t'} S(x,y) \geq \alpha$$

Recíprocamente, si $\min_{x,y \quad t \quad t'} S(x,y) \geq \alpha$, para cualquier $x,y \quad t \quad t'$ se cumple $S(x,y) \geq \alpha$. Por tanto todos los elementos de $t \quad t'$ son de la misma clase de equivalencia $H$. Como consecuencia tanto $t$ como $t'$ están contenidos en la misma clase, y son redundantes.

Esta formulación es la empleada en el trabajo de BUCKLES para introducir la redundancia, y derivar con algún esfuerzo el principio de interpretación única dado en la Prop.1.

Una vez detectada la redundancia es preciso formular operaciones orientadas a resolverla. En el caso clásico la repetición se resuelve mediante la eliminación. Aquí esta solución generaría pérdida de información. En efecto, cuando en una relación disponemos de dos tuplas cuyas componentes $k$-ésimas $t_k$ y $t'_k$ son redundantes, el dominio actual contiene a $t_k \quad t'_k$, si resolvemos eliminando una de ellas es posible que el dominio actual se vea afectado. Para

resolver la situación es preciso operar preservando la información existente en el dominio actual.

Definición 5 *Dadas dos tuplas válidas t y t' redundantes, la mezcla de ambas se obtiene al sustituirlas por una tupla cuyas componentes son* $t_k \quad t'_k$.

La definición es correcta ya que al ser redundantes las tuplas iniciales, los valores para cada componente pertenecen a la misma clase $H_k$, luego $t_k \quad t'_k \subseteq H_k$ y la tupla final es válida. Además la operación preserva la consistencia de la base de datos al pasar de una partición a otra más basta, es decir permite que la reorganización de una base de datos borrosa según una partición sea única.

Proposición 3 *Sea r una relación de tuplas válidas respecto a la partición* $\Pi$ *de* $D_1$ x...x $D_n$ *sin redundancia, al tomar una nueva partición más basta* $\Pi'$ *la reorganización por mezcla da lugar a una relación única.*

Demostración: Como para cada uno de los dominios la partición $\Pi_k$ es más fina que la $\Pi'_k$, una clase de equivalencia $H'_k$ en esta última incluye varias de la primera, es decir $H'_k = H_{1k} \quad ... \quad H_{pk}$. Consideramos en la nueva partición el conjunto *Tr* de tuplas de la relación redundantes respecto a la clase $H'_k$. Por ser válidas respecto a $\Pi_k$ se tendrá que

$$t \quad Tr \quad i \qquad \text{tal que} \qquad t_k \subseteq H_{ik}$$

y sus componentes *k*-ésimas estarán contenidas en algún $H_{ik}$. Tras la operación de la mezcla resultará que

$$\underset{t \quad Tr}{t_k \subseteq H_{1k} \quad ... \quad H_{pk} = H'_k}$$

Como en cada componente se verifica esta condición la tupla resulta válida. Además, operando de este modo para cada clase $H'_k$ obtenemos una relación única.

De este modo se puede partir de una relación con tuplas ordinarias, es decir una relación según la partición más fina, y por reorganización y mezcla obtener una única relación con tuplas válidas para cualquier partición, que necesariamente será más basta.

En el modelo de BP para un conjunto de similaridades $S_1,...,S_n$ las particiones dependen de los niveles $\alpha_k$ aplicados en cada dominio, y en tal caso podemos deducir de la proposición

Corolario 1 *Dado una similaridad para cada dominio, sea r una relación de tuplas válidas respecto al conjunto de niveles* $\alpha = (\alpha_1,...,\alpha_k,...,\alpha_n)$ *sin redundancia, y* $\alpha'_k \leq \alpha_k$, *entonces la reorganización por mezcla respecto a los niveles* $\alpha' = (\alpha_1,...,\alpha'_k,...,\alpha_n)$ *da lugar a una relación única.*

Demostración: Basta tener en cuenta que la ordenación en los niveles $\alpha$ sobre una misma similaridad *S* induce una ordenación de las particiones. En efecto, si $\alpha'_k \leq \alpha_k$, entonces $S_{\alpha_k} \subseteq S_{\alpha'_k}$ y $\Pi_\alpha$ es un refinamiento de $\Pi_{\alpha'}$. Teniendo en cuenta la proposición anterior se deduce el resultado.

Todo modelo de datos comporta una estructura de datos y un conjunto de operaciones que permita su manejo. En el modelo BP la base de datos relacional borrosa se completa con un álgebra relacional borrosa. Dicha álgebra debe incluir el álgebra relacional clásica como un caso y tener un conjunto completo de operaciones bien definidas.

En el caso clásico CODD demostró [14] que las operaciones *Unión, Diferencia, Producto, Selección* y *Proyección* forman un sistema completo y que las restantes operaciones se pueden obtener a partir de ellas. La adscripción a la operación clásica de un conjunto de niveles y

la subsiguiente mezcla para resolver la redundancia permite obtener a partir de relaciones borrosas una única relación borrosa en cualquiera de los cinco casos.

La operación de selección borrosa sobre *r*, que denotamos $\sigma_{\Theta,\alpha}(r)$, se redefine [9] en este modelo de tal modo que Θ representa una condición booleana y α es un conjunto de niveles que determinan la redundancia con arreglo a la cual se reorganizará la base de datos. Como la condiciones booleanas solo operan sobre datos elementales, la operación se puede descomponer de la forma siguiente: Formación de una relación con las interpretaciones o tuplas ordinarias, selección de tuplas ordinarias según la condición booleana y finalmente reorganización de la base en tuplas borrosas válidas para los niveles establecidos. Si llamamos *I* a la primera de las operaciones y $M_\alpha$ al última obtendremos

$$\sigma_{\Theta,\alpha}(r) = M_\alpha\ (\sigma_\Theta(I(r)))$$

Al ser las tres operaciones de que se compone, operaciones bien definidas, es decir que dan una única relación, podemos asegurar

Proposición 4 *Dada una relación borrosa r la operación selección* $\sigma_{\Theta,\alpha}(r)$ *del álgebra relacional borrosa está bien definida.*

De igual modo podemos definir las cuatro operaciones restantes incluyendo los niveles de redundancia que permitan la reorganización por mezcla, y por el mismo método demostrar su buena definición.

Es evidente, por otro lado, que las operaciones son una extensión del caso clásico, puesto que éste aparece cuando las operaciones *I* de interpretación y $M_\alpha$ de mezcla son la identidad de relaciones. Es fácil comprender que desde este punto de vista estamos ante un modelo relacional clásico al que se incorporan dos nuevas operaciones para el tratamiento de elementos no atómicos. En cierto modo es un modelo que evita la restricción impuesta por CODD en la primera forma normal. El punto fundamental que determina las limitaciones del modelo es que solo permite el tratamiento de condiciones booleanas ya que en realidad opera con la base relacional clásica subyacente al modelo.

### 4.2. Extensión de Shenoi y Melton

En el modelo BP se opta por la partición de dominios mediante relaciones de similaridad y niveles, y basados en dicha partición logramos las condiciones para la redundancia. Sabemos, sin embargo, que a partir de una relación de proximidad no necesariamente transitiva, también se puede formar una similaridad. Recuérdese que a dicha construcción la denominábamos clausura transitiva.

La propuesta de SHENOI y MELTON se basa en esta posibilidad. Para ello se parte de una relación binaria borrosa de proximidad, esto es reflexiva y simétrica, sobre cada dominio.

Definición 6 *Dada una proximidad P sobre un dominio D y un valor* α*, diremos que dos elementos x,y D son* α*-próximos, y denotamos* $xP_\alpha y$*, cuando* $P(x,y) \geq \alpha$.

En este caso no puede asegurarse que cuando $xP_\alpha y$ y $yP_\alpha z$ resulte $xP_\alpha z$. No obstante se puede construir una nueva relación a partir de *P* que es similaridad y que por tanto permite obtener una partición del dominio.

