Capítulo 4

Propuesta de Modelo Conceptual: FuzzyEER

"Nunca he podido contentarme con la lógica de lo blanco y de lo negro,
con la lógica de dos únicos valores contrapuestos. Me parece insuficiente.
Si una cosa no es negra, evidentemente puede ser blanca,
pero igualmente puede ser de un montón de colores"

Boris Vian (escritor y músico francés. 1920-1959)

Dentro de la problemática del diseño de bases de datos, los modelos de datos cumplen un importante rol, pues son las herramientas de abstracción que permiten representar la realidad, captando la semántica y restricciones de un sistema de información. Como se ha explicado anteriormente, esta tesis muestra un estudio en profundidad de la semántica que pueden representar: Aquellos atributos que muestren imprecisión, diferentes grados difusos que pueden representar los atributos, entidades y restricciones difusas, así como también, algunas restricciones difusas, agregación difusa, especializaciones difusas. En todas ellas se destaca el papel que desempeñan los modelos conceptuales de datos en el diseño de una base de datos.

Cada modelo conceptual de datos tiene semántica y notación, los que pueden adoptar varios formatos que incluyen textos y gráficos. Dicho modelo debe ser fácil de usar para propósitos determinados de un sistema de información que se desea especificar. Rumbaugh et al. (1999) dicen que un propósito fundamental de los modelos es que permiten "*captar y enumerar exhaustivamente los requisitos y el dominio de conocimiento, de forma que todos los implicados puedan entenderlos y estar de acuerdo con ellos"*. Por otro lado, Elmasri y Navathe (2000) dicen que "*un esquema conceptual es una descripción concisa de los requisitos de información de los usuarios, y que contienen descripción detallada de los tipos de entidades, vínculos (interrelaciones) y restricciones; éstas se expresan mediante los conceptos del modelo de datos de alto nivel*". Es por ello, que este trabajo se centra en la captura de aquellos datos que posean dominios imprecisos que actualmente no son representados por los modelos existentes. Sin embargo, éstos permiten la transformación o extensión para este objetivo con el uso de la teoría de conjuntos difusos. A su vez, usando cuantificadores difusos se logra flexibilizar algunas restricciones en el modelo propuesto.

En este capítulo se muestra una extensión de la definición y notación gráfica de componentes básicos del modelado, ER (Modelo Entidad Relación) y EER (Modelo Entidad Relación Extendido), generalmente llamado ER/EER. Para facilitar la incorporación de información imprecisa en un modelado de datos conceptual ER/EER, se introducen conceptos de lógica difusa a través de la teoría de conjuntos difusos y se formula un modelo de datos difusos FuzzyEER. Los componentes de modelado de datos que se extienden en este apartado son: Tipos de atributos difusos (tipo de información que se desea modelar), entidades difusas, interrelaciones difusas, participación difusa, completitud difusa, agregación difusa, especialización definida por un atributo difuso, así como también, añadir un grado (con distinto significado posible) a uno o varios atributos o a la entidad completa asociada al grado de toda la instancia de la entidad. Las restricciones difusas de nuestro modelo FuzzyEER son consideradas para las superclases y subclases en el mismo sentido como son utilizadas en el modelo ER/EER (Elmasri y Navathe 2002), sólo que estas restricciones serán de tipo difuso.

4.1 Atributos Difusos en el Modelo FuzzyEER

En el capítulo 2, se definieron varias formas de permitir la información imprecisa en entidades o interrelaciones difusas, como lo son: Valores difusos en los atributos, Grado en cada valor de un atributo, Grado en toda la instancia de la relación, Grado en un conjunto de valores de diversos atributos. A continuación para cada uno de estos casos se propone su representación gráfica y ejemplos en el modelo conceptual FuzzyEER.

4.1.1 Valores Difusos en los Atributos

Los valores difusos en los atributos hacen referencia a los distintos tipos de atributos de una entidad, que pueden modelar valores difusos (imprecisos), al igual que en el tercer nivel de Zviele y Chen (1986) pero de una forma más clara, son aquí definidos y clasificados según el tipo de referencial (ordenado o no) que tengan asociado. El tipo de valor para cada atributo puede ser muy variado, algunos de los más conocidos son: valor nulo, valor desconocido, valor indefinido, distribución de posibilidad, etc. (véase Tabla 3.1).

Los atributos clásicos no incluyen atributos cuyo dominio sea impreciso, considerándose que pueden ser tratados, pero con un formato especial, para representar atributos imprecisos (difusos). En el apartado 3.2.4 se clasificaron los *atributos difusos* en cuatro tipos: Tipo 1 para datos precisos que admiten tratamiento difuso, Tipo 2 para datos imprecisos en dominio de referencial ordenado, Tipo 3 para datos imprecisos con dominio de referencial no ordenado asociados a una relación de similitud y Tipo 4 que son semejante al Tipo 3 pero no asocian una relación de similitud.

En el modelo ER/EER los datos toman valores en un dominio específico y cada una de las propiedades o características que tienen un tipo de entidad o interrelación se denominan *atributos*. Así tenemos que, el conjunto difuso de posibles valores que puede tomar una determinada característica o atributo se denominará *dominio difuso*, hay que tener claro, que los atributos toman valores de uno o varios dominios.

La representación gráfica de los atributos se realiza mediante un círculo etiquetado y unido mediante un arco a la entidad a la pertenece (De Miguel et al. 1999). Estos autores proponen notación para varios tipos de atributos, por ejemplo, atributo identificador o clave denotado por un círculo relleno (negro), atributo simple denotado por un círculo normal, atributo compuesto denotado por un arco con un círculo normal que cruza todo los atributos que compone, todos ellos unidos al tipo de entidad o interrelación a la que pertenecen. Otros tipos de atributos se muestra la Figura 4.4.

Definición 4.1: Sea un tipo de entidad *E* con atributos $A_1, A_2, \ldots, A_n$ y sea $D_1, D_2, \ldots, D_m$ conjuntos finitos de los dominios respectivos para cada A_i, con i =1,..., n. Un **atributo difuso** es aquel que tiene un dominio impreciso (difuso) asociado al tipo de entidad *E*. Se representa gráficamente con un círculo normal para T1 y un círculo estrellado para T2, T3 y T4, con una línea que lo une a la entidad. Dicho de otra forma, en el caso de atributos clásicos se utiliza un círculo normal, y aquellos que admiten valores difusos se representan con un *círculo estrellado,* Además, al nombre del atributo se le debe anteponer el tipo de atributo que representa sea este T1, T2, T3 o T4. Opcionalmente, se puede incorporar junto al nombre del atributo las etiquetas lingüísticas que están definidas para ese atributo. En la Figura 4.1 se muestra un esquema para representar los cuatro tipos de atributos difusos que consideramos en esta definición.

Respecto a esos cuatro tipos de atributos difusos que aquí consideramos, hacemos las siguientes aclaraciones:

1. **Tipo 1:** D_j es el dominio y cada valor de D_j es *crisp* o A_j es un atributo clásico. Sin embargo, lo denominamos como atributo difuso porque este tipo de atributo admite que sobre él puedan efectuarse consultas difusas, y como en esas consultas difusas pueden utilizarse etiquetas lingüísticas, valores aproximados, etc. entonces, es necesario que esos conceptos estén definidos previamente.

2. **Tipo 2:** D_j es el dominio de conjuntos difusos definidos por una distribución de posibilidad μ que exprese un concepto difuso X. A cada uno de los valores de D_j, se le asocia un valor en un intervalo [0,1], representando el grado de pertenencia al conjunto, o la posibilidad de que ese valor sea cierto en la instancia al cual está asociado:

$$\mu_X(d), \text{ con } d \in U_j \tag{4.1}$$

   donde U_j es el dominio subyacente de D_j con j=1,...,n, es decir, el conjunto de valores *crisp* de dicho dominio.

3. **Tipo 3:** D_j es el dominio de conjuntos difusos o distribuciones de posibilidad definidas sobre un conjunto de términos, escalares o etiquetas lingüísticas, que no tienen asociada una distribución de posibilidad. Estas etiquetas se definen en un dominio no ordenado y sobre ellos se define una relación de similitud que permite medir el parecido existente entre cada par de etiquetas del dominio. La representación son distribuciones de posibilidad similares a la Ecuación 4.1, pero ahora el dominio U_j es el conjunto de etiquetas lingüísticas definidas, a las cuales la función μ le asocia un valor en el intervalo [0,1].

4. **Tipo 4:** D_j es un dominio similar al anterior pero en el que no existe una relación de similitud entre cada par de etiquetas del dominio.

*

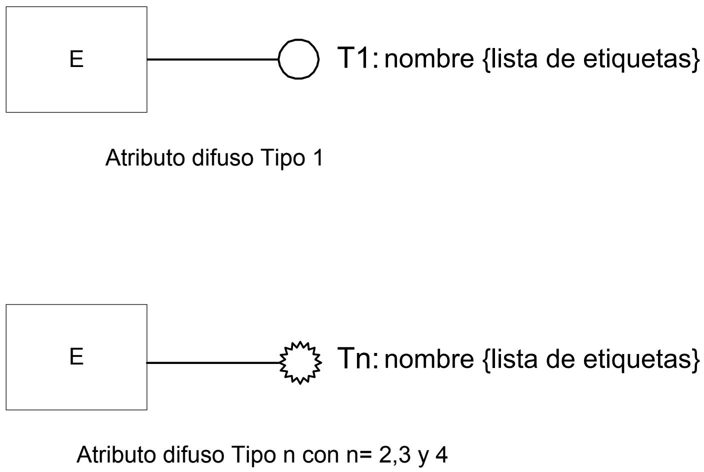


Figura 4.1: Representación gráfica de los tipos de atributos difusos T1, T2, T3 y T4.

.

Otra notación de nuestras primeras publicaciones es la presentada en el apartado 4.6, que permite una línea quebrada y sobre ella el tipo de atributo que representa, pero consideramos que la propuesta en la Figura 4.1 simplifica la notación y no es necesario incorporar la línea quebrada, pues con el círculo estrellado quedan claramente marcados. Se puede considerar también, que en los atributos difusos Tipo 3 y Tipo 4 como el dominio está restringido a las etiquetas o a una distribución de posibilidad sobre esas etiquetas, podemos establecer como obligatorio poner la lista de etiquetas, lo cual aclara el dominio de dicho atributo difuso.

Ejemplo 4.1: En una compañía de teatro se puede considerar una entidad *EMPLEADO* para representar los actores de la compañía. En esta entidad se pueden considerar de forma simplificada los siguientes atributos: {`DNI,Altura,Edad,Color_pelo,Color_piel`}. Algunos de estos atributos pueden estar caracterizados como atributos imprecisos. Veamos en detalle la definición de los atributos:

<u>Atributo Clásico</u> = `DNI`: Su dominio corresponderá a valores numéricos, siendo este atributo definido como clave primaria para la entidad Empleado.

<u>Atributo Tipo 1</u> = `Altura`: Su dominio puede ser definido como el conjunto de valores entre {0,..., 2.50}, identificado como un *atributo preciso*, pero se le pueden asociar etiquetas lingüísticas, como por ejemplo, "bajo", "mediano", "alto" siendo esto documentado en un diccionario de datos. En este caso se supone que si se conoce la altura de un empleado, se dice que, se conoce el valor de ese dato en forma precisa (exacta). En el caso de no conocer ese dato

se puede usar el valor Null en el sentido del SGBD (por ejemplo Oracle). Nótese que este atributo difuso podría ser de Tipo 2.

Atributo Tipo 2 = `Edad`: Su dominio corresponderá a un conjunto difuso sobre las edades consideradas como valores numéricos (referencial ordenado), permitiendo representar las etiquetas lingüísticas {"joven", "maduro", "mayor"}, definidas como *distribuciones de posibilidad*. En la Figura 4.2 pueden verse todas las etiquetas que considera este atributo difuso, los valores que cada una toma, y el grado de pertenencia asociado a cada valor en cada etiqueta. Obsérvese por ejemplo que la edad 26, tiene un grado de pertenencia 0.8 para la etiqueta "joven", que es un trapecio dado por sus cuatro valores característicos {0/15, 1/20, 1/25, 0/30}.

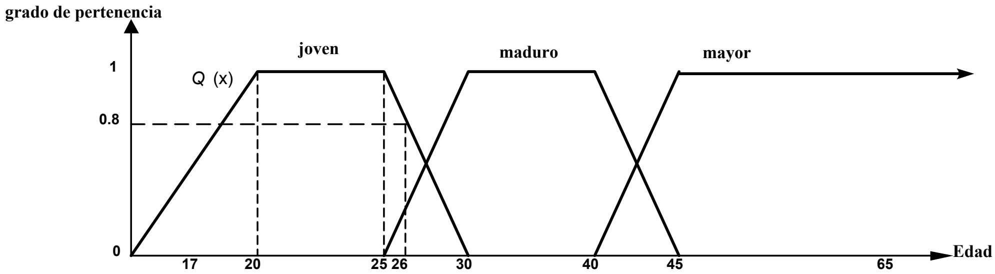


Figura 4.2: Distribuciones de posibilidad para las etiquetas lingüísticas del atributo difuso `T2:Edad`.

Atributo Tipo 3 = `Color_pelo`: Su dominio subyacente corresponderá al conjunto: {"rubio", "moreno", "pelirrojo"}, a cada una de estas etiquetas se debe asociar un valor de semejanza o similitud según corresponda en un intervalo [0,1]. Observe que los valores de ese dominio no están ordenados, es decir, no disponen de una relación de orden. Su *función de similitud* asociada está representada en la Tabla 4.1. El dominio real son valores simples de ese referencial (por ejemplo, 1/Rubio) o una distribución de posibilidad sobre el mismo (por ejemplo, {1/Rubio, 0.6/Pelirrojo}).

	Rubio	Moreno	Pelirrojo
Rubio	1	0.1	0.8
Moreno	0.1	1	0.3
Pelirrojo	0.8	0.3	1

Tabla 4.1: Función de similitud para atributo difuso `T3:Color_pelo`.

Atributo Tipo 4 = `Color_piel`: Su dominio corresponderá a añadir un único grado a una o varias etiquetas del dominio subyacente: {"blanca", "amarilla", "café", "negra"}. Por ejemplo, un valor válido es {0.8/blanca, 0.6/amarilla, 0.4/café, 0.1/negra}. A diferencia del caso anterior, en este atributo difuso no existe una tabla de similitud, sino que grados que se definen por cada atributo difuso en un intervalo [0,1].

La Figura 4.3, presenta la entidad *EMPLEADO* con la notación del modelo FuzzyEER, que define atributos difusos Tipo 1, Tipo 2, Tipo 3 y Tipo 4, correspondiendo al Ejemplo 4.1. Dicha entidad empleado puede ser de mucha utilidad si los empleados pertenecen a una compañía de teatro, estamos hablando de actores, en particular cuando se requiere de algunos rasgos especiales al representar alguna escena donde el actor debe cumplir con algunas características físicas del personaje.

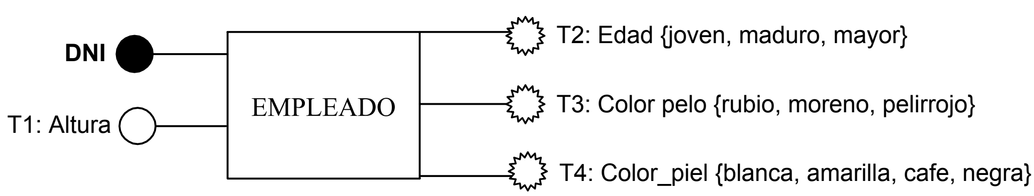


Figura 4.3: Entidad *EMPLEADO* con atributos difusos Tipo 1, Tipo 2, Tipo 3 y Tipo 4.

También se puede utilizar la clasificación de los atributos definidos en el modelo ER/EER, como atributo: Simple, compuesto, derivado, múltiple, etc. (por ejemplo para el atributo difuso Tipo 2, como lo muestra la Figura 4.4). Los atributos difusos que definen estas características, conservan la notación definida en De Miguel et al. (1999), dichos atributos son incorporados a la propuesta del modelo FuzzyEER. Entendiéndose por *atributo difuso derivado* aquellos que se obtienen a partir de otros ya existentes (estos pueden o no ser atributos difusos) y "*d*" es la regla de derivación documentada en el diccionario de datos. Un *atributo difuso simple* es aquel que toma un sólo valor en un dominio difuso en una entidad. Los *atributos difusos múltiples* (o multivaluado) toman más de un valor en un dominio difuso en una entidad definiéndose en su representación el número mínimo y máximo de valores, para el caso opcional el valor mínimo será 0 y este atributo puede no tomar valores y en el caso de obligatorio (mínimo) vale 1 y el atributo debe por lo menos tomar un valor en un dominio difuso. El atributo compuesto es el atributo definido sobre más de un dominio difuso, algunas de estas representaciones se muestran en la Figura 4.4.

Definición 4.2: Los **atributos difusos simple, compuesto, derivado, múltivaluado y compuestos** conservan la misma definición y representación gráfica cuando son difusos, esto es, que sólo hay que distinguir qué tipo de atributo difuso representa (T1, T2, T3, o T4) colocado delante del nombre del atributo y cambiar el círculo a estrellado (excepto para T1).

*

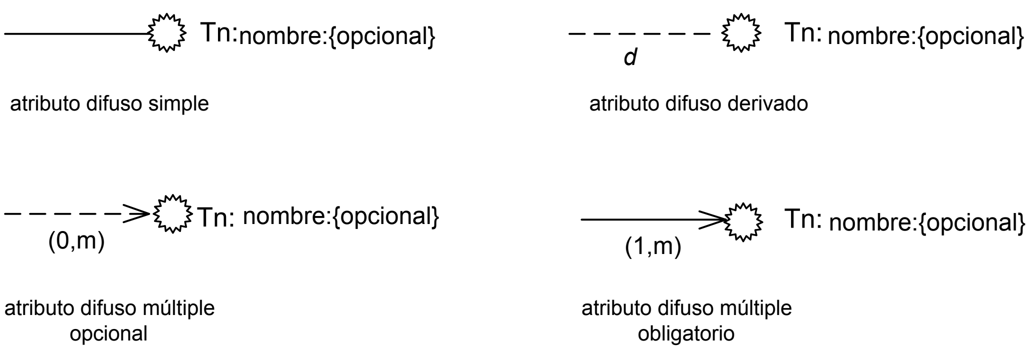


Figura 4.4: Representación gráfica de atributo difuso Tipo n, con n=2, 3 y 4, simple, derivado y múltiple para FuzzyEER (para n=1 es similar usando un círculo normal en vez de estrellado).