Definición 7 *Dada una proximidad P sobre un dominio D y un valor* α*, diremos que dos elementos x,z D son* α*-similares, y denotamos* $xP^+_\alpha z$*, cuando existe una sucesión* $y_1,y_2,\ldots,y_r \in D$ *tal que* $xP_\alpha y_1 P_\alpha y_2\ P_\alpha \ldots P_\alpha y_r\ P_\alpha z$

Que esta nueva relación es de equivalencia puede obtenerse directamente de la definición.

Proposición 5 ([41]) *Si P es una relación de proximidad en D entonces* $P^+_\alpha$ *es una relación de equivalencia para cualquier* α [0,1].

Demostración: La reflexividad es evidente ya que por ser $P$ reflexiva $P(x,x) = 1$, de donde resulta $P(x,x) \geq \alpha$, es decir $xP_\alpha x$. En el caso de que para $x,z$ se tenga $xP^+_\alpha z$, existirá una sucesión tal que $xP_\alpha y_1 P_\alpha y_2\ P_\alpha \ldots P_\alpha y_r\ P_\alpha z$

Como $P$ y $P_\alpha$ son simétricas podemos escribir $zP_\alpha y_r\ P_\alpha y_{r-1}\ P_\alpha \ldots P_\alpha y_1 P_\alpha x$ con lo que queda $zP^+_\alpha x$. La transitividad también se cumple ya que si $xP^+_\alpha z$ entonces existe una sucesión tal que $xP_\alpha y_1 P_\alpha y_2\ P_\alpha \ldots P_\alpha y_r\ P_\alpha z$ y si $zP^+_\alpha v$, existirá otra $zP_\alpha u_1 P_\alpha u_2\ P_\alpha \ldots P_\alpha u_s\ P_\alpha v$ y componiendo ambas cadenas se puede enlazar $x$ y $v$ $xP_\alpha y_1\ P_\alpha \ldots P_\alpha y_r\ P_\alpha z\ P_\alpha u_1\ P_\alpha \ldots P_\alpha u_s\ P_\alpha v$ Por tanto se tiene $xP^+_\alpha v$, y $P^+_\alpha$ es transitiva.

La demostración muestra la argumentación del autor, pero es fácil reformularla con los resultados obtenidos. En realidad la α-similaridad $P^+_\alpha$ es simplemente la clausura transitiva de $P_\alpha$, es decir $\hat{P}_\alpha$. Teniendo en cuenta que para $P$ proximidad sobre $D$ finito, se tiene que $\hat{P}_\alpha = \dot{P}_\alpha$, el resultado es evidente, pues siendo $\dot{P}$ una similaridad $\dot{P}_\alpha$ será una relación de equivalencia.

Tal y como vimos la reorganización de la relación por resolución de la redundancia daba lugar a una relación única gracias a la estructura ordinal de la familia $\{S_\alpha\}_{\alpha\ [0,1]}$. En el caso de $P^+_\alpha$ se obtiene un resultado equivalente.

Proposición 6 ([41]) *Para todo x,y D y valores* α,α' *tales que* α ≥ α', si $xP^+_\alpha y$ entonces $xP^+_{\alpha'} y$.

Demostración: Basta tener en cuenta que para cualquier relación $P$ y $x,y$ cualesquiera $P(x,y) \geq \alpha \geq \alpha'$, luego si $xP_\alpha y$ entonces $xP_{\alpha'} y$. A través de la sucesión el resultado puede ser extendido a $P^+_\alpha$.

Por tanto se pueden aplicar a este modelo los resultados sobre el álgebra relacional borrosa obtenidos en el modelo BP. Por otro lado este modelo queda incluido en el de SHENOI. La estrategia de reducción de SHENOI pasa por la proposición

Proposición 7 ([41]) *Dada una similaridad S sobre un dominio D y un valor* α [0,1], *entonces* $S_\alpha$ *y* $S^+_\alpha$ *generan las mismas clases de equivalencia.*

Demostración: Dado α [0,1], si $xS_\alpha z$ entonces $xS^+_\alpha z$. Si partimos de $xS^+_\alpha z$ existirá una sucesión $y_1,\ldots,y_r$ tal que $xS_\alpha y_1 S_\alpha y_2\ S_\alpha \ldots S_\alpha y_r\ S_\alpha z$ y por transitividad $xS_\alpha z$.

Una vez más la reformulación del resultado según la teoría de relaciones lo hace obvio, ya que cuando $S$ es similaridad coincide con su propia clausura transitiva, de modo que $\hat{S} = S$. Y al aplicar los cortes para cualquier α se tendrá $\hat{S}_\alpha = S_\alpha$.

En definitiva el modelo SM extiende el modelo BP al partir de relaciones más generales, pero básicamente vuelve a formar particiones de los dominios basadas en la definición de proximidades y en la fijación de niveles. La vía de paso entre ambos modelos es la construcción de la clausura transitiva de las proximidades.

## 5. MODELOS DE DATOS HETEROGÉNEOS

Las extensiones anteriores se aplican a dominios actuales finitos y a estructuras basadas en tipos homogéneos. Estas premisas limitan de manera excesiva la representación de

situaciones inciertas. Pensemos en la representación de números reales borrosos o en la de predicados como `joven', asociados a un dominio numérico. La teoría de la posibilidad [56][21] constituye un marco para la representación de tales situaciones, y partiendo de ella se han descrito modelos de datos para la construcción de bases borrosas. En este apartado describiremos la solución planteada por BUCKLES y PETRY [9] para el tratamiento de números borrosos, y a continuación expondremos el modelo de datos propuesto por PRADE y TESTEMALE, basado en distribuciones de posibilidad.

### 5.1. Ampliación del modelo BP con números borrosos

Evitamos una descripción de propiedades de los números borrosos que entenderemos aquí como conjuntos borrosos $q$ con la recta real como referencial e intervalos como soporte, y cuya función de pertenencia se describe mediante una función continua y cóncava $q(x)$ con $x$ R. Una definición más precisa y sus propiedades fundamentales pueden consultarse en el trabajo de DUBOIS y PRADE [21].

La ampliación del modelo BP a dominios no discretos formados por números borrosos exigía algún criterio para establecer la redundancia entre dos números cualesquiera $q,q'$ del dominio $Q$ de números borrosos. Este criterio debe aparejar la partición del dominio tras la fijación de un nivel α. Se trata aquí también de definir una similaridad $S$ en $Q$, para de este modo obtener relaciones de equivalencia $S_\alpha$.

La definición propuesta por BUCKLES y PETRY para establecer el `parecido' de dos números borrosos resulta ajena a la solución mostrada en el apartado anterior para dominios discretos, ya que no se parte de una relación binaria entre los elementos del referencial. Por otro lado su desarrollo es en muchos aspectos incompleto. De ahí que se haya hecho necesario reconstruir enteramente la línea de argumentación.

La primera idea sería que existe parecido a nivel α entre $q$ y $q'$ cuando el conjunto $(q \quad q')_\alpha$ es convexo.

Definición 8 ([9]) *Sea* α [0,1] *un valor cualquiera, entonces diremos que los números borrosos q,q' son* α*-próximos*[1], *y denotamos* $qM^\alpha q'$, *cuando para* $x,y \in (q \quad q')_\alpha$ *y para todo* β [0,1] *se cumple que* $z = \beta x + (1-\beta)y \quad (q \quad q')_\alpha$.

Sabemos que en la recta real con la topología ordinaria un subconjunto es convexo si y solo si es un intervalo cerrado. Por otro lado es intervalo cerrado si y solo si es conexo. Esta circunstancia permite asegurar que

Proposición 8 *Dado un valor* α*, diremos que los números borrosos q,q' verifican* $qM^\alpha q'$ *si y solo si* $(q \quad q')_\alpha$ *es conexo, es decir si* $q_\alpha \cap q'_\alpha \neq \emptyset$.