Un ejemplo de este tipo de atributo se encuentra en la Figura 4.7, los atributos experiencia y habilidad son atributos difusos simples, en cambio, el atributo difuso G^2 (experiencia, habilidad) es un atributo difuso derivado que representa un grado de incertidumbre.

La representación de los datos que se desea modelar en un sistema de información, considerando que estos tengan datos precisos o imprecisos, fue discutida en este apartado, pero se debe considerar que un grado puede tener distintos significados, a su vez, este tipo de grado requiere modelar uno o varios grados, según sea el significado de este grado en un esquema de datos dado, esto se discutirá en el siguiente apartado.

4.1.2 Grado en Cada Valor de un Atributo

En este caso, el valor de cada atributo puede tener asociado un grado, generalmente en un intervalo [0,1], que mida el *nivel de difuminado* de dicho valor. Los significados de estos grados pueden ser variados, como por ejemplo: de completitud, de incertidumbre, de posibilidad, etc.

Para mostrar la utilidad de los significados de grados en atributos difusos, se hace referencia a trabajos de algunos autores que han utilizado este concepto (Galindo, 1999; Galindo et al., 2001b; Ma et al., 2001; Medina et al., 1994). Cabe destacar que para cada uno de los grados, si toma el valor 1 representa la máxima expresión del grado, y si toma el valor 0 lo contrario.

El dominio de los grados suele estar delimitado al intervalo real [0,1]. También pueden considerarse etiquetas (como “mucho”, “normal”, etc.) definidos como conjuntos difusos sobre el intervalo [0,1], pero esto no aporta grandes ventajas, y sí, más complejidad técnica y semántica, por lo que no consideramos este tipo de grados. Nuestro interés se centra en representar algunos grados difusos definidos en el apartado 3.2.2, en este caso: grado de cumplimiento, grado de incertidumbre, grado de posibilidad, grado de importancia y grado de pertenencia. A continuación para cada uno de ellos se define su correspondiente notación gráfica.

Definición 4.3: Se define **grado difuso asociado a un atributo** con cierto significado. Este significado se expresa con la notación G^{n} o $G^{significado}$, donde ese “n” o su significado es uno de los que se definen a continuación u otros nuevos. Esa expresión se ubica delante del nombre de un atributo que define el grado unido por una flecha al atributo del que se le asocia. Véase Figura 4.5.

Sea una entidad *E* con atributos (A_1, A_2,...,A_n), para cada atributo A_i con $i \in \{1,..., n\}$, puede existir un grado difuso asociado a cada valor que dicho atributo que toma en cada instancia de *E*. El significado de ese grado se define a continuación:

- **Grado de pertenencia**: La pertenencia de un valor a una instancia concreta puede ser cuantificada por un grado. Se representa con G^{0} o $G^{Pertenencia}$.

- **Grado de cumplimiento** (satisfacción): En una instancia, una propiedad puede cumplirse con cierto grado entre dos extremos. Se representa con G^1 o $G^{Cumplimiento}$.

- **Grado de incertidumbre**: El grado de incertidumbre expresa la certeza o seguridad con que conocemos un dato determinado para una instancia concreta. Se representa con G^2 o $G^{Incertidumbre}$.

- **Grado de posibilidad**. Mide la posibilidad de la información que se está modelando para cada instancia de la entidad. Se representa con G^3 o $G^{Posibilidad}$.

- **Grado de importancia**: Distintos valores de un atributo pueden tener diferentes importancias, de forma que existan ciertos valores de cierto atributo más importantes que otros. Se representa con G^4 o $G^{Importancia.}$.

*

Obsérvese que la notación de este tipo de atributo difuso que contiene algún tipo de grado que representa, deberá considerar dos formas:

1. Si tiene una función Q(x) asociada que define el cálculo de un atributo, éste será un grado *difuso derivado*. Con esa función se obtiene el grado respectivo. O sea, la función permite calcular esos grados de forma automática basándose en el valor de otros atributos o en cualquier otra información almacenada o disponible en la base de datos. Esa función puede ser, por ejemplo, una distribución de posibilidad de una etiqueta lingüística.
2. Si no tiene definida una función es un *atributo difuso directo* (de derivado) que sólo debe llevar el nombre del atributo que contiene el grado difuso.

Aunque estos dos casos son producto de un cálculo para obtener el respectivo valor del atributo, el caso 1 es más complejo que el caso 2, ya que este último trata de un valor que se calcula de forma no automática. Por ejemplo, introducida por el usuario de la base de datos.

La Figura 4.5, muestra un ejemplo generalizado para representar los dos casos anteriormente mencionados con una notación gráfica en FuzzyEER de grado en un atributo, nótese que en la Figura 4.5 a) el atributo superior es un atributo simple no difuso pero derivado y en la Figura 4.5 b) el atributo superior es un atributo difuso no derivado. Para ambos se puede definir un tipo de grado difuso G^n, como lo muestra el atributo inferior.

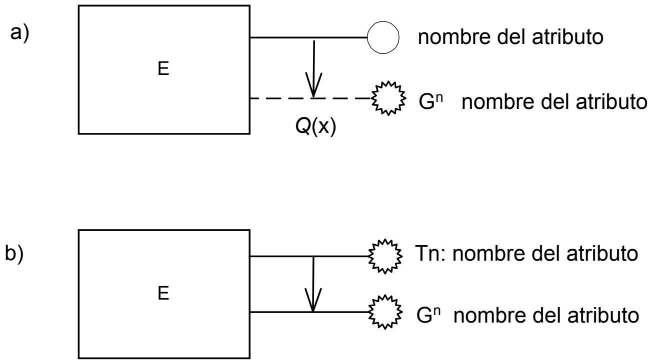


Figura 4.5: Representación gráfica del significado de los grados, *n* pertenece a {0, 1, 2, 3, 4}. a) Un atributo derivado, b) Un atributo no derivado.

Considérese, que el atributo difuso generalmente es diferente al grado que representa. Por ejemplo, decir que el valor del color del pelo tiene una importancia (o un grado de cumplimiento de cierta propiedad), pero este grado no es atributo Tipo 3, si no que representa un grado de importancia. Por tanto, cada grado debe estar asociado a un atributo cuyo nombre se sitúa tras G^n.

Ejemplo 4.2: Para un atributo difuso `T2:Edad` con etiquetas de la Figura 4.2 {joven, maduro, mayor}, definidas en un referencial ordenado perteneciente a una entidad *PERSONA*. Se puede definir un conjunto difuso considerando la etiqueta lingüística "joven" con una distribución de posibilidad dada por un trapecio con: $\{\alpha, \beta, \delta, \gamma\}$ correspondiente a: {0, 20, 25, 30}. Algunos de los significados de grados de atributos difusos que pueden definirse en este ejemplo son:

a) Persona con x edad pertenece al conjunto de personas jóvenes con Q(x), donde la función Q, es la función de pertenencia G^0 del conjunto difuso "joven".
b) Persona con posibilidad G^3 de que tenga entre 21 y 24 de edad es Q(x).
c) Para cada persona, la edad puede ser más o menos importante G^4 según su cargo, por ejemplo.

Para éste caso, el atributo difuso `T2:Edad` tiene una función de distribución asociada Q(x), por tanto es un atributo difuso derivado que representa algún grado. El esquema de la entidad *PERSONA* asociado al caso a) y b) están representado en la Figura 4.6, respectivamente.

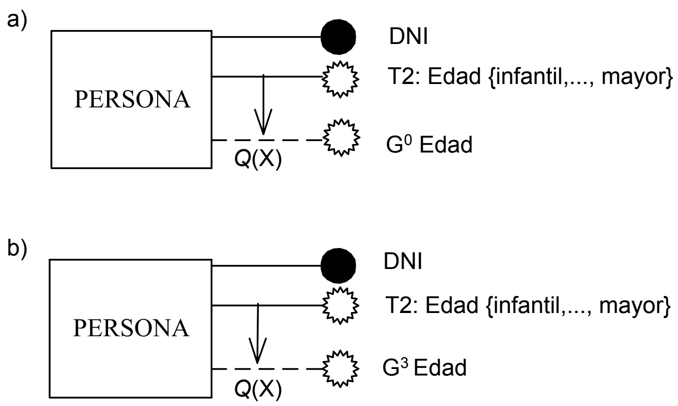


Figura 4.6: Representación en modelo FuzzyEER de significado de grado G^0 y G^3 para el atributo difuso `T2:Edad`.

En caso de que no sea definido el significado como 0, 1, 2, 3, o 4, el atributo representará por omisión el *grado de pertenencia* G^0.

4.1.3 Grado Asociado a los Valores de Diversos Atributos

Para este caso, se crea un atributo cuyo valor es un grado. Este grado está asociado a un conjunto de atributos pertenecientes a la entidad, siendo un caso poco usual pero puede dar expresividad si se incorpora a un modelo de datos, un ejemplo de este atributo se encuentra en el concepto de dependencia difusa planteada por Raju y Majumdar (1988).

Definición 4.4: Sea una entidad *E* con atributos (A_1, A_2,...,A_n), para un subconjunto *X* de atributos A_i con $i \in I$, siendo I un conjunto de índices tal que $I \subset \{1,..., n\}$. Ahora, puede existir un grado difuso asociado para cada valor de los atributos de *X* que toman en cada instancia de *E*. A este grado se le denomina **Grado asociado a los valores de diversos atributos**, y puede tener diversos significados como se expresó en la Definición 4.3. En lo referente a la notación, en un modelo FuzzyEER este tipo de atributo es representado como un atributo difuso (derivado o no), unido por una línea que atraviesa los atributos de *X* (los que están asociados a este grado) y junto a la línea, una punta dirigida al nuevo atributo difuso, tal y como se muestra la Figura 4.7. En general el grado del atributo será derivado de esos atributos.

*

Ejemplo 4.3: Supongamos los atributos {`DNI,Oficio,Experiencia,Habilidad`} de una entidad *EMPLEADOS*. Para este caso se considera, que el dominio del atributo `Oficio` es clásico u ordinario (alfabético), y los dominios de los atributos `Experiencia` y `Habilidad` son conjuntos difusos con etiquetas {aprendiz, normal, experto} para experiencia y las etiquetas {torpe, normal, hábil} para la habilidad, a modo de ejemplo en ambos casos.

A partir de estos dos atributos podemos añadir un grado al grupo de atributos {experiencia, habilidad}. El significado de este "grado de experticia" que se obtiene del conjunto {experiencia, habilidad} puede variar dependiendo del contexto. Por ejemplo, se puede dar:

- Grado de incertidumbre G^2: Representa en qué medida los valores de los atributos {`Experiencia,Habilidad`} son ciertos, dependiendo de cómo se hayan calculado esos dos valores (si se ha basado en muchas pruebas o en pocas). Véase G^2 de la Figura 4.7, donde "*d*" representa la fórmula de cálculo de derivación del atributo difuso.

- Grado de importancia G^4: Para un empleado esos atributos pueden ser más importantes que otros. Por ejemplo en el ámbito hospitalario, un cirujano requiere más experiencia y habilidad que un auxiliar administrativo.

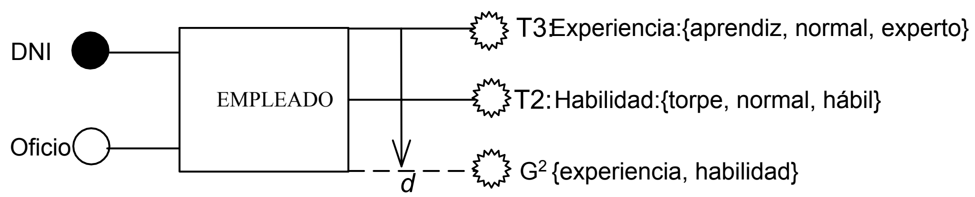


Figura 4.7: Gráfica de atributos para grado de conjunto de valores con grado G^2.

Otro ejemplo del uso de este tipo de caso, es que de dicha entidad se puede obtener un grado de "interés por la perfección".

4.1.4 Grado Difuso con Significado Propio

Otra forma de analizar los grados es considerar el caso que un determinado grado no tiene porqué estar asociado a otro atributo, sino que puede ser difuso y tener valor (o significado) por sí mismo, en otras palabras, se tienen dos formas de grados en atributos difusos: grado asociado a un atributo (que mide algo difuso de cierto atributo) o grado sin asociar a ningún atributo (que tiene información difusa con sentido por sí misma).

Por ejemplo, en una entidad *MEDICAMENTO*, puede tener diversos atributos (`Nombre, Tipo,Compuesto_1,Compuesto_2,Peligrosidad,Color`). Ahora, analicemos los atributos `color` y `peligrosidad`, considerados como grado de atributos difusos que toma valores en un intervalo [0.1] con las siguientes características:

- `Color` puede ser un atributo difuso con un grado asociado a cada valor que mide en qué medida ese color es "intenso".
- `Peligrosidad` medirá cuánto de peligroso es dicha sustancia.

O sea que, un `Grado_color` es un grado difuso que mide alguna característica difusa de un atributo, mientras que la `Peligrosidad` es un grado difuso que mide alguna característica difusa de la instancia. Nótese qué ambos atributos difusos no representan el mismo concepto de grado. Tal vez dependiendo del caso habría que especificar una notación distinta a la presentada en las definiciones 4.3 y 4.4 para el modelo FuzzyEER.

4.2 Entidades e Interrelaciones Difusas en FuzzyEER

En el apartado 3.2.2 se definen ciertos tipos de datos difusos que serán utilizados aquí para formalizar algunos conceptos, notaciones y ejemplos de una entidad difusa, ya sea ésta: regular (tratadas como grado en toda la instancia de la relación) o débil (dependencia de existencia o dependencia de identificación). Para el caso de las interrelaciones difusas se definen si tienen un atributo difuso que vincula las entidades que asocia.

4.2.1 Grado de Toda la Instancia de la Entidad (Entidad Difusa)

Como sabemos, una *entidad* es un objeto real o abstracto acerca del cual se quiere almacenar información en la base de datos, y se denomina a la estructura genérica *tipo de entidad*. Por otra parte, una *instancia (ocurrencia) de la entidad* es cada una de las realizaciones concretas (objetos o valores) de ese tipo de entidad. La representación gráfica de los tipos de entidades es un rectángulo etiquetado con el nombre del tipo de entidad que representa (Elmasri y Navathe, 2002; De Miguel et al., 1999).

En el modelo FuzzyEER una entidad difusa consiste en generar *tipos de entidades difusas*, de forma que cada instancia del tipo de entidad tenga un grado para medir la relación de esa instancia con su tipo de entidad (su grado de pertenencia a ese tipo de entidad, o bien, su grado de importancia dentro de ese tipo, su grado de certeza, etc.). Al igual que el apartado anterior una función $Q(x)$ que calcule los grados es opcional, si el usuario no introduce el valor de los grados.

Definición 4.5: Se define **tipo de entidad difusa** en un modelo FuzzyEER, como una entidad a la que se le añade un atributo que expresa un grado (con cualquier significado) definida como:

Sea E una entidad, con n instancias, $e_1, e_2, \ldots, e_n$ se define E como entidad difusa si $\exists$ una función μ_E definida sobre dichas instancias tal que $\forall\ e_i \in E$,

$$\mu_E\,(e_i) \in [0,1] \quad \text{con } i = 1,2,\ldots,n \tag{4.2}$$

La expresión $\mu_E\,(e_i)$ puede medir el grado con que la instancia e_i "*pertenece*" a E, aunque puede tener otros significados, como los expresados en la definición 4.3. La notación de un tipo de entidad difusa es representada con un rectángulo de líneas discontinuas. También se debe añadir el atributo difuso, con el significado del grado que represente, señalado con un círculo con línea discontinua. Véase Figura 4.8.

*

Esto es similar a los casos anteriores, pero aquí el grado está asociado a toda la instancia de cierto tipo de entidad y no exclusivamente a un valor particular de un atributo de una instancia. Por lo general, indicando el grado de pertenencia G^0, u otro tipo de grado que se desea representar. Esta definición consiste en generar *tipos de entidades difusas*, de forma que cada instancia del tipo de entidad tenga un grado para medir la relación de esa instancia con su tipo de entidad (su grado de pertenencia a ese tipo de entidad, o su grado de importancia dentro de ese tipo, su grado de certeza, etc.).

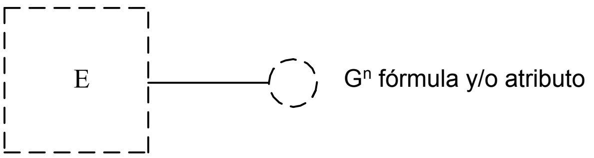


Figura 4.8: Gráfica de entidad difusa con grado en toda la instancia de la entidad.

Ejemplo 4.4: Se puede considerar una entidad difusa *EMPLEADO*, para la cual se define el correspondiente total de horas trabajadas que asignaría cierto *grado de pertenencia* a la entidad para cada empleado. Los empleados, que trabajan una cierta cantidad de horas distinta para cada uno, pertenecerán al tipo de entidad *EMPLEADO* con un cierto grado. Ese grado será calculado dividiendo el número total de horas trabajadas por el número mínimo de horas para que la pertenencia sea total. A partir de ese número mínimo de horas el grado de pertenencia será siempre 1. Se genera así, un atributo difuso derivado para obtener el grado de pertenencia. La Figura 4.9 representa un esquema que modela este ejemplo, donde *Q*(x) es el cálculo del grado, definido por: min{1, nº horas trabajadas/ nº mínimo de horas para la pertenencia total}.

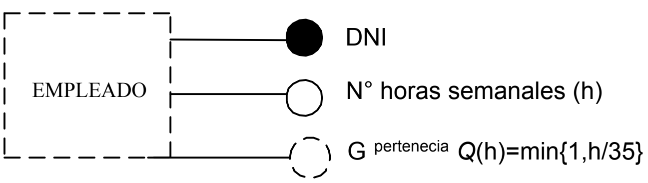


Figura 4.9: Entidad difusa con grado de pertenencia de número de horas trabajadas.

Así se tiene que, si "h" es el número de horas semanales trabajadas por cierto empleado, y 35 es el número mínimo de horas para ser considerado empleado totalmente (con grado 1), entonces se obtiene que el grado de pertenencia para dicho empleado viene dado por la expresión Q(h)=mín{1, h/35}. Así, un empleado que trabaje en la empresa 15 horas, será considerado como un empleado con grado 0.43, de forma que este grado pueda tenerse en cuenta en diversos cálculos (selecciones para distintos fines, gratificaciones, etc.).

Este caso representa un tipo de entidad difusa, que es un conjunto difuso y, por tanto, a cada instancia de este tipo de entidad le corresponderá un grado (de pertenencia, de importancia, etc.), La Figura 4.10, muestra la función Q(h) sobre el número de horas trabajadas. A partir de un número de horas mayor o igual a 35, las instancias tendrán una pertenencia total a la entidad empleado (grado 1).

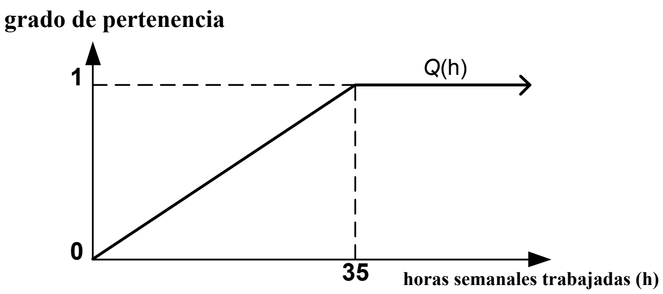


Figura 4.10: Distribución de posibilidad según el nº mínimo de horas trabajadas a la semana de la entidad *EMPLEADO*.

Por ejemplo, si se considera que las instancias son N1, N2, N3, N4,... con grados de pertenencia 0.5, 1, 0.25, 0.75,... respectivamente por cada empleado, entonces, la Figura 4.11 muestra un conjunto difuso sobre el dominio de instancias de la entidad *EMPLEADO*. Observe que se trata de un dominio discreto no ordenado, donde la forma de la función depende de cómo se ordenen los empleados.

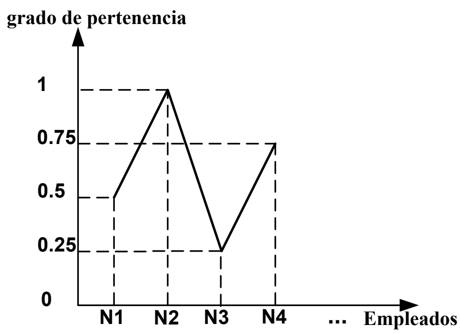


Figura 4.11: Conjunto difuso de las instancias de la entidad *EMPLEADO*.

4.2.2 Entidades Débiles Difusas en FuzzyEER

Existen dos clases de entidades, las *regulares*, que son aquellas que tienen existencia en la base de datos por sí mismas y las *débiles*, cuya existencia depende de una entidad regular por una interrelación de dependencia de identificación o una interrelación de dependencia de existencia. La representación gráfica de los tipos de entidades débiles son dos rectángulos concéntricos con su nombre en el interior conectadas a un tipo de interrelación marcada con una E (Existencia) o una ID (Identificación) definido en De Miguel et al. (1999).

También, en un modelo ER/EER, por lo general, ante la existencia de un atributo multivaluado en una entidad regular, se define un tipo de entidad débil por *dependencia de existencia*, esto quiere decir, que la entidad débil no puede existir si desaparece el tipo de entidad regular a la cual pertenece. Otro tipo es una *dependencia de identificación*, en el que el identificador de la entidad débil incluye el identificador de la entidad regular (Elmari y Navathe, 2002; De Miguel et al., 1999; Atzeni et al., 1999). En una representación del modelo FuzzyEER que considere un tipo de atributo difuso definido sobre etiquetas lingüísticas, se da el caso que este puede tomar más de un valor (atributo difuso multivaluado). Además, para atributos difusos con esta característica, se obtiene más de un grado de pertenencia al conjunto, dependiendo de la etiqueta asociada (véase Figura 4.13).

Definición 4.6: Se define una **entidad débil difusa** en dos casos:

a) Si una entidad débil se define por *dependencia de existencia*. Para este caso, si una entidad regular tiene uno o más atributos multivaluados y alguno es un atributo difuso, entonces en el modelo FuzzyEER se define una *entidad débil difusa,* identificada como E^F. La representación gráfica es denotada por dos rectángulos con una línea discontinua y asociado, por una parte, el atributo difuso multivaluado (que ejerce de clave parcial) y, por otra parte el grado, el cual opcionalmente puede tener algún significado, tal como se muestra en la Figura 4.12a). La Ecuación 4.3 muestra a E^F definiendo sus instancias a partir de las m instancias de la entidad propietaria E. Así, para cada instancia de E:

$$E^F \text{ tiene } n \text{ instancias: } z^F_{ij} = \langle \text{k}, b_j, \mu_{b_j}(\text{k}) \rangle, \quad \text{para } i = 1, 2, ..., m \qquad (4.3)$$

donde k es la clave de la entidad propietaria E, b_j son las n etiquetas con j=1,..., n. En este caso la clave de la entidad débil será la unión de los atributos de la clave de la entidad propietaria más el atributo b_j que señala la etiqueta.

b) Si una entidad débil es establecida por *dependencia de identificación*. La representación gráfica es denotada por dos cuadrados con líneas punteada, junto a esta un atributo con un círculo de líneas cortadas que índica el tipo de grado asociado al atributo, tal como lo muestra la Figura 4.12b). Este caso supone que la clave de la entidad débil será la unión de la clave de la entidad propietaria y una clave extra llamada clave parcial que debe tener la entidad débil. Así, los valores de esa clave parcial pueden repetirse en E^F si pertenecen a distinta instancia de E.

*

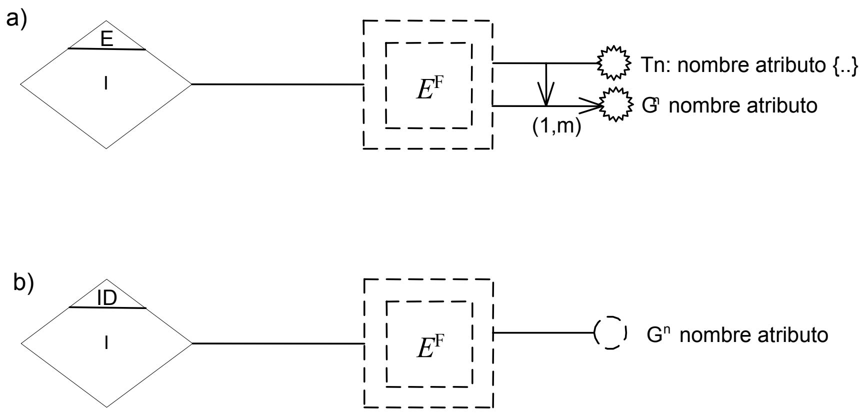


Figura 4.12: Gráfica de entidad débil difusa. a) Por interrelación de dependencia de existencia de atributo multivaluado con *m*>0, b) Por interrelación de dependencia de identificación.

Algunos autores (Elmari y Navathe, 2002) prefieren unir la entidad débil con el rombo de la interrelación *I* con una doble línea, indicando así que todos los elementos de la entidad débil deben estar relacionados con la entidad propietaria. Además, el rombo prefieren ponerlo con doble línea para indicar más claramente cual es la interrelación de la entidad propietaria.

Ejemplo 4.5: Se tiene una entidad *EMPLEADO* con los atributos (`DNI, Nombre_empleado,Año_contrato,Evaluación`). Para este ejemplo, el atributo `Evaluación` es una calificación que obtiene cada empleado por su desempeño durante un período. El atributo difuso Tipo 2 es definido como `T2:Evaluación y` tiene asociadas las etiquetas: {"deficiente", "regular", "bueno", "excelente"}. Desde este punto de vista, "Evaluación" es un atributo que presenta imprecisión, pudiendo en algunas instancias ser tratado como Tipo 1 con las mismas etiquetas definidas en el caso del atributo Tipo 2, y poder ser consultadas difusamente, o bien, puede transformarse a un dominio difuso, en ambos casos el conjunto de "deficiente", "regular", "bueno" y "excelente" son etiquetas lingüísticas de forma trapezoidal, como lo representa la gráfica de la Figura 4.13.

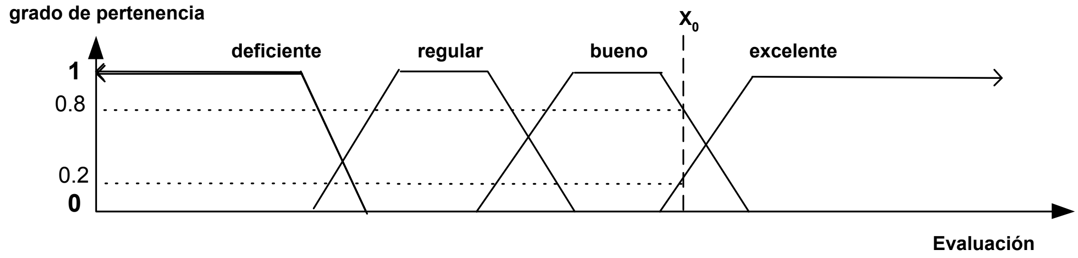


Figura 4.13: Representación de las etiquetas lingüísticas trapezoidales para el atributo difuso `T2:Evaluación` y $\mu_{excelente}(x_0)$=0.2 y $\mu_{bueno}(x_0)$=0.8.

En este ejemplo, un empleado puede tener un grado de pertenencia a más de una etiqueta en el atributo `Evaluación`, de forma tal, que se obtenga el grado de pertenencia más representativo que se ajuste al desempeño del empleado. Por ejemplo, si un empleado es evaluado *bueno*; que tan cerca de *excelente* o que tan cerca de *regular* es la evaluación. Considérese que no es lo mismo un "bueno" con un alto grado de pertenencia a "regular" que un "bueno" con un alto grado de pertenencia a "excelente".

Por otro lado, una forma de representar el diccionario de datos de los correspondientes valores que toman las etiquetas lingüísticas del atributo `Evaluación` se puede ver en la tabla 4.2, junto al cálculo de $\mu(b_j)$. La implementación en FSQL de los valores de una etiqueta lingüística semejante a este diccionario de datos, se encuentra en el apéndice IV Figura IV.1 en las tablas FUZZY_LABEL_DEF y FUZZY_OBJECT_LIST.

Atributo	*j*	Nombre etiqueta	Valor α	Valor β	Valor γ	Valor δ
Evaluación	1	Deficiente	1	2	4	6
	2	Regular	4	6	8	10
	3	Bueno	8	10	12	14
	4	Excelente	12	14	16	18

Tabla 4.2: Diccionario de datos de etiquetas lingüísticas del atributo difuso `T2:Evaluación`.

La entidad difusa *EMPLEADO*^{F}, se puede componer de los atributos `{DNI, Evaluación,Valor_μj(Evaluación)}`. La expresión `Valor_μj(Evaluación)`, con *j* siendo el valor del índice indicado en la segunda columna de la Tabla 4.2 contiene, el grado de pertenencia a cada etiqueta lingüística. La representación en un esquema utilizando el modelo FuzzyEER se encuentra en la Figura 4.14 a).

Un ejemplo de los valores que puede llegar a tomar una instancia de la entidad *EMPLEADO*, puede ser, <maría, 1991, x_0>, donde x_0 es el valor de un referencial ordenado, y dicha instancia forma una o más instancia al convertirse en *EMPLEADO*^{F}. Según la Figura 4.13, la instancia en cuestión generaría un valor para las etiquetas. "bueno" y "excelente", asociados a las instancias: <maría, bueno, 0.8> y <maría, excelente, 0.2>, además, para las etiquetas: "deficiente" y "regular", su grado de pertenencia es 0, y no tienen que expresarse en *EMPLEADO*^{F}.

Otro caso es que María tenga una evaluación tal que su valor es difuso, y toma el valor "bueno". Así, esa instancia generaría ahora tres instancias en *EMPLEADO*^{F}, según las tres etiquetas con las que "bueno" se relaciona:

$$\text{“bueno”} \begin{cases} \text{“regular” con grado 0.5} \\ \text{“bueno” con grado 1} \\ \text{“excelente” con grado 0.5} \end{cases}$$

En la Figura 4.14 a), dado que cada atributo identificador principal de Empleado, tendrá uno o más grados de pertenencia en *EMPLEADO*^{F}, se puede representar una cardinalidad máxima de *n*. Para una instancia de *E* puede asociarse *n* instancias en E^F, donde *n* es el número de etiquetas lingüísticas (b_j).

En la Figura 4.14 b), se muestra un ejemplo de una *entidad débil difusa* definida por *dependencia de identificación*, para el caso que un dueño de animales que tenga una cierta preferencia por cada uno de ellos. Ese grado de importancia G^4 es identificado en la entidad débil difusa mediante un atributo que considera "el grado de importancia de cada animal".

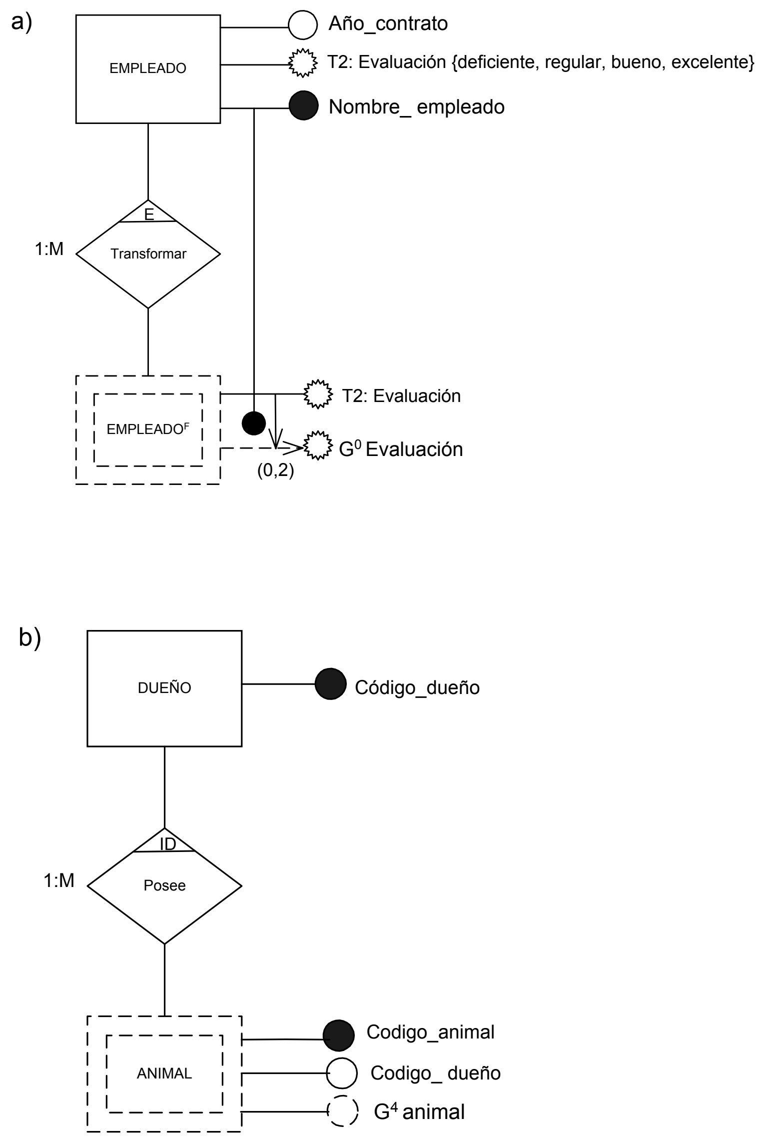


Figura 4.14: Ejemplo esquema FuzzyEER de entidad débil difusa. a) Con interrelación de dependencia de existencia, b) Con interrelación de dependencia de identificación.

Este tipo de entidad débil abarca y mejora la definición de Chaudhry et al. (1994 y 1999). Su ventaja principal es que acelera algunos tipos de consultas difusas al estar en la entidad E^F los grados ya calculados. Por ejemplo, ver los empleados que son "excelentes" (con grado mínimo 0.8), consiste en buscar en *EMPLEADO*^{F} aquellas instancias con evaluación = excelente y $G^0 \geq 0.8$. Otra opción es que esos grados también pueden ser calculados en vez de almacenarse, cuando sea necesario.

4.2.3 Interrelaciones Difusas en FuzzyEER

Hasta aquí se ha mostrado la definición, notación y algunos ejemplos de atributos, y entidades que se pueden representar en el modelo FuzzyEER. En este apartado presentamos una extensión al concepto de interrelación difusa considerando: atributos difusos y grados difusos.

Algunos autores tales como: Connolly et al. (1998) y Elmasri y Navathe (2002) llaman Vínculo o Tipo de interrelación (*relationship*) a la estructura genérica del conjunto de interrelaciones existentes entre dos o más tipos de entidad (en algunos casos se define una interrelación que asocian a la misma entidad, llamada interrelación recursiva), mientras que la ocurrencia (o instancia) de una interrelación será la vinculación existente entre las ocurrencias concretas de cada tipo de entidad que intervienen en la interrelación. Se representa el tipo de interrelación mediante un rombo etiquetado con el nombre de la interrelación, unido mediante arcos a los tipos de entidades asociadas.