Esta propiedad sugiere la posibilidad de definir de forma general, y sin dependencia de nivel, una relación borrosa de `parecido' entre números borrosos. Este es el punto de partida en que se sitúa la versión de SHENOI y MELTON.

Definición 9 ([43]) *Dados dos números borrosos q,q' definimos la relación P, mediante la función* $P(q,q') = \sup\{\alpha \; ½ q_\alpha \cap q'_\alpha \neq \emptyset\}$

1. Parece conveniente sustituir aquí la denominación original `similar' por próximo, teniendo en cuenta que la α-proximidad generará una proximidad en el sentido de ZADEH.

Esta relación generaliza la condición de α-próximos al resultar

Proposición 9 *Dados dos números borrosos q,q', si P es la relación anteriormente defini - da, se cumple q,q' son α-próximos si y solo si* $qP_\alpha q'$.

Demostración: Basta ver que $P_\alpha = M^\alpha$, y el resultado es consecuencia de las equivalencias

$$qP_\beta q' \quad \leftrightarrow \quad \sup\{\alpha \tfrac{1}{2} q_\alpha \cap q'_\alpha \neq \emptyset\} \geq \beta \quad \leftrightarrow \quad q_\beta \cap q'_\beta \neq \emptyset \quad \leftrightarrow \quad qM^\beta q'$$

Por tanto la α-proximidad se establece a través de una relación obtenida mediante corte de nivel de *P*. Por otro lado, dicha relación *P* es una proximidad en el sentido de ZADEH, i.e. es una relación reflexiva y simétrica, cuando se aplica a números borrosos normales. Dichos números deben cumplir la propiedad sup $q(x) = 1$.

Proposición 10 *Dados dos números borrosos normales q,q', si P es la relación anterior - mente definida, P es una proximidad.*

Demostración: La relación es reflexiva ya que

$$P(q,q) = \sup\{\alpha \tfrac{1}{2} q_\alpha \cap q_\alpha \neq \emptyset\} = \sup\{\alpha \tfrac{1}{2} q_\alpha \neq \emptyset\} = \sup q(x) = 1$$

La simetría a su vez es consecuencia inmediata de la simetría de la intersección.

Podemos ahora retomar la segunda definición dada por BUCKLES con vistas a construir la partición.

Definición 10 ([9]) *Sea* α [0,1] *un valor cualquiera, entonces diremos que los números borrosos q,q' son α-similares, y denotamos* $qP^+_\alpha q'$, *cuando existe una sucesión de cero o más números borrosos* $q_1,\ldots,q_r$ *tales que* $qP_\alpha q_1\ P_\alpha q_2\ P_\alpha \ldots\ P_\alpha q_r\ P_\alpha q'$

Teniendo en cuenta que partimos de una proximidad *P*, esta construcción permite razonar de igual modo que en el modelo SM. Realmente las relaciones definidas son cortes de nivel de la clausura transitiva $\hat{P}$ de *P*. Como esta relación es una similaridad, tendremos que $\hat{P}_\alpha = P^+_\alpha$ son relaciones de equivalencia y definen particiones en el conjunto de números borrosos normales.

Aunque no se cita expresamente la clausura transitiva y se desconocen sus propiedades, que abrevian considerablemente el desarrollo propuesto tanto en el modelo BP como en el SM, cabe indicar que en la ampliación de BUCKLES y PETRY a números borrosos la construcción de la partición resulta al construir la clausura transitiva de una proximidad. Aunque el desarrollo posterior de estas ideas adquiere coherencia y rigor en el modelo SM, es justo indicar que la relajación de las condiciones que SHENOI [41] postula ya había sido contemplada por BUCKLES. Por otro lado en ninguno de los modelos se establece la relación entre esta construcción y la clausura transitiva.

### 5.2. Modelo de Prade y Testemale

La introducción de las distribuciones de posibilidad (propiamente las funciones de distribución) como valores de atributo en una base de datos es llevada a cabo por UMANO [48] en 1982 en el sistema FREEDOM-0. En él se propone un modelo de datos, que incorpora la imprecisión a dos niveles: por un lado los datos se representan como distribuciones de posibilidad y por otro se precisa en cada registro un grado que refleja la pertenencia del registro a la relación o la interrelación de los valores.

El modelo de UMANO incorpora a los dominios $F(U_i)$, que representan las distribuciones de posibilidad sobre universos $U_i$, un valor para el tratamiento de valores indefinidos. El lenguaje de consulta no es propiamente relacional, se trata mas bien de un lenguaje orientado por la resolución lógica de predicados. En todo caso es interesante saber que en la instrucción QUERY} sobre la que se basan las consultas, la compatibilidad o inclusión entre dos predicados borrosos $F_1$,$F_2$ se valora mediante el índice

$$fcont(F_1,F_2) = \frac{\sum_i \min\,(F_1(u_i), F_2(u_i))}{\sum_i F_1(u_i)}$$

del que deriva un índice de igualdad

$$feq(F_1,F_2) = \min(fcont(F_1,F_2),fcont(F_2,F_1))$$

No tratándose de un modelo relacional en sentido estricto no se plantea propiamente la cuestión de la redundacia.

El modelo planteado por PRADE y TESTEMALE recoge algunas de las propuestas de UMANO para situarse en un marco más claramente relacional. Las relaciones dejan de ser borrosas, pero los datos son distribuciones de posibilidad representadas por subconjuntos borrosos normales. En este marco su trabajo se orienta a la representación de la información incompleta y a la formulación de operaciones que permitan crear un lenguaje de consulta para este modelo de datos.

Para la representación de la información incompleta se incorpora a cada dominio un elemento *e*, de modo que los dominios se denotan $\bar{D} = D \quad \{e\}$. Una relación se describe entonces como

$$r \subseteq F\,(\bar{D}_1)\ x...x\ F\,(\bar{D}_n)$$

a diferencia de las relaciones según UMANO

$$r : F\,(D_1) \quad \text{NULL}\ x...x\ F\,(D_n) \quad \text{NULL} \rightarrow [0,1]$$

No entraremos en este caso a describir la semántica del elemento *e*. Baste decir que su posibilidad junto con la de los restantes elementos permite clasificar los casos de información desconocida, no aplicable, etc., a diferencia del modelo UMANO en que se les daba un valor único NULL.

El álgebra relacional se extiende a fin de manejar distribuciones de posibilidad. Las operaciones conjuntistas se extienden de forma natural a los nuevos dominios, pero el operador de selección requiere un estudio más detallado.

La Θ-selección actúa sobre la relación seleccionando las tuplas que cumplen cierto tipo de condiciones. Estas condiciones representan consultas ambiguas, que pueden ser simples o compuestas.

Las condiciones simples representan comparaciones y admiten dos formas básicas $A\Theta a$ y $A\Theta B$. En la primera se representa, para un objeto o tupla dados, la comparación a través del comparador Θ entre la distribución correspondiente al atributo *A* y la distribución constante *a*. En la segunda la comparación se establece enter las distribuciones correspondientes a dos atributos *A* y *B*. El resultado de la selección es una doble relación con las tuplas que posiblemente satisfacen la condición y con las tuplas que necesariamente la satisfacen.

Las condiciones compuestas se obtienen al enlazar las condiciones simples con las conectivas lógicas *not, and* y *or*.

Dado un objeto o tupla $x$ de la base, si $A(x)$ representa el valor para el atributo $A$ dado por una distribución de posibilidad en $D$. La medida de la posible satisfacción de la condición $A\Theta a$ se establece determinando la posibilidad de que el valor para $A$ de la tupla $x$ esté en relación $\Theta$ con $a$. Si representamos ``estar en relación $\Theta$ con $a$" mediante la distribución $\Theta_a$, la posibilidad de la proposición ``$A(x)$ es $\Theta_a$" será

$$\Pi_{A(x)}(\Theta_a) = \sup_{d \quad D} \min (\pi_{A(x)}(d), \Theta_a(d)) \tag{1}$$

Por otro lado, y teniendo en cuenta que el comparador $\Theta$ se puede representar mediante una relación binaria borrosa, la comparación con el borroso que representa a $a$ podría reflejarse con el subconjunto$\Theta_a$

$$\Theta_a(d) = \sup_{d' \quad D} \min (\Theta(d,d'), a(d')) \tag{2}$$

Este conjunto viene a representar la proyección sobre una componente de la intersección entre la relación $\Theta$ y la extensión cilíndrica de $a$, es decir

$$\Theta_a = \text{proj}(\Theta \cap c(A))$$

La integración de las ecuaciones (1) y (2) daría

$$\Pi_{A(x)}(\Theta_a) = \sup_{d,d' \quad D} \min (\pi_{A(x)}(d), \Theta(d,d'), a(d'))$$

Por su parte, la medida de necesidad se obtiene a partir de su definición

$$N_{A(x)}(\Theta_a) = 1 - \Pi_{A(x)}(\Theta_a^c)$$

y da lugar a

$$N_{A(x)}(\Theta_a) = \inf_{d \quad D} \max (1 - \pi_{A(x)}(d), \Theta_a(d))$$

Los valores obtenidos permiten construir las dos componentes del operador selección para la condición $A\Theta a$ sobre la relación $r$ como

$$\sigma\, \Pi(r; A\Theta a) = \Pi_{A(x)}(\Theta_a)$$

$$\sigma\, N(r; A\Theta a) = N_{A(x)}(\Theta_a)$$

y escribir

$$\sigma(r; A\Theta a) = (\sigma\, \Pi(r; A\Theta a), \sigma\, N(r; A\Theta a))$$

De forma semejante se obtiene las medidas de posibilidad y necesidad para la comparación entre los valores de dos atributos $A$ y $B$ como

$$\Pi_{A(x),B(x)}(\Theta) = \sup_{d,d' \quad D} \min (\Theta(d,d'), \pi_{A(x)}(d), \pi_{B(x)}(d'))$$

$$N_{A(x),B(x)}(\Theta) = \inf_{d,d' \quad D} \max (\Theta(d,d'), 1 - \pi_{A(x)}(d), 1 - \pi_{B(x)}(d'))$$

y permiten generar $\sigma(r; A\Theta B)$. El estudio del álgebra relacional se completa con las condiciones compuestas y con el análisis y extensión de las restantes operaciones. Los criterios son semejantes a los expresados en la selección y no resultan relevantes para el estudio que nos proponemos.

A modo de resumen es preciso destacar cómo las únicas relaciones consideradas entre los elementos de $D$ son los comparadores $\Theta$ y que son éstos los que desarrollan comparadores entre distribuciones. El problema de la redundancia no puede establecerse en términos de similaridad entre tuplas, ya que no se suponen similaridades elementales en cada dominio. Como alternativa el modelo propone la medida de igualdad entre distribuciones según nivel

$$\sup_{d \quad D} \frac{1}{2}\pi(d) - \pi'(d) \frac{1}{2} \leq \varepsilon_D$$

## 6. MODELOS INTEGRALES

Con las propuestas para la integración de los modelos de Buckles-Petry y Shenoi-Melton en el marco de la representación de datos mediante distribuciones de posibilidad, se pretende unir a la idoneidad de esta representación las ventajas de la partición a través de nivel, y su repercusión en el tratamiento borroso de la redundancia. Para ello es preciso definir similaridades entre subconjuntos borrosos a base de las similaridades o proximidades elementales de los dominios. El problema de la extensión de relaciones adquiere, por tanto, en este marco nuevo relieve e importancia y las soluciones planteadas por los autores en diversos trabajos (cf. [23], [24]) podrían servir de guía para la evaluación de los modelos de redundancia.

Con este criterio denominamos modelos integrales a los que integran datos en forma de distribuciones de posibilidad y recurren a la partición según un nivel como criterio de clasificación en el dominio. A esta categoría pertenecen modelos como el de Rundensteiner et al., el modelo revisado de Prade y Testemale, y los trabajos de Chen. Trataremos brevemente el primero, incluyendo las críticas de Chen, y pasaremos a discutir el segundo en el marco de la teoría de relaciones.

### 6.1. Modelo de Rundensteiner, Hawkes y Bandler

En este modelo se propone una generalización de la redundancia tal y como se plantea en el modelo BP, extendiéndola a valores de atributo representados por distribuciones de posibilidad. Como instrumento se considera una relación de parecido (resemblance) en cada uno de los dominios $D_i$ base de las distribuciones de posibilidad. Dichas relaciones $res_i$ se definen sobre elementos, no sobre distribuciones, con propiedades reflexiva y simétrica lo que nos permite asimilarlas a proximidades en $D_i$.

La fijación de niveles $\alpha = (\alpha_1,...,\alpha_n)$ para cada uno de los componentes de la relación permite obtener si no una partición de los dominios, sí un recubrimiento de los mismos mediante las clases resultantes de la relación $res_\alpha$.

Para que dos tuplas sean redundantes se propone una doble condición: que cada uno de los elementos de las respectivas distribuciones se asemejen, es decir que pertenezcan a la misma clase inducida por $res_\alpha$, y que globalmente las distribuciones sean próximas. Estas condiciones se reflejan en la definición

Definición 11 ([40]) *Dada una relación borrosa* $r \subseteq F(D_1) \times ... \times F(D_n)$, *donde cada tupla x tiene por valor en el atributo* $A_i$ *la distribución de posibilidad* $\pi_{Ai}$ *sobre el dominio* $D_i$; *consideremos para cada* $i = 1,...,n$ *una relación* $res_i$ *de parecido (proximidad) en* $D_i$ *y fijemos un umbral* $\alpha_i$, *y un umbral de encaje* $\beta$ [0,1], *entonces dos tuplas x,x' de r son redundantes según* $\alpha = (\alpha_1,...,\alpha_n)$ *y* $\beta$ *cuando se cumple para* $i = 1,...,n$

$$\min_{\substack{u,v \quad x_{Ai} \; x'_{Ai} \\ \pi_{A_i}(v) > 0,\ \pi'_{Ai}(v) > 0}} res_i\,(u,v) \geq \alpha_i$$

$$\min_{d \quad D_i} (1 - \tfrac{1}{2}\pi_{Ai}\,(d) - \pi'_{Ai}\,(d)\tfrac{1}{2}) \geq \beta$$

A la vista de la definición destacan los siguientes hechos: 1) El segundo criterio, de encaje entre distribuciones, es semejante al utilizado por PRADE y TESTEMALE, y además no deriva de las proximidades $res_i$ introducidas en los distintos dominios, sino que es independiente de ellas; 2) el reagrupamiento en clases es independiente del grado $\pi_{Ai}\,(u)$ que tenga cada elemento *u*. Unicamente se exige que no sea nulo. Y por último 3) aunque la resolución se establezca por unión no podemos asegurar que la relación resultante sea única.

Debemos a CHEN la formulación de ejemplos que muestran algunas de las inconsistencias intuitivas del modelo. Para ello se toma de RUNDENSTEINER la relación de parecido *res* definida para el dominio $D = \{Mb, B, M, D, Md\}$ mediante una matriz para la que en particular se tiene $res(Mb,D)=0.1$. Al aplicar la definición anterior resulta que si $\alpha > 0.1$ las distribuciones $\pi = \{0.2/Mb, 1.0/D\}$ y $\pi' = \{0.2/Mb, 1.0/D\}$ pese a ser iguales no son redundantes. En efecto, aunque el segundo criterio es válido para todo $\beta$, en el primero se obtendría

$$min(res(Mb,Mb),res(Mb,D),res(D,D)) = \min(1,0.1,1) = 0.1 < \alpha$$

y por tanto fallaría por la elección de $\alpha$.