También en De Miguel et al (1999) definen el concepto de *atributo de las interrelaciones* teniendo en cuenta que si una interrelación 1:N tiene un atributo asociado, este puede llevarse a la entidad cuya cardinalidad es máxima N. Semánticamente en ocasiones puede conservarse el atributo dependiendo de la interrelación, siendo esta última definición la que es utilizada en el modelo de datos FuzzyEER.

Definición 4.7: Un tipo de interrelación I con una o más entidades se considera **Interrelación difusa** si tiene un atributo difuso que relacione las entidades asociadas a *I*. Su representación gráfica es un rombo con línea cortada con un atributo que establece dos tipos de interrelación difusa:

a) Un Tipo de atributo difuso T1, T2, T3 o T4, que enlaza un tipo de entidad con las etiquetas de un tipo de atributo difuso, las cuales son vistas como instancias de un tipo de entidad. Ese tipo de atributo difuso tendrá sus características propias. Por ejemplo, una relación de similitud si es Tipo 3, y esa relación de similitud "puede" declarase explícitamente, a través de la interrelación difusa. Opcionalmente entre paréntesis el número máximo de valores asociados al atributo difuso. Véase Figura 4.15 a).

b) Un Tipo de grado difuso con un significado propio G^0, G^1, G_2, G^3, G^4 de un atributo difuso. Este tipo de interrelación difusa enlaza o relaciona dos o más tipos de instancias, asociando un grado a ese enlace, para cada pareja de instancias relacionadas. Véase Figura 4.15b)

La Definición 4.7 tienen asociada la notación (min,max) correspondiente a los tipos de entidades.

*

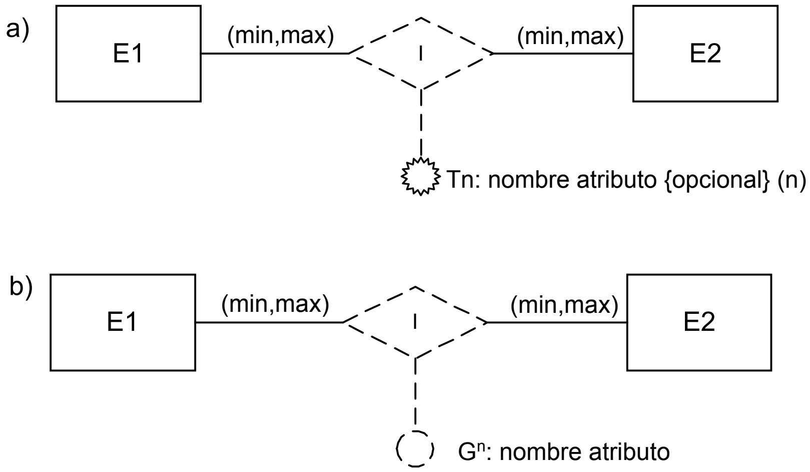


Figura 4.15: Notación de FuzzyEER para interrelaciones difusas. a) asociadas a un atributo difuso b) asociada a un grado.

Ejemplo 4.6: Se tiene la entidad *BARRIO***,** que puede establecer un esquema de {`Código_barrio`, `Nombre_barrio`, `Calidad`}. Los atributos de código y nombre son definidos, para este caso, atributos *crisp*. El atributo `Calidad` del barrio se identifica como atributo difuso Tipo 3, dado por un dominio perteneciente al conjunto {"malo", "regular", "bueno", "excelente"} que es previamente establecido.

Cada inmueble puede estar situado en un punto tal que pertenezca a varios barrios, o bien, que pertenezca a uno pero relativamente cerca de otro u otros barrios. Por ejemplo, para un inmueble se puede indicar que su barrio toma la siguiente distribución de posibilidad {0.5/Norte, 1/Oriente, 0.2/Plaza_España}, indicando que está situado propiamente en la zona del barrio de Oriente más cerca del barrio Norte que del barrio de la Plaza de España. Este es un caso especial, porque *BARRIO* es una entidad relacionada con la entidad *INMUEBLE*, de forma que un inmueble puede estar relacionado con varios. A la vez, un barrio para un inmueble concreto, puede tener un grado con el que ese inmueble pertenece a ese barrio. Por tanto, el atributo difuso `T3:Barrio_colindante`, genera un atributo difuso que relaciona las entidades *INMUEBLE* y *BARRIO* en la interrelación *pertenece* con un máximo de 3 barrios asociados a cada inmueble. Véase Figura 4.16.

La relación de proximidad sobre el atributo `T3:Barrio` puede representarse como interrelación difusa *cerca_de* como lo muestra la Figura 4.16.

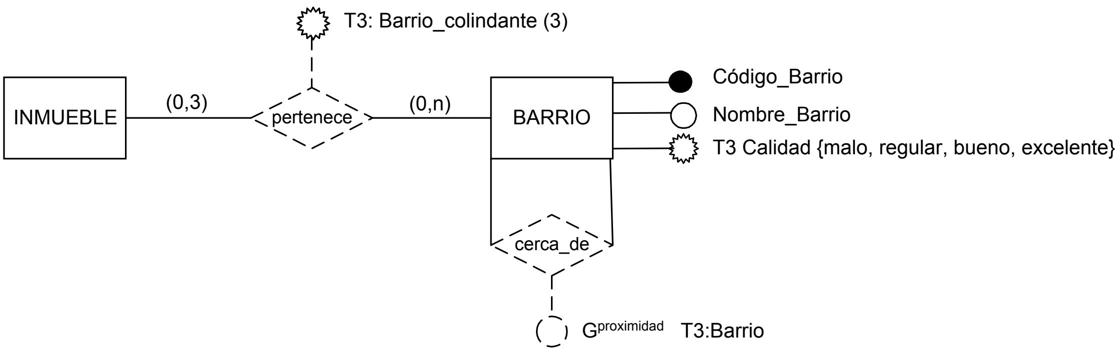


Figura 4.16: Interrelaciones difusas por atributos difuso y tipo de grado difuso.

Un ejemplo más detallado de este caso se encuentra en el apartado 5.2.

4.3 Restricciones para Interrelaciones Difusas en FuzzyEER

Una de las características que se pueden encontrar sobre las interrelaciones (también llamadas tipos de vínculo) en un modelo conceptual ER/EER son las restricciones en interrelaciones, las que limitan las posibles combinaciones de entidades o (tipos de entidad) que pueden participar en cada elemento de la interrelación. Estas restricciones se determinan a partir de las especificaciones de un sistema de información de un determinado universo del discurso. Se pueden distinguir tres tipos principales de restricciones de interrelación en un modelo ER/EER: participación, tipo de correspondencia según De Miguel et al. (1999), también llamadas Razón de cardinalidad por Elmasri y Navathe (2002), y la notación (min,max) o cardinalidad (min,max). Esta última, más que un tipo de restricción distinta es una notación especial de las dos anteriores, que abarca a ambas y las mejora.

En este apartado se pretende flexibilizar dichas restricciones de un modelo ER/EER. Todo ello, para facilitar la incorporación de información difusa o imprecisa en las restricciones de las interrelaciones o asociaciones (participación, tipo de correspondencia y notación (min,max)) para un modelo FuzzyEER. En este caso se utiliza la noción de cuantificador difuso aplicado a las restricciones que son extendidas. A continuación, se describe una representación gráfica para expresar estos conceptos, los esquemas resultantes serán también parte del modelo FuzzyEER propuesto, así como también un apartado de los umbrales utilizados para los cuantificadores que se utilizan en la definición de la restricción.

4.3.1 Umbrales de los Cuantificadores Difusos en la Aplicación de Restricciones Difusas en FuzzyEER

En el contexto de las bases de datos, los cuantificadores difusos (definidos en el apartado 3.2.5) han sido usados para flexibilizar consultas. Por ejemplo, en una base de datos difusa utilizando el operador división del álgebra relacional flexibiliza este tipo de consulta (Galindo et al., 2002). En el contexto de los modelos de datos conceptuales, los cuantificadores difusos son expresiones sobre el número de instancias que satisfacen una determinada condición o porcentaje de éstas con respecto al total de instancias de una entidad. En este sentido esta tesis usa cuantificadores difusos a los que denotamos con una Q, en las restricciones de interrelaciones, y son definidos en un diccionario de datos de una base de datos.

Se establecen dos tipos de umbrales, $\gamma \in [0,1]$ indicando el grado mínimo de difuminado denotado como $Q[\gamma]$. También se definen umbrales indicando un "área de aviso" entre dos umbrales denotado como $Q[\gamma, \delta]$ como lo indica la Figura 4.17.

Consiguientemente, el significado de ciertas restricciones asociadas a los valores de γ, Φ (Φ definido en la Ecuación (3.4)) en una base de datos se puede obtener a partir de las siguientes ecuaciones.

$$Q(\Phi) \geq \gamma \tag{4.4}$$

Cada vez que la base de datos se modifica, el SGBD evalúa el valor de Φ y se comprueba si la Ecuación 4.4 se satisface:

Si Q es una función creciente se puede cambiar la Ecuación 4.4, donde se obtiene que:

$$\Phi \geq Q^{-1} (\gamma) \tag{4.5}$$

Igualmente si Q es una función decreciente, se obtiene que:

$$\Phi \leq Q^{-1} (\gamma) \tag{4.6}$$

Por supuesto que el cuantificador Φ debe ser definido en el diccionario de datos que acompaña al modelo conceptual de datos.

En general, si Q es un cuantificador difuso relativo entonces, la Ecuación 4.5 expresa que la restricción debe satisfacerse en la base de datos en un *porcentaje* mínimo de $100Q^{-1}(\gamma)$. En ese caso, también será posible expresar ese *porcentaje* en vez del cuantificador Q y del umbral γ. Aunque esto pueda parecer más simple, hay que hacer notar que se pierde la expresividad intuitiva y natural del cuantificador y que esto no es válido con cuantificadores difusos absolutos ni para cuantificadores relativos que no sean crecientes.

Además, puede establecerse otro valor opcional δ, superior al grado de cumplimiento γ, de la siguiente forma: $Q[\gamma,\delta]$, tal que $\gamma<\delta$. El valor δ es más restrictivo y establece que cuando se incumpla la restricción con ese valor superior, o sea, entre el intervalo $[\gamma,\delta]$ se genera un "área de aviso" que permitida de una restricción determinada dando un aviso que se esté en ella. Véase Figura 4.17.

Por tanto, un cuantificador puede aparecer de tres formas en un modelo de datos que muestre la participación en interrelaciones. Estas son:

1. Sin umbral, (por defecto $\gamma=0.5$) por ejemplo, "casi_todo" (Figura 3.7).
2. Con umbral γ representado por [γ], por ejemplo "casi_todo[0.2]" (Figura 4.20).
3. Con dos umbrales γ y δ, con $\gamma<\delta$ representado por [γ, δ], por ejemplo "casi_todo [0.2, 0.6]" (Figura 4.22).

Para el caso 1 se escoge el valor 0.5, por ser el punto medio del dominio de los umbrales, que son [0,1].

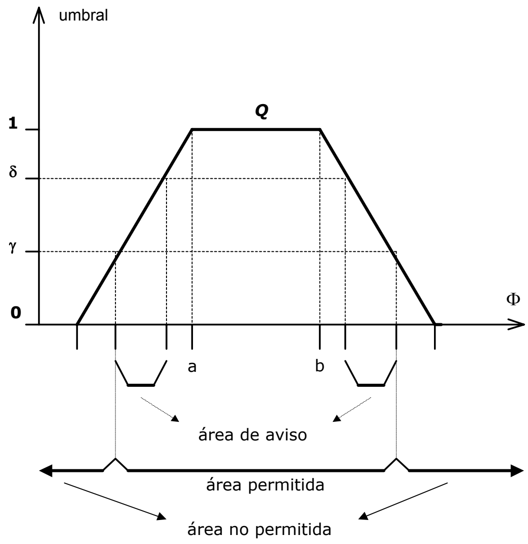


Figura 4.17: Umbrales γ y δ para cuantificadores difusos y áreas generadas de "aproximadamente entre a y b".

Cada uno de estos casos representa distintas formas de restricciones sobre un grado de verdad de un domino, ya sea el cuantificador relativo o absoluto.

4.3.2 Participación Difusa de Interrelaciones en FuzzyEER

Elmasri y Navathe (2000) dicen que la restricción de participación especifica si la existencia de una entidad depende de que esté relacionada con la otra entidad a través de un tipo de vínculo. La participación de una entidad en una interrelación puede ser total (si cada entidad debe relacionarse forzosamente con la otra u otras entidades de la interrelación) o parcial (si dicha interrelación no es obligatoria para todas las entidades pertenecientes a ese tipo). Elmasri y Navathe (2000) y Connolly et al. (1998) representan la participación total por una doble línea uniendo la entidad correspondiente con el rombo de la interrrelación, mientras que la participación parcial se representa con una línea simple.

En el modelo FuzzyEER se propone la participación de interrelación difusa utilizando (principalmente) un cuantificador difuso relativo, pero también se pueden aplicar cuantificadores absolutos. La representación se hace utilizando una *línea interceptada por un arco*, adjuntando el cuantificador utilizado, seguido opcionalmente, de un umbral o grado de cumplimiento γ entre corchetes para dicho cuantificador, *cuyo valor por defecto es 0.5*.

Definición 4.8: Sean E_1 y E_2 dos entidades interrelacionadas con I. Se define una **restricción de participación difusa** de E_1 en I, y se representa con un cuantificador Q sobre la línea que une E_1 con I, siendo esta línea sustituida por una línea normal cruzada con un arco etiquetado con Q La notación de esta definición se muestra en la Figura 4.18. Está restricción establece que:

$$Q(\Phi) \geq \gamma \tag{4.7}$$

donde γ es el umbral mínimo establecido para Q, y Φ es:

$$\Phi = \begin{cases} a & \text{si } Q = Q_{\text{abs}} \\ a/b & \text{si } Q = Q_{\text{rel}} \end{cases} \tag{4.8}$$

Siendo a el número de instancias de E_1 relacionadas con E_2, y b siendo el número total de instancias de E_1.

*

Si Q tiene dos umbrales $[\gamma,\delta]$ entonces define un "área de aviso", como se explicó en el apartado 4.3.1. Por tanto, siempre que exista un "área de aviso", en la implementación de ésta restricción, habrá que generar un "mensaje" cuando la condición establecida por la restricción se cumpla y, a la vez, no se cumpla la misma restricción utilizando el umbral δ en vez del umbral γ (véase Figura 4.17).

O sea, la condición establecida por la restricción define un "área permitida" y como el "área de aviso" está dentro de un área permitida, en el "área de aviso" debe forzosamente cumplirse esa condición. Como δ es más restrictivo que (γ,δ) pudiera ser que la condición δ no se cumpla, y en ese caso es cuando se genera el "mensaje".

Para la Definición 4.8, por ejemplo, un "área de aviso", está definida cuando se cumple la Ecuación 4.7 y no se cumple que:

$$Q(\Phi) \geq \delta \tag{4.9}$$

o lo que es lo mismo, que se cumpla que:

$$Q(\Phi) \geq \gamma \ \wedge \ Q(\Phi) < \delta \tag{4.10}$$

Resumiendo, si la Ecuación 4.10 se cumple, se genera un “mensaje”, si se cumple la Ecuación 4.9 la restricción en curso *está totalmente permitida*, y en cualquier otro caso, la restricción *no está permitida*.

Igualmente, como se trató en el apartado 4.3.1, si Q no tiene umbral que γ, se supone γ=0.5, y si Q tiene dos umbrales $[\gamma,\delta]$, el umbral δ establece un “área de aviso” ($Q(\Phi) < \delta$), dentro del “área de permitida”.

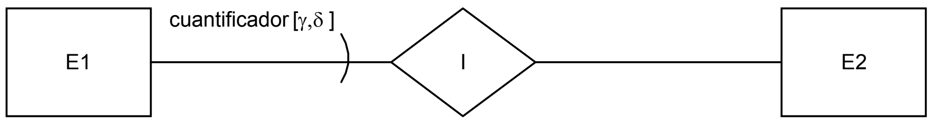


Figura 4.18: Notación en el modelo FuzzyEER de participación difusa en interrelaciones.

Nótese que en la Figura 4.18 la participación de la entidad E_1 y la interrelación I es difusa, en cambio, la participación entre la entidad E_2 y la interrelación I es una participación parcial.

Participación de Interrelaciones con Cuantificador Relativo

Como se vio en el apartado 4.3.1, el cuantificador relativo se aplica a la división entre el número de elementos que cumple cierta condición y el número total de elementos existentes. Si se considera el cuantificador “casi_todo” (Figura 3.7), al cual se le pueden aplicar algunos grados de cumplimientos como los que se muestran en las Figuras 4.20 y 4.22, para establecer distintas “áreas permitidas”.

Ejemplo 4.7: Supongamos una entidad *EMPLEADO* y otra *PROYECTO* unidas a través de una interrelación *Trabaja_en*. La participación de la entidad *EMPLEADO* en esa interrelación puede ser representada a través de un cuantificador difuso relativo como “casi todo” (Figura 3.7), indicando que “*Casi todos los empleados trabajan para algún proyecto*”. La Figura 4.19, representa estas dos entidades, la interrelación entre ellas, además de la restricción difusa del cuantificador “casi_todo”.

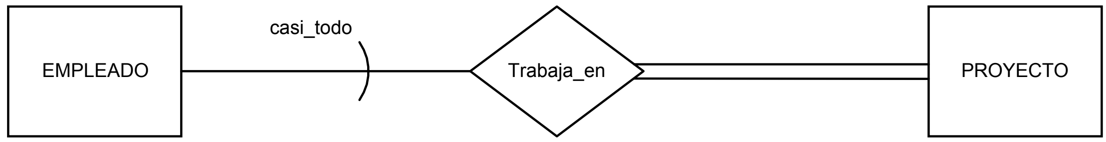


Figura 4.19: Representación gráfica de restricción de participación en interrelación con cuantificador difuso "casi_todo".

Ahora bien, en el Ejemplo 4.7 al cuantificador "casi_todo" se le puede incorporar un umbral γ, como por ejemplo "casi todo [0.2]", mostrado en la Figura 4.20. Dicha figura, indica el grado mínimo [0.2] con el que este cuantificador "casi todo" debe cumplirse la restricción en el modelo de datos.

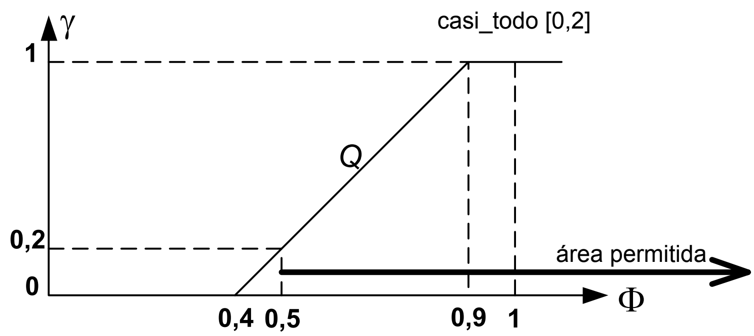


Figura 4.20: Cuantificador difuso "casi_todo [0.2]".

Es decir, si se divide el número de empleados que trabajan para algún proyecto, entre el número de empleados total de la base de datos, el resultado Φ debe ser, de acuerdo con la Ecuación 4.5, un valor mayor o igual a $Q^{-1}(0.2) = 0.5$, ya que ese es el valor para que el cuantificador "casi_todo", alcance el valor de cumplimiento de 0.2: $Q(0.5) = 0.2$. A partir de ese valor 0.5 este cuantificador obtiene un grado mayor a 0.2, que era la restricción impuesta por el umbral γ en el cuantificador inicial.

En este ejemplo, el valor 0.5 obtenido por la expresión $Q^{-1}(0.2)$ indica la restricción que es impuesta en el modelo de datos. Es decir, un mínimo del 50% de los empleados deben trabajar para algún proyecto. La representación gráfica de este caso se encuentra en la Figura 4.21.

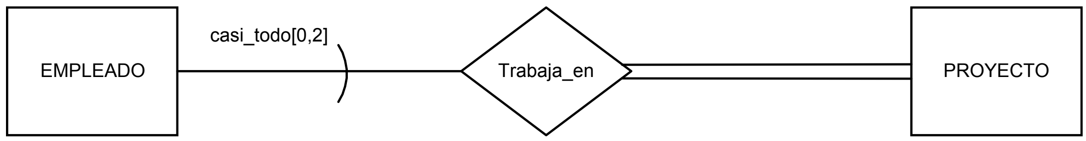


Figura 4.21: Representación gráfica de interrelación de participación difusa "casi_todo[0.2]".

Otra restricción que se le puede incorporar al Ejemplo 4.7, es el uso de dos umbrales, por ejemplo, el cuantificador "casi_todo[0.2,0.6]" que determinará un "área de aviso" entre los umbrales 0.2 y 0.6, teniendo como "área permitida" a partir del umbral 0.2, tal como lo muestra la Figura 4.22. Para umbrales menores a 0.2 la restricción no estará permitida.

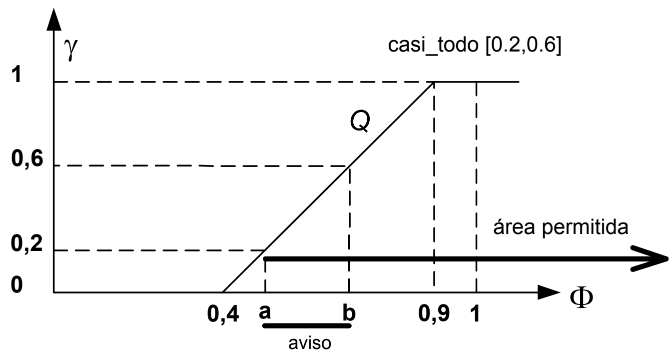


Figura 4.22: Cuantificador "casi_todo [0.2, 0.6]" para participación difusa.

El modelo de datos conceptual en FuzzyEER que acompaña al cuantificador difuso relativo "casi_todo [0.2, 0.6]", se muestra en la Figura 4.23.

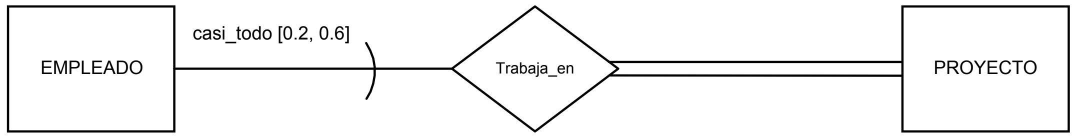


Figura 4.23: Representación gráfica de cuantificador "casi_todo [0.2, 0.6]" para la participación difusa en el modelo FuzzyEER.

Lo anterior es importante, ya que, en algún modelo puede ser interesante establecer varios cuantificadores difusos como restricción de participación difusa (véase Figura 4.25). La Figura 4.24, muestra dos ejemplos de uso de dos cuantificadores con umbrales iguales y distintos.

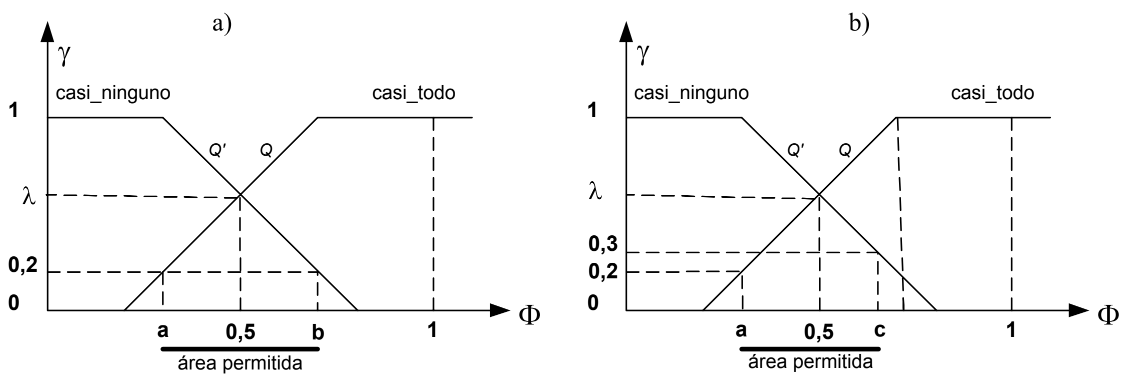


Figura 4.24: Dos cuantificadores relativos, "casi ninguno" y "casi todo" con umbral [0.2] y umbral [0.2,0.3].

La Figura 4.24a), el "área permitida" para el cuantificador difuso "casi_todo[0.2], está dada por $\Phi \geq$ a con a $Q^{-1}(0.2)$, y para el cuantificador difuso "casi_ninguno [0.2]", el "área permitida" es $\Phi \leq$ b (con b= $Q^{-1}(0.2)$). Esto implica que, Φ pertenece al intervalo [a,b] y en ese intervalo se cumplen los dos cuantificadores con grado mayor o igual que [0.2].

La representación en un modelo FuzzyEER para este caso se encuentra en la Figura 4.25, donde, a la entidad *EMPLEADO* e interrelación *Trabaja_en* se aplica la restricción de participación difusa de los cuantificadores "casi_ninguno" y "casi_todo", ambos con umbrales [0.2]. La restricción de participación entre la entidad *PROYECTO* y la interrelación *Trabaja_en* es total.

La Figura 4.24 b), el "área permitida" para el cuantificador "casi_todo [0.2]", está dada por $\Phi \geq a$ y, para "casi_ninguno [0.3]", el área permitida es $\Phi \leq c$ (con $c=(Q')^{-1}(0.3)$). Esto implica que, Φ pertenece al intervalo [a,c], y en ese intervalo se cumplen los dos cuantificadores con grado [0.2] y [0.3], respectivamente.

Obsérvese que a pesar de la aparente contradicción entre estos dos cuantificadores, en la Figura 4.24 existe un "área permitida". Sin embargo, si elevamos ambos umbrales por encima del nivel λ, entonces no existe esa área permitida y la definición de la restricción del modelo será no permitida o estará mal definida.

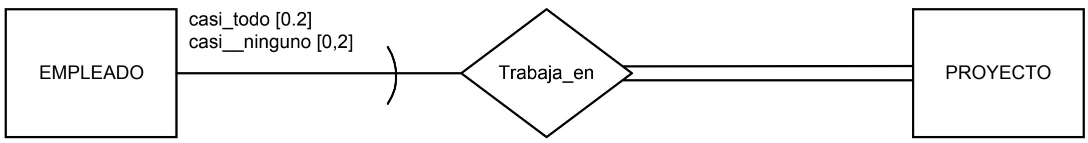


Figura 4.25: Restricción de participación difusa con dos cuantificadores en FuzzyEER.

El área permitida anterior puede también conseguirse, utilizando un único cuantificador del tipo "aproximadamente la mitad" (con un γ adecuado). Sin embargo, aunque el "área permitida" por ambas técnicas sea idéntica, no lo será el grado con que se cumple la restricción, y ese grado puede ser importante, como restricción de participación, en algunas aplicaciones.

Como se observa, el cuantificador difuso expresado en este tipo de restricciones también puede ser absoluto, aunque debido al significado de una restricción de participación éste será, en general, relativo ya que el número de instancias de la entidad afectada puede variar mucho.

Restricción de Participación Difusa en Interrelaciones con Cuantificador Absoluto

En caso de un cuantificador difuso absoluto, como por ejemplo "muchos" (véase Figura 4.26), o "aproximadamente entre 100 y 200". Este cuantificador restringirá la cantidad de instancias que interrelacionan dos o más entidades. En el Ejemplo 4.8, que veremos a continuación, se restringirá la cantidad de empleados que tienen algún proyecto asignado para el cual trabajar.

El cuantificador absoluto puede ser utilizado de tres formas como se describe en el apartado 4.3.1. Por ejemplo, un cuantificador "muchos" puede tener un "área permitida" con un umbral de [0.2], obteniendo el cuantificador absoluto "muchos [0.2]", siendo el "área permitida" a partir de b=Q^{-1}(0.2), como lo muestra la Figura 4.26. También se puede dar el caso que tenga un intervalo de grados, como: "muchos [0.2, 0.6]".

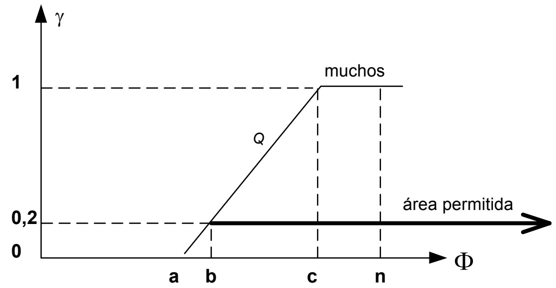


Figura 4.26: Representación de cuantificador "muchos" con grado de cumplimiento [0.2].

Ejemplo 4.8: Siguiendo con el Ejemplo 4.7, si usamos el cuantificador "muchos [0.2]", la restricción representa que *"Muchos empleados tienen proyectos con grado de cumplimiento 0.2"*. Además, si en la Figura 4.26, los valores a, b, c toman los valores 3, 9, 25 respectivamente, para representar el cuantificador "muchos", el "área permitida" será a partir de *nueve empleados*, lo que significa que al *menos nueve empleados* deben tener algún proyecto. La Figura 4.27 muestra el modelo FuzzyEER asociado a ésta restricción de participación.

Esto indica que deben existir muchos empleados que tengan proyectos asignados. Para ello, se debe cumplir que $n \geq$b (véase Figura 4.26) donde n es el número de instancias de la entidad *EMPLEADO*, relacionados con "tiene" proyectos asignados.

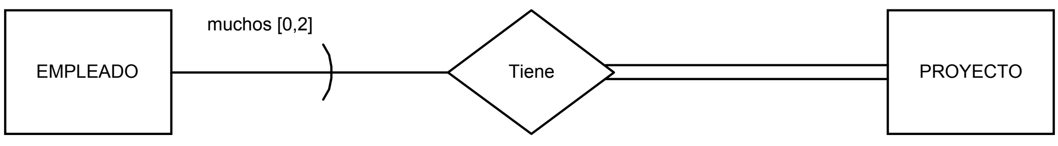


Figura 4.27: Modelo FuzzyEER de restricción de participación difusa para cuantificador difuso absoluto "muchos [0.2]".

Con los ejemplos anteriormente mostrados, se puede concluir que una restricción de participación difusa no es tan restrictiva como la restricción de participación total, ni tan permisiva como la restricción de participación parcial, por lo que las restricciones de participación difusas extienden el modelo EER, permitiendo una nueva expresividad que no se encuentra en el modelo tradicional.

Por otro lado, es conveniente, por lo general, utilizar cuantificadores relativos en el modelo FuzzyEER, ya que la cantidad total de instancias de una entidad pueden variar constantemente, y el cuantificador relativo, al estar definido respecto a ese total, se muestra flexible respecto del número de instancias involucradas en la relación difusa. En cambio, el cuantificador absoluto fija (o restringe) la cantidad de instancias involucradas, y es inflexible aunque se modificara el número total de instancias en la entidad.

4.3.3 Tipo de Correspondencia Difusa en FuzzyEER

En el modelado clásico, la restricción de tipo de correspondencia (también llamada, razón de cardinalidad), restringe el número de elementos de la interrelación en los que puede participar una entidad. Los tipos de correspondencia más comunes en el caso de interrelaciones binarias, donde se expresa si la relación entre entidades, es de "*uno a uno*" (1:1), de "*uno a muchos*" (1:N), o de "*muchos a muchos*" (N:M), que se definen en Elmasri y Navathe (2002) y Connolly et al. (1998) como *Razón de cardinalidad*, mientras que, De Miguel et al. (1999) la llama *tipo de correspondencia*. Esta restricción en el modelo FuzzyEER se usará como *tipo de correspondencia difusa*.

La restricción que permite expresar el tipo de correspondencia como un valor difuso utilizando (principalmente) un cuantificador difuso absoluto, corresponderá a un tipo de correspondencia difusa. La representación se hará escribiendo ambos cuantificadores justo bajo el rombo que representa la interrelación entre ambas entidades y separados por el carácter ":", con el siguiente formato:

Cuantificador difuso 1 : Cuantificador difuso 2

El cuantificador situado a la izquierda del separador ":" corresponderá a la entidad situada a la izquierda, y el cuantificador de la derecha para la otra entidad. Opcionalmente, puede ponerse uno o dos grados de cumplimiento entre corchetes [γ, δ], para cada cuantificador con el mismo significado y valor por defecto que el explicado anteriormente para las restricciones de participación difusas.

Definición 4.9: Sean E_1 y E_2 dos entidades interrelacionadas con R como lo muestra la Figura 4.28, suponemos que e_i con i=1,2,...,b_1, son las instancias de E_1 y w_j con j=1,2,...,b_2, son las instancias de E_2. Una **restricción de tipo de correspondencia** consiste en dos cuantificadores difusos separados por el carácter ":", de la forma "Q_1:Q_2", que afecta respectivamente a E_1 y E_2, haciendo cumplir las restricciones 4.11 y 4.13:

$$Q_1(\Phi_{1i}) \geq \gamma_1 \qquad \forall\, i = 1,2,...,b_1 \tag{4.11}$$

donde γ_1 es el umbral mínimo establecido por Q_1. Por otra parte, b_1 es el número total de instancias de E_1, y Φ_{1i} con i = 1,2,..., b_1, está definido como:

$$\Phi_{1i} = \begin{cases} a_i & \text{si } Q_1 \text{ es absoluto} \\ a_i/b_2 & \text{si } Q_1 \text{ es relativo} \end{cases} \tag{4.12}$$

siendo a_i el número de instancias de E_2 con las que se relaciona la instancia $e_i \in E_1$, y b_2 el número total de instancias de E_2. Se cumple que:

$$Q_2\,(\Phi_{2j}) \geq \gamma_2 \qquad \forall\, j = 1,2,..., b_2 \tag{4.13}$$

donde γ_2 es el umbral mínimo establecido para Q_2. Por otra parte, Φ_{2j} con j = 1,2,...,b_2, es definido como:

$$\Phi_{2j} = \begin{cases} a_j & \text{si } Q_2 \text{ es absoluto} \\ a_j/b_1 & \text{si } Q_2 \text{ es relativo} \end{cases} \tag{4.14}$$

siendo a_j el número de instancias de E_1 con las que se relaciona la instancia w_j de E_2.

*

Generalizando lo anterior, el cuantificador Q_1 debe cumplirse para todos los valores de cuantificación Φ_{1i} con i=1,2,...,b_1, o dicho de otra forma, los valores de Φ_{1i} deben estar en un "área permitida" de Q_1 según su umbral γ_1 establecido. También debe cumplirse Q_2 para todos los valores de cuantificación Φ_{2j} con j = 1,2,...,b_2, es decir, los valores Φ_{2j} deben estar en el "área permitida" de Q_2 según su umbral γ_2, establecido.

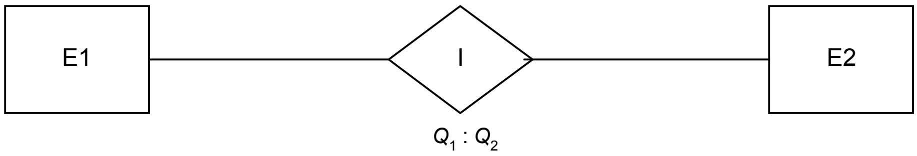


Figura 4.28: Representación de tipo de correspondencia difusa en el modelo FuzzyEER.

Ejemplo 4.9: Se tienen dos entidades *EMPLEADO* y *PROYECTO* unidos por la interrelación *Trabaja_en*, ahora supongamos que, la entidad *EMPLEADO* está a la izquierda de la interrelación *Trabaja_en*, y la entidad *PROYECTO* está a la derecha, entonces una restricción de tipo de correspondencia difusa podrá ser, por ejemplo: menos_de_aprox_ 3 [0.8] : aprox_8 [0.4]. Véase Figura 4.29. En dicha figura, se ha considerado la participación parcial o total según la entidad que sea a modo de ejemplo.