Por otro lado, si en la matriz tenemos $res(Mb,B) = 0.8$, dos distribuciones como $\pi = \{1/Mb, 0.85/B\}$ y $\pi' = \{0.9/B\}$ cumplirán el primer criterio para cualquier $\alpha \geq 0.8$, pero en el segundo tendremos $\min(1 - (1 - 0), 1 - (0.9 - 0.85)) = 0$

Ambos resultados muestran la discordancia de los criterios y el carácter escasamente intuitivo de la definición.

### 6.2. Nuevos criterios de redundancia

Como ya hemos indicado, la elaboración de criterios para determinar la redundancia de las tuplas y para establecer el encaje de una tupla con una condición ambigua es un campo idóneo para la aplicación de la teoría de relaciones extendidas [24]. En ambos casos, si el modelo de datos utiliza distribuciones de posibilidad, nos vemos abocados a comparar subconjuntos borrosos y en definitiva a construir pseudoigualdades de segundo orden (cf. [25]).

Algunas de las paradojas observadas llevaron a PRADE y TESTEMALE a revisar el criterio de redundancia ofrecido en su modelo. En un nuevo trabajo [39] analizan posibles criterios de similaridad entre subconjuntos borrosos. En este caso parten de una relación de similaridad *S* en el dominio *D* de las distribuciones, que derivaría de una métrica en *D*.

En lo sucesivo supongamos que las distribuciones están representadas por dos subconjuntos borrosos *A,B* en *D*. La primera propuesta planteada, aplicando al grado de pertenencia de *S* el principio de extensión, sería

$$\tilde{S}\,(A,B)(z) = \sup_{s(x,y)=z} \min(A(x),B(y))$$

Del estudio del caso en que *A,B* son ordinarios infieren que el valor que determina la similaridad entre ambos es

$$\min_{\substack{x \quad A \\ y \quad B}} S\,(x,y)$$

y que en términos borrosos se refleja con la expresión

$$S^e\,(A,B) = \min_{x,y \quad D} \max\,(S(x,y), 1 - A(x), 1 - B(y))$$

La aparición de ciertos casos antiintuitivos desaconseja este valor como índice de desencaje de los subconjuntos. En su lugar proponen la aplicación de una extensión borrosa de la distancia de Hausdorff.

Partiendo de la definición de distancia de Hausdorff en el caso ordinario como el valor

$$H\,(A,B) = \max\,(\sigma(A,B),\, \sigma(B,A))$$

donde $\sigma(A,B)$ representa el valor

$$\sigma(A,B) = \inf\,\{\alpha\ ½A \subseteq_{y \quad B} \{x\ ½S(x,y) > \alpha\}\}$$

Con base en σ se construye el número

$$\tilde{\sigma}\,(A,B)\,(r) = \sup\,\{\alpha ½r = \sigma(A_\alpha, B_\alpha)\}$$

y finalmente se obtiene

$$\tilde{H}\,(A,B) = \tilde{m}ax\,(\tilde{\sigma}(A,B),\, \tilde{\sigma}(B,A))$$

utilizando el operador extendido *m̃ax* a los números borrosos resultantes.

La evaluación de la similaridad se realiza previa fijación en el dominio de un umbral Λ, que en este caso también es borroso. El criterio de redundancia resulta entonces

$$\tilde{H}\,(A,B) \subseteq \Lambda$$

Finalmente se propone una resolución de la redundancia por unión de los subconjuntos. La discusión posterior revela que la relación resultante de estas resoluciones en una base de datos no es única.

## 7. MODELOS DE SISTEMAS DEDUCTIVOS BORROSOS

Hasta ahora hemos centrado nuestra atención en los distintos modelos de representación de la información imprecisa, poniendo especial énfasis en el problema planteado por la redundancia de información. Es momento de retomar el problema de la consulta imprecisa, si bien aplicado a un contexto o modelo de datos borrosos simple. En este marco, además de la implementación de variantes borrosas de lenguajes tan convencionales como SQL realizada por BOSC ET AL. y que hemos considerado anteriormente, han ido apareciendo formas de interrogación basadas en métodos deductivos. Se trata de sistemas PROLOG, aquí llamados fProlog, que tienen su origen en el sistema FRIL de procesamiento de reglas ponderadas y en el lenguaje PRUF de proposiciones y predicados borrosos.

### 7.1. Sistema FRIL

La creación de sistemas deductivos conlleva un modelo de representación de datos imprecisos. En el caso de FRIL, BALDWIN [2] propone la representación de predicados

mediante relaciones borrosas y la incorporación de medios para calcular la incertidumbre de predicados compuestos y lograr cierto tipo de inferencia aproximada. En posteriores trabajos los predicados aparecen como cláusulas de Horn [33] y el sistema se postula como una versión de PROLOG que incorpora al procedimiento inferencial el factor de incertidumbre en hechos y reglas.

En palabras de sus autores, FRIL `es un lenguaje de programación lógica que permite la inferencia en condiciones de incompletitud y de incertidumbre, además del proceso de razonamiento puramente deductivo de Prolog'. En la representación de los hechos y reglas no se afecta un grado sino un intervalo o margen de incertidumbre. Dependiendo de los contextos cabe interpretar el margen con criterios probabibilísticos o posibilísticos, asociando en este último caso la cota inferior al grado de posibilidad y la superior al grado de necesidad. Tal diversidad de interpretaciones otorga al instrumento gran interés y flexibilidad a la hora de reflejar las condiciones de la información imprecisa.

Uno de los puntos en que el lenguaje se aparta del PROLOG convencional es que en él podemos asignar márgenes de certeza tanto a una cláusula como a su negación. De hecho esta alternativa nos sitúa extramuros de la lógica clásica, pero da pie a la representación de lógicas no standard.

El fProlog desarrollado posteriormente por este equipo conserva del anterior sistema la estructura de la base de datos y el sistema de tratamiento por listas, de modo que caben interrogaciones en PROLOG ponderado e interrogaciones con FRIL. El sistema se basa en cláusulas de Horn ponderadas o borrosas y el mecanismo de inferencia se basa en la resolución de reglas mediante la búsqueda en profundidad. Sin embargo, el sistema no implementa la posibilidad de interrogaciones borrosificadas con modificadores lingüísticos tales como `muy' o `bastante', que son características de otros sistemas más evolucionados.

### 7.2. Sistemas PROLOG borrosos

Otro sistema próximo a PROLOG aparece propuesto en el trabajo de KANAI ET AL. [29], y una primera implementación se da en MUKAIDONO [35]. También UMANO progresa a una variante de PROLOG [49] desde su sistema de bases de datos borrosas FREEDOM-0 [48]. Por citar finalmente una de las que subsanan algunos de los problemas planteados, hablaremos de la versión de PROLOG borroso que fue introducida por LI y LIU [31]. En esta versión la estructura y mecanismo de inferencia es muy semejante a la del fProlog desarrollado por BALDWIN.

El intérprete de PROLOG borroso, que llegó a implementarse en una plataforma MicroVAX II, contiene el PROLOG ordinario como un caso particular. En él un programa se compone de tres tipos de elementos:

f-hechos, que se reflejan

$$P(t_1,...,t_n) : -[f]-$$

viniendo a significar que $P(t_1,...,t_n)$ es cierto en un grado $f$.

f-reglas, que denotamos

$$P : -[f] - Q_1,...,Q_n.$$

Aquí, si suponemos que el grado de certeza de $Q_i$ es $[[Q_i]]$, el grado de certeza α de todas las condiciones sería

$$\alpha = \begin{cases} \min \{[[Q_i]] \ ½ i = 1 \ldots n\} & si \quad n \neq 0 \\ 1 & si \quad no \end{cases}$$

y el de *P* se obtendría combinando α y *f* mediante un operador tipo norma, tal como el producto, el mínimo u otro.

f-objetivos, descritos como

$$? - [f] - Q_1,\ldots,Q_n.$$

donde *f* es una constante real en [0,1] que fija el umbral de certeza con que se satisfacen las condiciones, o bien

$$? - [F] - Q_1,\ldots,Q_n.$$

donde *F* representa una variable que se instanciará con los grados de cumplimiento de $Q_1 \ \ldots \ Q_n$.

En el caso de disponer el siguiente conjunto de hechos y reglas borrosas sobre la magnitud de ciudades y su importancia.

```
impor(paris) :-[0.9]-.
impor(washington) :-[1]-.
impor(napoles) :-[0.6]-.
impor(pernambuco) :-[0.4]-.
impor(londres) :-[0.9]-.
magni(X) :-[0.7]- impor(X).
```

podemos plantear distintas cuestiones sobre la magnitud de las ciudades de la base instalada. Si suponemos que el grado de confianza se construye a partir del de inferencia *f* y el de las condiciones α con el producto $f \cdot \alpha$ tendremos:

```
?-[0.6]-magni(X).
X=paris
X=washington
X=londres
```

para las ciudades grandes con un grado de confianza de 0.6. Si queremos conocer el grado con que es grande *pernambuco* haremos

```
?-[F]-magni(pernambuco).
F=0.28
```

Al formular un objetivo más general resulta

```
?-[F]-magni(X).
X=paris    F=0.63
X=washington    F=0.9
X=napoles    F=0.42
X=pernambuco    F=0.28
X=londres    F=0.63
```

Los autores desarrollaron también un sistema de interrogación del tipo SQL sobre modelos simples de bases de datos borrosas, que implementaron en ORACLE. Este sistema, denominado fSQL sirve de instrumento efectivo para la interrogación, al traducirse los objetivos en fProlog a consultas fSQL. Como consecuencia de ello el sistema permite: 1) un completo tra-

tamiento de valores desconocidos, indefinidos y nulos en los predicados de consulta; 2) el establecimiento de una ponderación o grado de certeza de las cláusulas fijado por medio de etiquetas lingüísticas, y 3) el manejo de operadores de comparación más flexibles o borrosos, que podemos interpretar como `aproximadamente mayor', `mucho menor', `casi igual', etc.

### 7.3. Extensiones lingüísticas

El empleo de modificadores y cuantificadores lingüísticos suele venir fijado por medio de una lógica lingüística basada en subconjuntos borrosos. Al igual que en el caso anterior, algunas otras soluciones al problema de la consulta imprecisa se basan en el lenguaje PRUF para la representación semántica de proposiciones borrosas del lenguaje natural. Este lenguaje fue introducido por ZADEH en 1978 y está basado en el cálculo de restricciones borrosas [54], un método desarrollado tras la introducción de su teoría de la posibilidad para ofrecer básicamente un lenguaje de manejo de relaciones borrosas. Este lenguaje difícilmente puede asimilarse a un álgebra relacional borrosa ya que su propósito es más amplio al ofrecer, en palabras de ZADEH, ``un sistema para la caracterización del significado de una proposición por un procedimiento que actúa sobre una colección de relaciones borrosas en una base de datos y devuelve una distribución de posibilidad".

De hecho ZADEH en dicho lenguaje distingue cinco tipos de proposiciones borrosas. Las primeras, denominadas *simples* o tipo I, son aserciones tales como

$$X \text{ es } f$$

donde $X$ representa el sujeto y $f$ una adjetivación vaga o de medida imprecisa tal como `viejo' o `grande', pero no `treintañero'.

Las proposiciones tipo II se conocen como *modificadas.* En ellas se introducen modificadores de la adjetivación tales como `bastante' o `muy', dando un formato

$$X \text{ es } mf$$

donde $m$ representa el modificador y $f$ el adjetivo.

El tercer grupo, tipo III, lo forman las proposiciones *compuestas* tales como

$X_1$ es $f_1$ y $X_2$ es $f_2$ o bien
$X_1$ es $f_1$ y $X_2$ es $f_2$, incluso
si $X_1$ es $f_1$ entonces $X_2$ es $f_2$

Un cuarto grupo, formado por las proposiciones *cuantificadas,* quedaría tipificado como

$$qX \text{ son } f$$

donde $q$ representa un cuantificador generalizado, e incluye aserciones tales como 'La mayoría de los rusos son calvos' o `Casi todos los topos son ciegos'. En ambos ejemplos se muestran casos de cuantificaciones que no corresponden a los convencionales cuantificadores existencial (`existe') y general (`para todo').

Finalmente, las proposiciones pueden ser también *calificadas.* Formaremos este tipo V con aserciones como `Es seguro que Juan está sordo' o `Se cree que las noches son oscuras', que responden al formato

$$\text{Es } rg \text{ que } S$$

donde *S* representa una proposición borrosa de los tipos I, II o III, *r* es un calificador que aprecia la verdad, la probabibilidad o la posibilidad de la aserción y *g* un adverbio que la matiza.

Basados en esta clasificación LI y LIU avanzan con su fSQL y su fProlog en la creación de consultas sobre proposiciones de los tipos I y II. Con este fin fijan las etiquetas o valores lingüísticos que describen la posibilidad (grado de certeza o grado de inferencia) de las aserciones y reglas. En concreto

{claro, MUY posible, posible, BASTANTE posible,...}

Cada uno de los adjetivos `claro', `posible' se asocia con una distribución de posibilidad representada por un subconjunto borroso de [0,1]. Por ejemplo para `posible' se propone

$$posible(x) = \begin{cases} 0 & si \quad 0 \leq x \leq 0.5 \\ 2(\frac{x-0.5}{0.2})^2 & si \quad 0.5 < x \leq 0.6 \\ 1 - 2(\frac{x-0.7}{0.2})^2 & si \quad 0.6 < x \leq 0.8 \\ 2(\frac{x-0.9}{0.2})^2 & si \quad 0.8 < x \leq 0.9 \\ 0 & si \quad 0.9 < x \leq 1 \end{cases}$$

Los modificadores a su vez también vienen representados por medio de subconjuntos borrosos. Es común, siguiendo la pauta presentada en PRUF, hacer

$$(MUY\ p)(x) = (p(x))^2$$

Sobre estas bases se puede llegar a introducir reglas tales como

accidente(Ind, Carret) :-[MUY posible]- bebedor(Ind), peligro(Carret).

que se apoyan en una base que describe la propensión a la bebida de distintos conductores y el grado de peligro de las distintas carreteras. La combinación de los distintos elementos en las consultas permite conocer por ejemplo a partir de un valor umbral cuales son las carreteras peligrosas para los distintos individuos.

En un trabajo reciente TAKAHASHI [45] propone agotar el programa de representación de proposiciones reflejado en la clasificación de ZADEH. El resultado es la construcción de un lenguaje con fórmulas basadas en proposiciones borrosas (FQL) que sirve de fundamento a un lenguaje de consulta borroso sobre bases relacionales.

A diferencia de todos los sistemas de consulta anteriormente tratados, en los que se propone la borrosificación del álgebra relacional introducida por CODD, aquí el enlace con las bases de datos se basa en el cálculo de dominios (cf. [47]) que el propio Codd propone también como lenguaje de consulta. A tal fin se describen las fórmulas y la semántica que regirán en FQL. En dicho lenguaje las fórmulas atómicas son de tres tipos:

$R(x_1,...,x_n)$, que denota que las variables $x_i$ correspondientes a los distintos dominios forman una tupla en la relación *R*.

$x_i \theta x_j$ que representa a posibles relaciones de comparación borrosa entre variables, tales como `es casi igual a' o `es mucho mayor que'.

$\alpha\theta x_j$ o $x_i\ \theta a$ que representan las mismas comparaciones pero frente a un valor constante nítido o borroso.