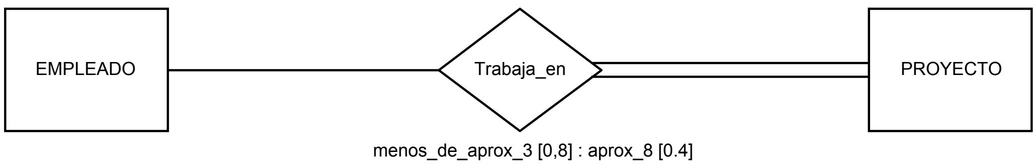


Figura 4.29: Representación gráfica de restricción tipo de correspondencia difusa menos_de_aprox_ 3 [0.8] : aprox_8 [0.4].

Las restricciones de participación difusa mostradas en la Figura 4.29, expresan la condición de que cada empleado trabajará en un máximo de aproximadamente *tres* proyectos, y cada proyecto tendrá aproximadamente *ocho* empleados. Además, exigiendo que en ambas restricciones se cumplan con los grados de verdad mínimos indicados entre corchetes, [0.8] y [0.4] respectivamente. El valor de verdad perteneciente a este tipo de correspondencia lo muestra la Figura 4.30, donde Φ_1 es el número de proyectos en los que trabaja un empleado. Así, todos los empleados deben tener su Φ_1 en el "área permitida".

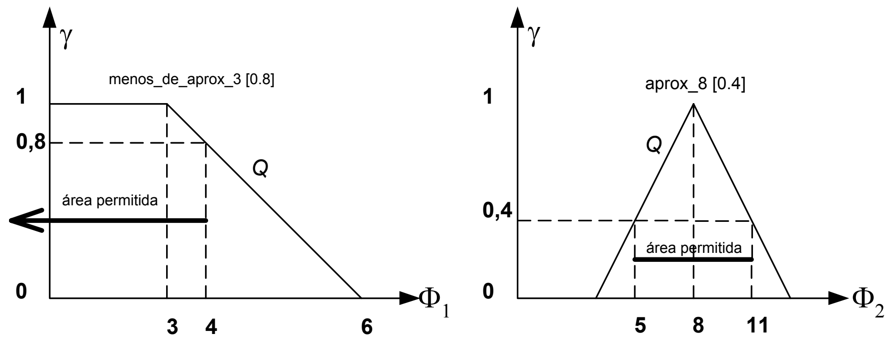


Figura 4.30: Cuantificadores "menos_de_aprox_3 [0.8]" y "aprox_8 [0.4]".

Por otra parte, Φ_2 es el número de empleados asignados a cierto proyecto. También en este caso todos los proyectos deben tener su Φ_2, dentro del "área permitida".

Por otra parte, en la Figura 4.30, como se observa, el cuantificador difuso de este tipo de restricciones también puede ser relativo, aunque debido al significado de una restricción de tipo de correspondencia, éste será, en general absoluto. En caso de que sea un cuantificador difuso relativo, el cuantificador indicará la cantidad de instancias de la otra entidad con las que se interrelaciona cada entidad con respecto al número total de instancias de la otra entidad que existen. Así, si en el Ejemplo 4.7 usamos el cuantificador "casi_todo" a la izquierda esto significa que "*cada empleado deberá trabajar en casi todos los proyectos existentes*".

Incluso, en algún modelo puede ser interesante establecer varios cuantificadores difusos en algún o en ambos lados de una restricción de tipo de correspondencia difusa.

Si una interrelación tiene grado mayor a dos, puede ponerse el cuantificador difuso de tipo de correspondencia asociado a cada entidad al lado del arco que une esa entidad con la interrelación. Si hubiera ya un cuantificador para la restricción de participación difusa (línea zigzag), entonces, para evitar ambigüedad se debe poner delante del cuantificador de tipo de correspondencia un texto que identifique a que entidades pertenece la restricción.

4.3.4 Notación (min,max) Difusa en FuzzyEER

Una notación alternativa de ER/EER con la que también se puede especificar las restricciones estructurales, consiste en asociar un par de números enteros (min, max) a cada participación en de un tipo de entidad *E*, en una interrelación *I*, donde $0 \leq \min \leq \max$ y $\max \geq 1$. Los números indican que, para cada entidad *e* de *E*, *e* debe participar en por lo menos en "min" y como máximo en "max" elementos de la interrelación de *I* en todo momento.

La representación de la restricción de tipo de correspondencia difusa, para el caso de la participación difusa, se analiza cuando ésta se refiere a todas las instancias de una entidad, y no para cada una individualmente. En el modelado de FuzzyEER, tanto "min" como "max", pueden tomar valores que sean cuantificadores difusos, ya sean relativo o absoluto, tanto para el mínimo como para el máximo.

Definición 4.10: Sean E_1 y E_2 dos entidades interrelacionadas por *I*, denotaremos como e_i para i=1,2,...,b_1 las instancias de E_1, y con w_j para j=1,2,...,b_2 las instancias de E_2. Se define una **restricción usando la notación (min, max) difusa** de E_1 en *I* y se representa escribiendo junto a la línea que une E_1 con *I*, dos cuantificadores (por lo general absolutos) entre paréntesis, ($Q_{\min}$, $Q_{\max}$), como lo muestra la Figura 4.33. Esta restricción establece que:

$$\lambda_{\min} \leq \Phi_{\min,i} \quad \wedge \quad \lambda_{\max} \geq \Phi_{\max,i} \qquad \forall\, i = 1,2,...,b_1 \tag{4.15}$$

donde, b_1 el número total de instancias de E_1 y,

$$\Phi_{\min,i} = \begin{cases} a_i & \text{si } Q_{\min} \text{ es absoluto} \\ a_i/b_2 & \text{si } Q_{\min} \text{ es relativo} \end{cases}$$
$$\Phi_{\max,i} = \begin{cases} a_i & \text{si } Q_{\max} \text{ es absoluto} \\ a_i/b_2 & \text{si } Q_{\max} \text{ es relativo} \end{cases} \tag{4.16}$$

En este caso a_i es el número de instancias de E_2 con las que se relaciona la instancia e_i que pertenecen a E_1, y b_2 es el número total de instancias de E_2. Además, tenemos que:

$$\lambda_{\min} = \min\{\alpha : \alpha = (Q_{\min})^{-1}(\gamma_{\min})\} \tag{4.17}$$
$$\lambda_{\max} = \max\{\beta : \beta = (Q_{\max})^{-1}(\gamma_{\max})\}$$

donde, γ_1 y γ_2 son los umbrales mínimos establecidos para Q_{min} y Q_{max} respectivamente.

*

Por lo tanto, el intervalo $[\lambda_{min},\lambda_{max}]$ es el "área permitida". El "área de aviso" se define, si están demarcados dos umbrales δ_{min} y δ_{max} para Q_{min} y Q_{max} respectivamente, cuando se cumple la condición de la Ecuación 4.15 y no se cumple que:

$$\lambda'_{min} \leq \Phi_{min,i} \quad \wedge \quad \lambda'_{max} \geq \Phi_{max,i} \quad \forall\, i = 1,2,...,b_1 \tag{4.18}$$

donde,

$$\lambda'_{min} = \min \{\alpha : \alpha = (Q_{min})^{-1}(\delta_{min})\} \tag{4.19}$$

$$\lambda'_{max} = \max \{\beta : \beta = (Q_{max})^{-1}(\delta_{max})\}$$

Así, el área de aviso es la unión de dos intervalos: $[\lambda_{min},\lambda'_{min}] \cup [\lambda'_{max},\lambda_{max}]$. Sabemos que: $\lambda_{min} < \lambda'_{min}$ y que $\lambda'_{max} < \lambda_{max}$ porque $\gamma_{min} < \delta_{min}$ y $\gamma_{max} < \delta_{max}$ respectivamente, y los cuantificadores están definidos como funciones convexas.

Obsérvese que las Ecuaciones 4.19 son iguales que las Ecuaciones 4.17, si cambiamos γ_{min} y γ_{max}, por δ_{min} y δ_{max} respectivamente. Además, en la Ecuación 4.18 se puede aplicar la ley de De Morgan quedando:

$$\lambda'_{min} > \Phi_{min,i} \quad \vee \quad \lambda'_{max} < \Phi_{max,i} \quad \forall\, i = 1,2,...,n \tag{4.20}$$

*

Consideraciones a tener en cuenta: En la elección de algunos cuantificadores a usar, en cualquiera de las restricciones anteriores, hay que considerar que el cuantificador difuso sea representativo, por lo que se hacen los siguientes alcances:

a) Si $Q_{min}(0) \geq \gamma_{min}$ no usar este cuantificador $Q_{min}[\gamma_{min}]$, usar 0 en el lugar de Q_{min}: $[0,Q_{max}]$.
b) Si $Q_{max}(\psi) \geq \gamma_{max}$, donde ψ es el valor máximo que puede tomar el dominio subyacente del cuantificador Q_{max} (si Q_{max} es relativo, ψ=1), entonces no debe usarse este cuantificador $Q_{max}[\gamma_{max}]$, sino que se debe usar la letra "N" en su lugar, denotando una relación de "muchos": $[Q_{min}$, N].

c) Además, si $Q_{\min}$ y $Q_{\max}$ son del mismo tipo (relativos o absolutos), implica que: $\Phi_{\min,i} = \Phi_{\max,i}$.

En síntesis, se trata de no usar cuantificadores que especifiquen mal la restricción, como por ejemplo: [aprox_3_o_menos, casi_todos].

Ejemplo 4.10: En el contexto de las entidades *EMPLEADO* y *PROYECTO*, pueden ponerse las siguientes restricciones utilizando la cardinalidad difusa. Estas restricciones están representadas en la Figura 4.33. En el lado de la entidad *EMPLEADO* puede ponerse la siguiente restricción de cardinalidad:

(0, aprox_3 [0.25, 0.75])

Indicando que un empleado puede no pertenecer a ningún proyecto (0 como mínimo), y como máximo puede pertenecer a aproximadamente 3 proyectos. Los dos valores, tras el cuantificador, indican que si éste se cumple con un grado mayor a 0.75 se permitirá la tarea asociada a la restricción normalmente (Figura 4.31 área permitida), si se cumple con un grado entre 0.25 y 0.75 entonces se avisará al usuario pero se permitirá la tarea (Figura 4.31 área de aviso) y, si la restricción se cumple con un grado inferior a 0.25, querrá decir que se está incumpliendo la restricción con un grado no permitido y, por tanto, la tarea en curso no debe permitirse (Figura 4.31 área No permitida).

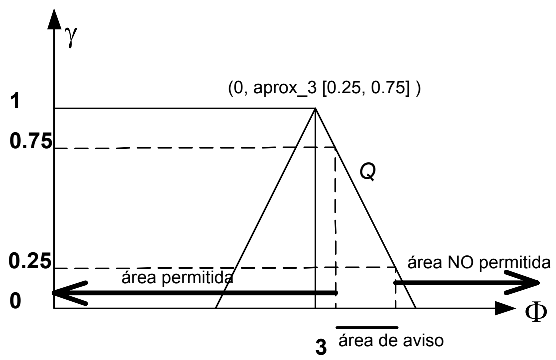


Figura 4.31: Representación de la restricción de cardinalidad (0, aprox_3[0.25,0.75]).

Ahora si, en el lado de la entidad *PROYECTO* del mismo ejemplo, se pone la siguiente restricción (min, max):

(aprox_2, aprox_8 [0.25])

Esto indican que un proyecto tiene que tener como mínimo aproximadamente 2 empleados (con grado 0.5 como mínimo). Si el cuantificador “aprox_2”, está definido como un triángulo, según la Figura 3.1 con n = 2 y margen = 2, entonces el valor 0.5 se consigue con el valor mínimo 1, por lo que este cuantificador con el grado mínimo 0.5 garantiza la participación total de la entidad *PROYECTO* en la interrelación *Trabaja_en*. Esta eventualidad es también indicada por la doble línea que une la entidad con la interrelación. La Figura 4.32 muestra el intervalo permitido por esta restricción.

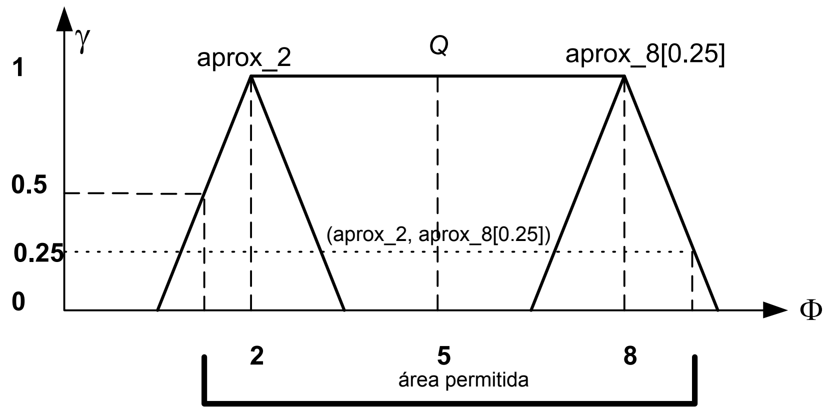


Figura 4.32: Cuantificador para la restricción (min, max) (aprox_2, aprox_8[0.25]).

Por otra parte, el número de empleados en cada proyecto está restringido a un máximo de aproximadamente 8 (con grado mínimo 0.25). La Figura 4.33 muestra el esquema en FuzzyEER resultante.

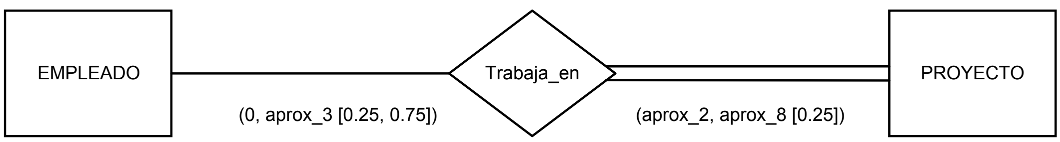


Figura 4.33: Representación gráfica de restricción de cardinalidad difusas con notación (min, max).

4.3.4.1 Análisis entre Restricciones de Tipo de Correspondencia y Notación Difusa (min,max) en FuzzyEER

Con respecto a las restricciones de tipo de correspondencia (apartado 4.3.3) está claro que la semántica de ambas notaciones es diferente. Como se observa que en el Ejemplo 4.9, el cuantificador "aprox_8" indica que un Proyecto debe tener aproximadamente 8 empleados, mientras que en el Ejemplo 4.10, el mismo cuantificador, indica que un Proyecto puede tener como máximo aproximadamente 8 empleados.

Tanto la participación, como el tipo de correspondencia no difusas de una entidad en una interrelación se pueden expresar a través de esta restricción de cardinalidad con la notación (min, max), donde min y max indican respectivamente, el número mínimo y máximo de instancias de la entidad que participan en la interrelación, definidas en De Miguel et al. (1999); Elmasri y Navathe (2002) y Connolly et al. (1998).

En general, una restricción de tipo de correspondencia difusa con cualquier tipo de cuantificador puede representarse mediante la cardinalidad difusa usando dos valores (min, max). La restricción de tipo de correspondencia difusa expresada en el Ejemplo 4.9, puede expresarse en notación (min, max) difusa, haciendo que el valor de min sea igual al de max y ambos tomando el valor de "aprox_8".

La expresividad también es equivalente hacia el lado contrario con dos excepciones:

- Si la notación (min, max) utiliza dos cuantificadores del mismo tipo (absolutos o relativos), entonces esa restricción puede expresarse mediante una restricción de tipo de correspondencia difusa mediante un cuantificador (también convexo) que englobe a ambos. Por ejemplo, la restricción del Ejemplo 4.10, puede expresarse con la notación de cardinalidad difusa utilizando en lugar de los cuantificadores "aprox_2" y "aprox_8" un cuantificador más amplio como "aprox_entre_2_y_8". De igual forma si estos cuantificadores tienen algún umbral, por ejemplo, "aprox_entre_2_y_8[0.25]" (véase Figura 4.32).

- Si la notación (min, max) utiliza dos cuantificadores de distinto tipo (uno absoluto y otro relativo), entonces esa restricción no puede expresarse con un único cuantificador de tipo de correspondencia difusa. Esto es debido a que, al ser cuantificadores de distinto tipo tienen distinto dominio (Ecuación 4.6) y no pueden unirse en otro cuantificador que englobe a ambos.

Debido a su significado, la cardinalidad con notación (min, max) utilizará preferentemente cuantificadores absolutos, aunque también son aceptados los relativos al igual que ocurre en las restricciones de tipo de correspondencia difusa.

4.3.4.2 Análisis entre Restricción de Participación y Notación Difusa (min,max) en FuzzyEER

Por otra parte, no son excluyentes la cardinalidad con notación (min, max) y una restricción de participación difusa (apartado 4.3.2). Mientras una restricción de participación difusa establece una condición sobre las instancias de una entidad de manera global, la cardinalidad restringe la interrelación de cada instancia de forma individual con la otra entidad participante.

Por esto, las notaciones más útiles son la notación (min, max) difusa y la notación para las restricciones de participación difusa (con cuantificadores difusos relativos principalmente). La notación para las restricciones de tipo de correspondencia difusa, puede eliminarse para ser expresadas utilizando cardinalidad con notación (min,max) difusa.

En el modelo ER/EER clásico, la notación (min,max) sustituye a las otras dos notaciones para las restricciones de participación y de tipo de correspondencia, ya que si min=0 se trata de una participación parcial, y si min > 0 se trata de una participación total. Por otra parte, si max=1 la interrelación será 1:1 ó 1:N (en el lado del 1) y si max=N se tratará de una interrelación N:M ó 1:N (en el lado de N).

En cambio, en el modelo FuzzyEER con restricciones difusas la notación (min,max) añade expresividad al modelo conceptual, pero sólo puede sustituir al tipo de correspondencia difusa y, en algunos casos, a la participación. La Figura 4.34 muestra la representación de la restricción de participación "casi_todo", con la notación (min,max) igual a (0,aprox_3[0.25,0.75]) con min=0.

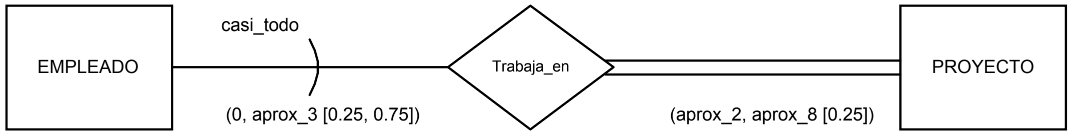


Figura 4.34: Restricción de participación y cardinalidad difusa en FuzzyEER.

A continuación se analiza el esquema en FuzzyEER de la Figura 4.34, que muestra la notación propuesta en la participación difusa con notación difusa (min,max).

Si tenemos las siguientes notaciones (min,max) en el lado de la entidad *EMPLEADO*, sus significados son los que se indican a continuación:

a) (0, aprox_3 [0.25,0.75]): Cada empleado trabaja como mínimo en cero (o ningún) proyecto, y como máximo en aprox_3[0.25,0.75].
b) (1, aprox_3[0.25,0.75]): Cada empleado trabaja como mínimo en un proyecto, y como máximo en aprox_3[0.25,0.75].
c) (casi_todo, aprox_3[0.25,0.75]): Cada empleado trabaja como mínimo en casi todos los proyectos, y como máximo en aprox_3[0.25,0.75].

El caso c) puede entrar en contradicción fácilmente, o sea, puede darse que no exista "área permitida". El cuantificador "casi_todo" para la participación es: "*casi todos los empleados están relacionados con proyectos*", no siendo lo mismo que lo representado en una notación (min,max) teniendo como min="casi_todo". Luego para la Ecuación 4.6:

- Si $\min\{\alpha:\alpha = (Q_{min})^{-1}(\gamma_1)\} > 0$ participación es total, caso b) y c).
- Si $\min\{\alpha:\alpha=(Q_{min})^{-1}(\gamma_1)\} = 0$ participación parcial se puede difuminar en la participación difusa, y queda representado en el Ejemplo 4.10 de la Figura 4.34.

De este modo, si la cardinalidad mínima es min=0 y la participación es parcial, es necesario que el cuantificador de ésta quede expresada en el modelo (como lo muestra la Figura 4.34), pero si la cardinalidad mínima es 1 implica que la participación es total, por lo tanto "no" es necesario colocar la notación del cuantificador a menos que sea una restricción acotada.

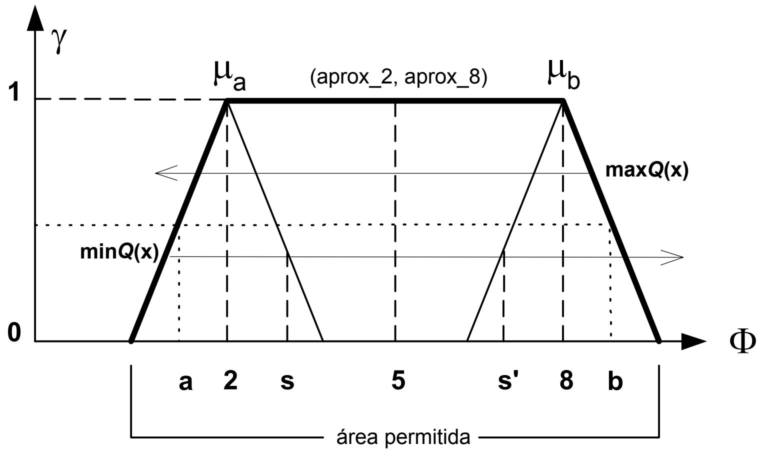


Figura 4.35: Representación de función de distribución para cuantificadores de cardinalidad (min, max).

Para la Figura 4.35, $\mu_{(a,b)}$ (x)= sup min (μ_a (s), μ_b(s') para todo s ≤ x ≤ s'). La función de distribución entre a y b está en línea más gruesa.

4.4 Jerarquías de Generalización o Especialización Difusas en FuzzyEER

Algunos autores, definen extensiones del modelo ER conocidas como EER. El modelo EER abarca todos los conceptos de modelado del ER comentados en los apartados anteriores, además, incluye los conceptos de especialización, generalización, categorías y subclases compartidas en jerarquías. Desafortunadamente, no existe una terminología estándar para estos conceptos, de modo que se utilizará los términos de uso más frecuente.

De Miguel et al. (1999) dicen que la generalización es el tipo de interrelación que existe entre un tipo de entidad y los tipos de entidad más específicos que dependen de él. En el mundo real es muy habitual la descomposición de un tipo de entidad, creándose de esta forma una jerarquía de tipos de entidad, donde se puede distinguir una *superclase* del cual dependen varias *subclases*, también algunos autores utilizan el concepto superentidad y subentidad o supertipo y subtipo.

Elmasri y Navathe (2000 y 2002) dicen que el modelo EER abarca todos los conceptos de modelo ER y además, incluye los conceptos de subclase y superclase para los conceptos relacionados de especialización y generalización, además, de los tipos de unión mediante el uso de categorías y los tipos de intersección mediante el uso de subclases compartidas. El modelo FuzzyEER utiliza, para las jerarquías, el concepto superclase y subclase.

Subclases y **superclases**: En muchos casos, un tipo de entidad tiene varias agrupaciones adicionales de entidades que son significativas y se deben representar explícitamente, por su importancia, en una aplicación de base de datos. Cada una de estas agrupaciones es una *subclase* del tipo de entidad, y cada tipo de entidad es la *superclase* de una de estas subclases. Se llama a la relación entre una superclase y cualquiera de sus subclases un vínculo superclase/subclase,. o simplemente vínculo, clase/subclase.

Empleo de subclases en el modelado de datos: Hay dos razones principales para incluir vínculos superclase/subclase en un modelo de datos. La primera es que ciertos atributos pueden aplicarse a algunas de las entidades del tipo de entidad (superclase), pero no a todas. Se define una subclase para agrupar las entidades a las que se aplican estos atributos. Los miembros de la subclase pueden compartir la mayor parte de sus atributos con los demás miembros de la superclase. La segunda razón para usar subclases es que en algunos tipos de vínculo (o interrelaciones) sólo pueden participar instancias que sean miembro de la subclase.

Tanto la especialización como la generalización están definidas por subclases y superclases que representan un conjunto de entidades, y por tanto, es un tipo de entidad; es por ello que las superclases y las subclases (al igual que los tipos de entidades) se representan como rectángulos en los diagramas ER/EER.

En la extensión del modelo FuzzyEER proponemos algunas restricciones difusas para la especialización y la generalización en los siguientes apartados, en especial para la completitud y disjunción (solapamiento). La propuesta del modelo FuzzyEER, se genera a partir de algunos ejemplos, donde se trata de utilizar restricciones difusas, considerando el caso de la especialización y no la generalización ya que este es un proceso inverso, pero sin diferenciación alguna final.

4.4.1 Restricción de Completitud Difusa en la Especialización de FuzzyEER

Elmasri y Navathe (2000) dicen que, una restricción existente entre una superclase y una subclase sobre la especialización se denomina *restricción de completitud* (completeness), la cual puede ser total o parcial. Una *restricción de completitud total en una especialización* especifica que, toda entidad de la superclase debe ser miembro de alguna subclase de la especialización, esto se representa en los diagramas ER/EER con una línea doble que conecte la superclase al círculo que separa las subclases. Se usa una línea sencilla para indicar una *restricción parcial en una especialización,* indicando que existen instancias de la superclase que no pertenezca a ninguna de las subclases.

Nuestro modelo FuzzyEER se basa en el uso de cuantificadores en una restricción completitud, expresada de la siguiente forma:

Definición 4.11: Sea E una superclase y $S_1,S_2,...,S_n$ el conjunto de n subclases que definen una **restricción de completitud difusa** cuando se utiliza un cuantificador difuso Q en la especialización. La representación se hará utilizando una *línea con un arco* que une la superclase con el vínculo de especialización, indicando a su lado el cuantificador utilizado, seguido opcionalmente, de uno o dos grados de cumplimiento entre corchetes [γ,δ] para dicho cuantificador, con el mismo significado y valor por defecto que el explicado en los apartados anteriores. La Figura 4.36 muestra la representación gráfica.

Esta restricción establece que:

$$Q(\Phi) \geq \gamma \qquad (4.21)$$

donde γ es el umbral establecido para Q y Φ es:

$$\Phi = \begin{cases} a & \text{si } Q = Q_{\text{abs}} \\ a/b & \text{si } Q = Q_{\text{rel}} \end{cases} \qquad (4.22)$$

siendo a, el número de instancias de E que pertenecen también a “alguna” subclase, y b el número de instancias totales de E.

*

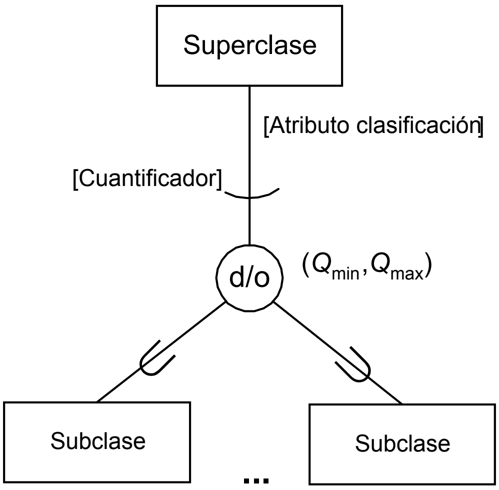


Figura 4.36: Representación de la restricción de completitud difusa con cuantificador difuso.

Observe, que este cuantificador debe ser por lo general relativo, aunque como se indicó en el caso de restricciones de participación (apartado 4.3.2) también pueden ser cuantificadores difusos absolutos.

A su vez, Elmasri y Navathe (2002) dicen que, una restricción existente entre una superclase y subclases sobre la especialización, se denomina *restricción de disyunción*, que especifica que las subclases de una especialización deben ser *disjuntas*. Esto significa que las instancias de la superclase pueden ser miembro de cuando más una de las subclases de la especialización. La representación es una “d” dentro del círculo que significa disyunción o que las subclases son conjuntos disjuntos. Por otro lado, si las subclases no son disjuntas, sus conjuntos de entidades pueden *solaparse*; esto es, que una misma instancia de la superclase puede ser miembro de más de una subclase de la especialización. Esto se indica colocando una “o” en el círculo para identificar una *restricción solapada (overlapping)*. En el apartado 4.4.3 se exponen restricciones disjuntas y solapadas difusas.

Otra forma de restricción en una especialización, puede ser definida por un atributo, que indica la restricción de disyunción, si el atributo con que se define el predicado de pertenencia es monovaluado (univaluado).

Ahora bien si, las restricciones de completitud y disyunción son independientes, tenemos cuatro tipos de especialización: disjunta total, disjunta parcial, solapada total, solapada parcial.

En el caso de nuestra propuesta hay dos alternativas: disjunta con completitud difusa, solapada con completitud difusa.

Ejemplo 4.11: Supongamos una entidad *EMPLEADO* que es una superclase con dos subclases *EMPLEADO FIJO* y *EMPLEADO TEMPORAL*, definidas por el atributo discriminante `Tipo_Contrato`. Un cuantificador difuso relativo como "casi todo" indicará que: "*Casi todos los empleados deben ser fijos o temporales, pero pueden existir otros tipos de empleados (en prácticas, becarios, etc.)*". Véase Figura 4.37.

En el ejemplo de la Figura 4.37, la especialización es disjunta (con una "d" en el círculo) ya que no puede haber un empleado con dos tipos de contrato distintos. Sin embargo, las restricciones de completitud difusas también pueden aplicarse a especializaciones solapadas (con una "o" en el círculo). Esto se muestra en el Ejemplo 4.12.

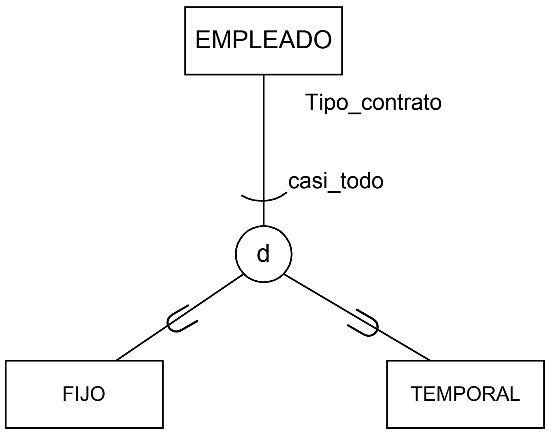


Figura 4.37: Ejemplo de restricción de completitud difusa definido por un atributo `Tipo_contrato` en una especialización disjunta y cuantificador "casi_todo".

Ejemplo 4.12: Supongamos una entidad *EMPLEADO* que es una superclase con varias subclases definidas por el atributo `Tipo_habilidad` de: *PROGRAMADOR_GESTIÓN*, *PROGRAMADOR_SISTEMAS*, *PROGRAMADOR_INTERNET*, *DISEÑADOR_GRÁFICO*, *CONTABLE, ANALISTA*. Un cuantificador difuso relativo como "casi_todo" indicará que: "*Casi todos los empleados deben tener alguna o algunas de las habilidades expresadas en las subentidades*". Véase Figura 4.38.

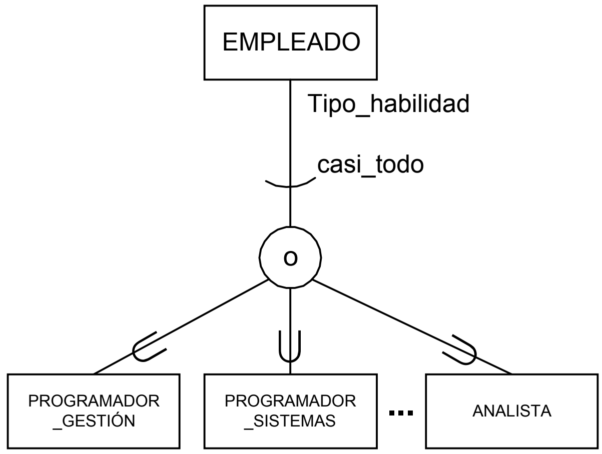


Figura 4.38: Restricción de completitud difusa en especialización solapada con cuantificador difuso casi_todo.

4.4.2 Restricción de Cardinalidad Usando Notación Difusa (min,max) en Especializaciones Solapadas en FuzzyEER

En el caso de una especialización con solapamiento, se puede además establecer de forma flexible el número mínimo y máximo de subclases a las que puede pertenecer cada miembro de la superclase (Varas et al. 1998). Esto puede fácilmente expresarse utilizando la notación (min, max) difusa que se colocará junto al vírculo que contiene la letra "o" (*overlapping*).

Definición 4.12: Sea E una superclase y S_1, S_2,...,S_b el conjunto de b subclases en una especialización solapada. Supongamos que E tiene n instancias: $e_1,e_2,...,e_n$. Una **restricción de cardinalidad difusa con notación (min,max) en una especialización solapada**, denotada con dos cuantificadores difusos, ($Q_{\min}$,$Q_{\max}$), situados junto al vínculo de la especialización que tiene la "o". Esta restricción establece que:

$$\lambda_{\min} \leq \Phi_{\min,i} \quad \wedge \quad \Phi_{\max,i} \leq \lambda_{\max}, \quad \forall\, i = 1,2,...,n \tag{4.23}$$

donde se tiene que:

$$\Phi_{\min,i} = \begin{cases} a_i & \text{si } Q_{\min} \text{ es absoluto} \\ a_i/b & \text{si } Q_{\min} \text{ es relativo} \end{cases}$$
$$\Phi_{\max,i} = \begin{cases} a_i & \text{si } Q_{\max} \text{ es absoluto} \\ a_i/b & \text{si } Q_{\max} \text{ es relativo} \end{cases} \tag{4.24}$$

tal que:

- a_i es el número de subclases a las que pertenece la instancia e_i de E.
- b es el número de subclases de E.
- $\lambda_{\min} = \min \{\alpha : \alpha = (Q_{\min})^{-1}(\gamma_{\min})\}$.
- $\lambda_{\max} = \max \{\beta : \beta = (Q_{\max})^{-1}(\gamma_{\max})\}$.
- $\gamma_{\min}$ y $\gamma_{\max}$ son los umbrales mínimos establecidos para $Q_{\min}$ y $Q_{\max}$ respectivamente.

*

Para calcular el "área de aviso", o cuando debe enviar un "mensaje" impuesta por la restricción, se usan los umbrales $\delta_{\min}$ y/o $\delta_{\max}$ en $Q_{\min}$ y $Q_{\max}$ respectivamente, se tiene que cumplir la Ecuación 4.24, y además cumplir la siguiente ecuación:

$$\lambda'_{\min} > \Phi_{\min,i} \quad \vee \quad \lambda'_{\max} < \Phi_{\max,i}\,, \quad \forall\ i = 1,2,\ldots,n \tag{4.25}$$

Además, aplicando ley de De Morgan, debe cumplir la siguiente ecuación:

$$\lambda'_{\min} \leq \Phi_{\min,i} \quad \wedge \quad \lambda'_{\max} \geq \Phi_{\max,i}\,, \quad \forall\ i = 1,2,\ldots,n$$

donde:

$$\lambda'_{\min} = \min \{\alpha : \alpha = (Q_{\min})^{-1}(\delta_{\min})\} \tag{4.26}$$
$$\lambda'_{\max} = \max \{\beta : \beta = (Q_{\max})^{-1}(\delta_{\max})\}$$

Ejemplo 4.13: Considerando el Ejemplo 4.12, se puede establecer una restricción difusa de cardinalidad sobre la especialización solapada, tal como: (aprox_2, aprox_la_mitad). Esto establece la restricción de que cada empleado deba figurar en un mínimo de *aproximadamente 2* tipos de habilidades y como máximo, en *aproximadamente la mitad* de los tipos de habilidades existentes. Véase Figura 4.39.