Las restantes fórmulas se forman a partir de éstas inductivamente siguiendo los cinco patrones presentados en PRUF. Como interpretación para dichas fórmulas en términos de distribuciones de posibilidad se proponen las descripciones y operaciones en términos de subconjuntos borrosos que allí aparecen, y que ya habían sido utilizadas por LI y LIU, aunque extendidas aquí a los tipos III, IV y algunos casos de V.

Para formular consultas con FQL se presentan dos modalidades. Para la primera consideramos una relación *r* sobre *n* atributos con dominios $D_1,..., D_n$, y donde *x* representaría una tupla genérica $x_1,...,x_n$ de la relación. En este caso la formulación sería

SELECT < algo > FROM *r* WHERE *Form(x)*

donde *algo* representa valores de los dominios y *Form(x)* una fórmula cualquiera del lenguaje que no sea cuantificada, es decir no del tipo IV. Los grados de contraste que derivan de la fórmula se van obteniendo con arreglo a la semántica propuesta en PRUF y desarrollada para FQL.

El segundo tipo se desarrolla para el tratamiento de consultas sobre fórmulas del tipo IV y tiene una sintaxis muy similar. En ambos casos, como el grado de contraste resulta un valor en [0,1], cabría formular consultas que ofrezcan tanto la tupla que alcanza mayor grado en el contraste como la serie de tuplas que superan al contrastar un valor umbral previamente prefijado.

Enfoques similares, es decir con una interpretación lingüística en términos de teoría de la posibilidad, pueden encontrarse en sistemas expertos tales como los presentados por ZEMANKOVA y KANDEL en [58]. En este caso, sin embargo, el análisis es más detallado en el modelo de datos y de consulta borrosos empleado que en el acoplamiento efectivo a una base de datos, sin que llegue a describirse de forma efectiva un lenguaje de consulta en sentido propio.

Algunas de las carencias presentes en dicho trabajo fueron abordadas por ZEMANKOVA en el sistema FILIP [60]. El sistema constituye un modelo integral para el tratamiento inteligente de la información basado en ideas muy semejantes a las expuestas. El sistema no se plantea la adaptación o aproximación a los modelos actuales para bases de datos con lo que sería cuestionable su acoplamiento y aprovechamiento de los datos disponibles en las bases de datos convencionales.

## 8. CONCLUSIONES

Es difícil evaluar el impacto de una teoría hasta que no aparecen las primeras controversias o las primeras aplicaciones. En el caso del tratamiento de la incertidumbre mediante lógica borrosa han llegado antes las segundas que las primeras. Este enfoque, alternativo al tratamiento probabilístico, se ha mostrado de gran utilidad tanto en los sistemas de control automático como en los sistemas de información.

En el ámbito de la información los métodos borrosos permiten abordar viejos problemas de representación, como por ejemplo la incorporación de información desconocida, incompleta o parcialmente cierta. También ofrecen la posibilidad de generar instrumentos de consulta más flexibles, capaces de dar índices de fiabilidad de respuesta. Además, es posible con ellos reconsiderar las reglas de inferencia convencionales, ya que no en vano tienen como base una lógica, no convencional sino polivalente, pero que constituye como tal una extensión de la lógica binaria o booleana.

Sin embargo, las ventajas del marco de representación obtenido en los sistemas de información por una generalización del modelo relacional de bases de datos se ven veladas por una mayor complejidad en los procedimientos básicos y por la aparición de nuevos problemas. Nuestro trabajo muestra al hilo de la breve historia de estos métodos en qué modo quedan relacionados los problemas de la representación de la vaguedad y la redundancia de datos. La redundancia o sobreinformación resulta de la incapacidad para distinguir entre datos o informaciones a los que su ambigüedad convierte en similares. La necesidad de filtrar la información obliga a introducir métricas o criterios de similaridad para aplicarlos a los datos. La representación de éstos mediante subconjuntos borrosos, y en última instancia con funciones de pertenencia, reconduce la cuestión al marco de los espacios métricos.

Otro de los problemas aquí abordado ha sido el estudio de sistemas de consulta o contraste, así como la capacidad de inferencia en los modelos de datos anteriormente citados. Hemos razonado sobre las dificultades para engranar métodos borrosos en sistemas convencionales. Aun partiendo de esquemas con una base muy estructurada como el modelo relacional, cuya transición a modelos clausales como DATALOG resulta además bastante natural, nos volvemos a encontrar, en el curso de su adaptación para la información incierta, con los problemas iniciales de representación. Algunos de ellos se han pretendido resolver ateniéndose en la formulación de consultas a la tipificación y semántica que para las proposiciones borrosas propuso ZADEH en su lenguaje PRUF.

No parece que este tipo de soluciones puedan darse como definitivas. El propósito de ZADEH parece consistir en dar una orientación funcional a la estructura inherente en el lenguaje natural para de este modo poder hacerlo operativo en los sistemas automatizados. Pero está por verse si la propia estructura del lenguaje puede quedar englobada en dicho marco funcional, sin contar con el factor temporal y con su capacidad evolutiva. Parece, en opinión de los lingüistas, que el elemento activo en la regeneración de la lengua, y la base de los procesos creativos en ella, tiende a crear nuevas improntas de ambigüedad.

## 9. REFERENCIAS

[1] ANVARI M., ROSE G.F., Fuzzy relational databases, en *Analysis of Fuzzy Information Vol. II* (Bezdek J.C. ed., CRC Press, Boca Raton 1987) 203-212.

[2] BALDWIN J.F., A fuzzy relational inference language for expert systems, en *Procs. 13th. IEEE Int. Symp. on Multiple-valued logic* (Kyoto, 1983) 416-423.

[3] BOSC P., GALIBOURG M., HAMON G., Fuzzy quering with SQL: Extensions and implementation aspects, *Fuzzy Sets and Systems* 28 (1988) 333-349.

[4] BOSC P., PIVERT O., Fuzzy quering in conventional databases, en *Fuzzy Logic for Management of Uncertainty* (Zadeh L.A., Kacprzyk J. eds., J. Wiley, N. York 1992) 645-671.

[5] BOSC P., LIETARD L., PIVERT O., Quantified statements and database fuzzy querying, en *Fuzziness in Database Management Systems*, (Bosc P., Kacprzyk J. eds., Physica Verlag, Heidelberg, 1995) 275-308.

[6] BOSC P., DUBOIS D., PIVERT O., PRADE H., Flexible queries in relational databases - The example of the division operator, *Theoretical Computer Science*, 171 (1997) 281-302.

[7] BUCKLES B.P., PETRY F.E., A fuzzy representation of data for relational databases, *Fuzzy Sets and Systems* 7 (1982) 213-226.

[8] BUCKLES B.P., PETRY F.E., Fuzzy databases and their applications, en *Fuzzy Information and Decision Processes* (Gupta M.M.- Sanchez E. eds., North-Holland, Amsterdam 1982) 361-371.

[9] BUCKLES B.P., PETRY F.E., Extending the fuzzy database with fuzzy numbers, *Information Sciences* 34 (1984) 145-155.

[10] BUCKLES B.P., PETRY F.E., SACHAR H.S., Design of similarity-based relational databases, en *Fuzzy Logic in Knowledge Engineering* (Prade H., Negoita C. eds., Verlag TUV Rheinland, Köln 1986) 3-17.

[11] BUCKLES B.P., PETRY F.E., Generalized database and information systems, en *Analysis of Fuzzy Information Vol. II* (Bezdek J.C. ed., CRC Press, Boca Raton 1987) 177-201.

[12] CERI S., GOTTLOB G., TANCA L., *Logic Programming and Databases*, Springer Verlag, Berlin, 1990.

[13] CODD E.F., A relational model of data for large shared data banks, *Communications of ACM* 13 (1970) 377-387.

[14] CODD E.F., Relational completeness of data base, en *Data Base Systems* (Courant Computer Science Symposia Series, Prentice-Hall, 1972).

[15] CHANG C.L., Interpretation and execution of fuzzy programs, en *Fuzzy Sets and their Applications to Cognitive and Decision Processes* (Zadeh et al. ed. Academic Press, New York 1975) 191-218.

[16] CHEN G., VANDELBULCKE J., KERRE E.E., A step towards the theory of fuzzy relational database design, en *Proceedings of IFSA 91* (Brussels 1991) 44-47.

[17] CHEN G., VANDELBULCKE J., KERRE E.E., A general treatment of data redundancy in a fuzzy relational data model, *Journal of Am. Soc. for Inf. Science* 43 (1992) 304-311.

[18] CHEN G., KERRE E.E., VANDELBULCKE J., Fuzzy functional dependency and its axiomatic system in a fuzzy relational data model, en *Proceedings of IPMU 92* (Mallorca 1992) 313-316.

[19] DATE C.J., *An Introduction to Database Systems, Vol. I* Addison-Wesley, Reading Mass, 1990.

[20] DUBOIS D., PRADE H., *Fuzzy Sets and Systems: Theory and Applications,* Academic Press, Orlando, 1980.

[21] DUBOIS D., PRADE H., *Possibility Theory*, Plenum Press, New York, 1988.

[22] GRANT J., Null values in a relational data base, *Inf. Process. Lett.*, 6 (1979) 156-157.

[23] GONZÁLEZ L., MARÍN A., Fuzzy Correspondence and Extension of Relations, *Australian Journ. of Intell. Inform. and Process. Systems*, 4 (1995) 46-55.

[24] GONZÁLEZ L., MARÍN A., Extending fuzzy relations, *Fuzzy Sets and Systems* 69 (1995) 157-169.

[25] GONZÁLEZ L., MARÍN A., Pseudoigualdades de segundo orden, *Formula,* 4 (1997) 183-197.

[26] KAUFMANN A., *Introduction to the theory of fuzzy subsets* Academic P., New York, 1975.

[27] KERRE E.E., CHEN G., An overview of fuzzy data models, en *Fuzziness in Database Management Systems*, (Bosc P., Kacprzyk J. eds., Physica Verlag, Heidelberg, 1995) 23-41.

[28] KLIR G.J., FOLGER T.A., *Fuzzy Sets, Uncertainty and Information* Prentice-Hall, Englewood Cliffs N.J., 1988.

[29] KANAI N., ISHIZUKA M., Prolog-ELF incorporating fuzzy logic, *Trans. of Information Process. Society of Japan*,27-4 (1986) 4.

[30] KUNII T.L., DATAPLAN: An interface generator for database semantics, *Information Sciences* 10 (1976) 279-298.

[31] LI D., LIU D., *A Fuzzy Prolog Database System*, J. Wiley, New York, 1990.

[32] LIPSKI W., On databases with incomplete information, *J. of the As. of Comp. Mach.* 28 (1981) 41-70.

[33] MARTIN T.P., BALDWIN J.F., PILSWORTHB.W, The implementation of fPROLOG - a fuzzy PROLOG interpreter, *Fuzzy Sets and Systems* 23 (1987) 119-129.

[34] MEDINA J.M., VILA A., GEFRED. Un modelo generalizado de Bases de Datos Relacionales Difusas, en *Actas II Cong. esp. Tecn. y Log. Fuzzy* (Boadilla del Monte, Noviembre 1992) 233-252.

[35] MUKAIDONO M., Fundamentals of Fuzzy Prolog, *Intern. Journ. of Approximate Reasoning*, 3 (1989) 179-193.

[36] NOVAK V., *Fuzzy Sets and their Applications*, Adam Hilger, Bristol, 1986.

[37] OVCHINNIKOV S., Similarity relations, fuzzy partitions, and fuzzy orderings *Fuzzy Sets and Systems* 40 (1991) 107-126.

[38] PRADE H., TESTEMALE C., Generalizing Database Relational Algebra for Treatment of Incomplete or Uncertain Information and Vague Queries, *Information Sciences* 34 (1984) 115-143.

[39] PRADE H., TESTEMALE C., Fuzzy relational databases: Representational issues and reduction using similarity measures, *Journal of Amer. Soc. Information Sciences* 38 (1987) 118-126.

[40] RUNDENSTEINER E.A., HAWKES L.W., BANDLER W., On Nearness Measures in Fuzzy Relational Data Models, *Int. Journ. Approximate Reasoning* 3 (1989) 267-298.

[41] SHENOI S. MELTON A., Proximity relations in the fuzzy relational database model, *Fuzzy Sets and Systems* 31 (1989) 285-296.

[42] SHENOI S. MELTON A., FAN L.T., An equivalence classes model of fuzzy relational databases, *Fuzzy Sets and Systems* 38 (1990) 153-170.

[43] SHENOI S. MELTON A., An extended version of the fuzzy relational database model, *Information Sciences* 52 (1990) 35-52.

[44] TAHANI V., A conceptual framework for fuzzy query processing - a step toward very intelligent database systems *Inform. Process. Management* 13 (1977) 289-303.

[45] TAKAHASHI Y., A fuzzy query language for relational databases, en *Fuzziness in Database Management Systems*, (Bosc P., Kacprzyk J. eds., Physica Verlag, Heidelberg, 1995) 365-384.

[46] THALHEIM B., Fundamentals of database and knowledge-based systems, en *Enc. of Comp. Sci. and Technology, Vol.29*, M. Dekker, 1994.

[47] ULLMAN J. D., *Principles of Database and Knowledge-base Systems*, Computer Science Press, New York, 1989.

[48] UMANO M., FREEDOM-0: a fuzzy databse system, en *Fuzzy Information and Decision Processes* (Gupta M.M.- Sanchez E. eds., North-Holland, Amsterdam 1982) 339-349.

[49] UMANO M., Establishing a Prolog by the concept of Fuzzy Set, en *31th National Congress of the Information Processing Society of Japan*, 1985, 9M-4.

[50] UMANO M., FUKAMI S., Perspectives of Fuzzy Databases, *Japanese Journal of Fuzzy Theory and Systems* 3 (1991) 75-91.

[51] YAGER R.R., On ordered weighted averaging aggregation operators in multicriteria decision-making, *IEEE Transactions on Systems, Man and Cybernetics* 18 (1988) 183-190.

[52] ZADEH L.A., Fuzzy Sets, *Information and Control* 8 (1965) 338-353.

[53] ZADEH L.A., Similarity relations and fuzzy orderings, *Information Sciences* 3 (1971) 177-200.

[54] ZADEH L.A., Calculus of fuzzy restrictions, en *Fuzzy Sets and Their Applications to Cognitive and Decision Processes* (Zadeh et al. ed. Academic Press, New York 1975) 1-39..

[55] ZADEH L.A., The concept of a linguistic variable and its application to approximate reasoning Part I, *Information Sciences* 8 (1975) 199-249.

[56] ZADEH L.A., Fuzzy Sets as a basis for a theory of possibility, *Fuzzy Sets and Systems* 1 (1978) 3-28.

[57] ZADEH L.A., PRUF: A meaning representation language for natural languages, *Int. Journ. Man-Mach. Studies* 10 (1978) 395-460.

[58] ZEMANKOVA M., KANDEL A., *Fuzzy relational databases - a key to expert systems*, Verlag TUV Rheinland, Koln, 1984.

[59] ZEMANKOVA M., Implementing aspects of human reasoning in intelligent information system design, en *Fuzzy Logic in Knowledge Engineering* (Prade H., Negoita C. eds., Verlag TUV Rheinland, Köln 1986) 73-92.

[60] ZEMANKOVA M., Filip: A fuzzy intelligent information system with learning capabilities, *Information Systems* 14 (1989) 473-486.