Obsérvese que los cuantificadores difusos de la cardinalidad (min, max), pueden ser de cualquier tipo (absolutos o relativos). Por supuesto, en este caso también cada cuantificador utilizado puede ir seguido, opcionalmente, de uno o dos grados de cumplimiento entre corchetes [γ,δ] para dicho cuantificador, con el mismo significado y valor por defecto que el explicado anteriormente.

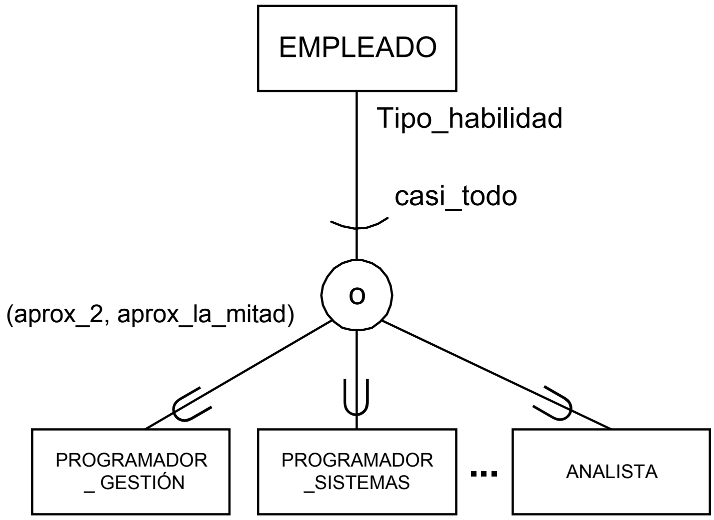


Figura 4.39: Restricción de completitud difusa y notación (min,max) para especialización solapada.

4.4.3 Restricción Difusas Disjuntas o Solapadas sobre Especializaciones en FuzzyEER

En las especializaciones puede ser interesante incluir una extensión para expresar el caso en el que las instancias de la superclase pertenezcan a cada subclase, usando etiquetas lingüísticas, tales como “muchos” o “pocos”, de forma más simple, dando un grado de pertenencia en un intervalo [0,1]. En tanto, cada subclase es considerada como un *conjunto difuso de la superclase*, asumiendo que los elementos no están claramente definidos y pueden pertenecer al conjunto difuso con un cierto grado de pertenencia en el intervalo [0,1]. Esto consiste en ver las subclases como entidades difusas. Así, el conjunto de los Analistas del ejemplo anterior, es un conjunto difuso, ya que un empleado puede pertenecer a este conjunto con cierto grado, según su *tipo de habilidad*. Por otro lado, esto complementa lo expuesto por Marín et al. (2000).

Esta extensión puede ser aplicada en las restricciones de especialización disjuntas o solapadas considerando la siguiente definición:

Definición 4.13: Sea E una superclase en una especialización con n subclases. Se dice que esa **especialización es disjunta difusa** cuando al menos una de las subclases es una entidad difusa (definida en apartado 4.2.1), y para cualquier instancia $e \in E$ existe un máximo de una subclase S_i con $i \in \{1,2,...,n\}$ tal que:

$$\mu_i(e) > 0 \tag{4.27}$$

donde $\mu_i(e)$ es el grado de pertenencia de la instancia e a la subclase S_i. O sea, como toda especialización disjunta debe cumplirse que: $S_1 \cap S_2 \cap ... \cap S_n = \emptyset$.

Esta restricción se representa usando “fd” dentro del círculo de la especialización, o vínculo que une la superclase a las subclases.

*

Por supuesto, si S_i es una subclase no difusa, entonces: $\mu_i(e)=1$ si e pertenece a S_i, y $\mu_i(e)=0$ si $e \notin$ a S_i.

Definición 4.14: Sea E una superclase en una especialización con n subclases S_1, S_2,...,S_n. Se dice que esa **especialización solapada difusa** cuando al menos una o más de las subclases son entidades difusas, y para cualquier instancia e de E puede existir una o más subclases S_i con $i \in \{1,2,...,n\}$ tal que: $\mu_i(e) > 0$ donde $\mu_i(e)$ es el grado de pertenencia de la instancia e a la subclase S_i.

Esta característica se representa usando “fo” dentro del círculo de la especialización, o vínculo que une la superclase a las subclases. Véase Figura 4.40.

*

La Definición 4.14 es más flexible que la Definición 4.13, porque el hecho que una instancia e perteneciente a E, puede significar que pertenece a varias subclases con distintos grados. La notación grafica propuesta para las subclase difusas será un rectángulo con línea cortada (o punteada). Véase Figura 4.40.

Es bueno señalar, que esto no implica que toda subclase sea un conjunto difuso. Una especialización difusa, ya sea disjunta o solapada, puede tener subclases con o sin la restricción difusa.

La restricción de especialización difusa de la Definición 4.12, tiene dos puntos de vistas:

a) **Desde el punto de vista de la subclase**: Las subclases son conjuntos difusos y sus dominios respectivos son todas las instancias de la superclase, por ejemplo, cada instancia de la superclase tiene un grado de pertenencia a cada subclase, incluido el valor cero. Véase Tabla 4.3. Supongamos S es una subclase de E, entonces el conjunto difuso es representado por:

$$\{\ \mu_S(e_1)/\ e_{1,},\ \mu_S(e_2)/\ e_2\ ,\ \ldots,\ \mu_S(e_m)/\ e_m\ \} \tag{4.28}$$

donde $e_i,\ i=1,\ldots,$ n son todas las instancia de la superclase de E, y $\mu_S(e_i)$ es el grado de pertenencia de e_i a la subclase S.

b) **Desde el punto de vista de las instancias de la superclase**: Cada instancia de la superclase puede pertenecer a una o más subclases y este grado de pertenencia es medido con un conjunto difuso asociado, siendo el respectivo dominio de este conjunto difuso, el conjunto de todos los nombres de la subclase. Sea S_j, con j=1,…, n subclases en E, entonces el conjunto difuso de las instancias de e_i es:

$$\{\mu_{S1}(e_1)/\ S_{1,},\ \mu_{S2}(e_i)/\ S_2\ ,\ \ldots,\ \mu_{Sn}(e_i)/\ S_n\ \} \tag{4.29}$$

donde $\mu_{Sj}(e_i)$ con j=1,… ,n, es el grado de pertenencia de e_i a la subclase S_j. Nótese que en la especialización disjunta el número de subclases para una instancia de la superclase es 1.

Instancias/subclases	S_1	S_2	…	S_n
e_1	$\mu_{S1}(e_1)$	$\mu_{S2}(e_1)$	…	$\mu_{Sn}(e_1)$
e_2	$\mu_{S1}(e_2)$	$\mu_{S2}(e_2)$	…	$\mu_{Sn}(e_2)$
….	…	…	…	…
e_i	$\mu_{S1}(e_i)$	$\mu_{S2}(e_i)$	…	$\mu_{Sn}(e_i)$
…	…	…	…	…
e_m	$\mu_{S1}(e_m)$	$\mu_{S2}(e_m)$	…	$\mu_{Sn}(e_m)$

Tabla 4.3: Representación de conjuntos difusos sobre la especialización con *n* subclases y *m* instancias de superclases.

Obsérvese, que tanto el caso a) como el caso b) trabajan con conjuntos difusos con diferentes dominios discretos. En efecto, según Tabla 4.3, el caso a) es representado por el conjunto difuso dado por las columnas (instancias) y las filas (subclases), dan el caso b).

Ejemplo 4.14: La Figura 4.40, indica que nuestro esquema conceptual puede ser usado para representar la restricción solapada difusa de los empleados respecto a las subclases de sus tipos de habilidades. Esto es planteado *desde el punto de vista de la subclase,* donde los *programadores de sistemas* son un conjunto difuso (por ejemplo, un empleado puede pertenecer a este conjunto con un cierto grado de pertenencia). Sin embargo, el conjunto de *contables* no está definido, en este caso, como un conjunto difuso (un empleado puede o no pertenecer a este conjunto). En cambio, *desde el punto de vista de las instancias de la superclase*, un empleado en particular tiene un conjunto difuso definido sobre las subclases. Por ejemplo: un empleado que es un experto en *programación de gestión*, puede también tener un tipo de habilidad en otro tipo de *aplicaciones* y, por ejemplo, menos tipo de habilidad como *analista.* Esto podría ser representado en una base de datos como el conjunto difuso siguiente: {1/programador_gestión, 0.8/programador_sistemas, 0.3/analista}. Note que el dominio respectivo es el nombre de todas las subclases y que ese tipo de conjunto difuso puede representarse con un atributo difuso Tipo 3 o 4 (véase apartado 3.2.4), según exista o no la relación de similitud entre cada una de las subclases.

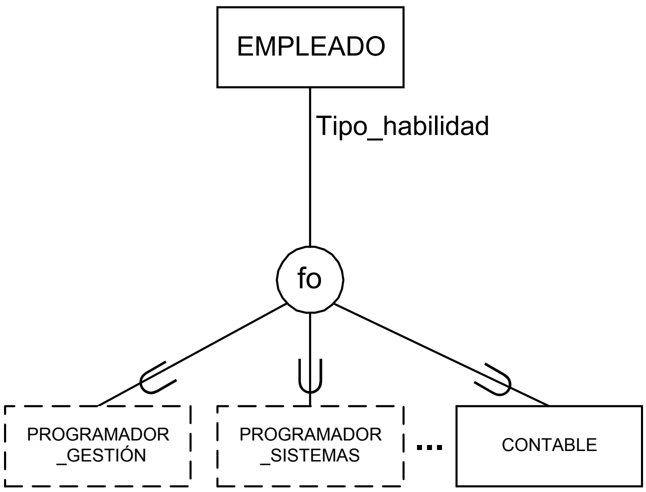


Figura 4.40: Ejemplo de especialización difusa solapada con algunas subclases difusas.

Este tipo de representación de restricciones en un modelo conceptual permitirá obtener respuesta de la base de datos del tipo: "*Seleccione el nombre de los <u>mejores</u> manejadores de aplicaciones entre aquellos que no son asignados a <u>muchos</u> proyectos y quienes son al <u>menos</u> analistas regulares*".

Está restricción no evita el uso de restricciones difusas de completitud o cardinalidad como las explicadas anteriormente. Sin embargo, cuando estos tipos de restricciones están mezcladas con restricciones sobre una especialización disjunta difusa o solapada difusa, deben estudiarse, cómo definir un método para una implementación en un SGBD, que asegure por completo el cumplimiento de estas restricciones, ya que, al ser la pertenencia de cada subclase difusa puede ser complejo contestar a cuántas subclases pertenece una determinada instancia. Un método simple y efectivo es usar el concepto de cardinalidad de un conjunto, esto se explica en el próximo apartado.

4.4.3.1 Método de Validación de una Restricción de Completitud en una Especialización Disjunta o Solapada Difusa: Semántica de este Caso Mixto

Si una **restricción de completitud difusa existe**, y se puede validar mediante un cálculo el hecho que cada instancia de la superclase pertenece a "alguna" subclase. Por ejemplo, para decidir si "*casi todas* las instancias de la superclase pertenecen a alguna subclase", necesitamos saber qué instancias pertenecen a alguna subclase, (valor *a* de la Definición 4.11). El problema es que esa pertenencia es ahora difusa, dada por un grado. Nosotros proponemos que este grado de pertenencia a "alguna" subclase pueda ser calculado de dos formas, considerando el conjunto difuso formado d*esde el punto de vista de las instancias de la superclase*, planteado anteriormente:

a) Usando el máximo grado de pertenencia de esta instancia a cualquier subclase, es decir la *altura* (altura se define en el apartado 3.2.6) de ese conjunto difuso. Tabla 4.4.

b) Usando la cardinalidad de ese conjunto difuso, o sea, sumar todos los grados de pertenencia, o usar medidas generalizadas, como la energía de ese conjunto difuso, Tabla 4.4.

Subclases/Max-suma de grado	Máximo grado	Suma de grado
S_n	$Max\ (\mu_{Sn}(e_m))$	$\Sigma\mu_{Sn}(e_m)$
S	Hgt	*Card*

Tabla 4.4: Validación de restricción de completitud en una especialización difusa.

En la Tabla 4.4, *n* es el número total de subclases y *m* es el número total de instancias, siendo los dominios de las subclases de la Tabla 4.4 discretos (subclases), por otro lado, se puede fijar un límite mínimo λ (umbral) para decidir si una instancia pertenece a alguna clase, de forma que las instancias que pertenezcan a alguna subclase con un grado insignificante, no sean tomadas en cuenta al evaluar esta restricción de completitud.

Así, en este caso, para calcular el valor de *a* de la Definición 4.11, tenemos que contar cuantas instancias de *E* pertenecen *difusamente* a *alguna* subclase.

Esa pertenencia difusa está resuelta por las dos opciones anteriores. Ahora el problema es contar esos elementos. Para ello proponemos las siguientes opciones, teniendo en cuenta que *b* es el número de instancias de la superclase *E* y *a* es representado por:

$$a = \sum_{i=1,\ldots,b} \alpha_i$$

la definición de α_i esta dada por las siguientes cuatro opciones:

a) **Opción 1**:

$$\alpha_i = \begin{cases} \alpha_i = \mathrm{Hgt}(I_i) \text{ si } \mathrm{Hgt}(I_i) \geq \lambda \\ 0 \text{ en otro caso} \end{cases}$$

Donde I_i es el conjunto difuso formado para la instancia e_i perteneciente a *E* desde el puno de vista de las instancias y Hgt representa la función ALTURA (definida en 3.2.6) de un conjunto difuso. El valor λ es el umbral mínimo para despreciar las instancias que pertenecen con poco grado.

b) **Opción 2**:

$$\alpha_i = \begin{cases} \alpha_i = 1 \text{ si } \min\{1, \mathrm{Card}(I_i)\} \text{ siendo } \mathrm{Card}(I_i) \geq \lambda \\ 0 \text{ en otro caso} \end{cases}$$

Donde Card es la suma de todos los grados de pertenencia de las instancias e_i en las distintas subclases.

c) **Opción 3**:

$$\alpha_i = \begin{cases} 1 \text{ si } \mathrm{Hgt}(I_i) \geq \lambda \\ 0 \text{ en otro caso} \end{cases}$$

d) **Opción 4**:

$$\alpha_i = \begin{cases} 1 \text{ si } \mathrm{Card}(I_i) \geq \lambda \\ 0 \text{ en otro caso} \end{cases}$$

Por otro lado, para todos los valores que correspondan a un 0, en las cuatro opciones, no se considera aquellas instancias que pertenezcan con “poco” grado a “alguna” subclase.

Por supuesto, existen en la literatura más métodos para calcular este tipo de casos, pero las cuatro opciones propuestas son suficientemente eficientes y abarcan un amplio espectro de resultados. O sea, podemos ordenar las opciones según el resultado de su conteo, de menor a mayor, con λ fijo:

$$\text{Opción 1} \leq \{ \text{Opción 2, Opción 3}\} \leq \text{Opción 4}$$

Las opciones 2 y 3 no pueden ordenarse, porque aunque siempre $\mathrm{Card}(I_i) \geq \mathrm{Hgt}(I_i)$, la Opción 3 “suma 1” (si la ALTURA supera el umbral λ), mientras que la Opción 2 “suma un valor que es como máximo 1” (si la cardinalidad supera el umbral λ).

4.4.3.2 Método de Validación de Restricción con la Notación (min,max) Difusa en una Especialización Disjunta o Solapada Difusa: Semántica de este Caso Mixto

Si una **Notación (min,max) difusa existe**, entonces para validar se debe calcular el número de subclases a las que pertenece cada instancia e_i con i=1,2,...,n de la superclase valor a_i de la Definición 4.12. Por ejemplo, para decidir si el número de subclases de una instancia de una superclase está entre “aproximadamente 2” y “aproximadamente la mitad” de las subclases (usando la notación (min, max)). Ahora bien, calcular ese número no es simple, porque la pertenencia es ahora difusa. Se propone que este problema puede ser resuelto de dos maneras:

a) Sumar los distintos grados, usando la cardinalidad del conjunto difuso formado *desde el punto de vista de las instancias de la superclase* o usar medidas generalizadas, como la energía de ese conjunto difuso $a_i = \mathrm{Card}\ (I_i)$.

b) Contar el número de subclases con un grado de pertenencia mayor a un valor mínimo (usualmente 0 igual a:

$$a_i = \sum_{i=1,\ldots,b} \alpha j \text{ con } j=1,\ldots, n \text{ y,}$$

$$\alpha_j = \begin{cases} 1 \text{ si } \mu_j(e_i) \geq \lambda \\ 0 \text{ en otro caso} \end{cases}$$

donde, b es el número de subclases y μ_j (e_i) es el grado de pertenencia de la instancia de la superclase e_i a la subclase s_j.

Una vez validado este número, el sistema debe chequear si el número satisface la restricción de cardinalidad difusa.

4.4.4 Atributos Difusos que Definen Especializaciones en FuzzyEER

Elmasri y Navathe (2002) dicen que, en algunas especializaciones se puede determinar con exactitud las entidades que se convertirán en miembro de cada subclase, para lo cual se específica una condición en término del valor de algún o algunos atributos de la superclase. Tales subclases se llaman *subclases definidas por predicado* (o definidas por condición). Para representar las subclases definidas por predicado se escribe la condición del predicado junto a la línea que conecta la subclase a su superclase. Si la condición de pertenencia de todas las subclases de una especialización es definida en términos del mismo atributo de la superclase, se dice que la especialización misma es una especialización *definida por el atributo*. La especialización definida por un atributo se representa colocando el nombre del atributo de definición junto a la línea que va unida al círculo o vínculo de la superclase. Cuando no tenemos una condición que determine la pertenencia, se dice que la especialización está *definida por el usuario*. La pertenencia a tales subclases la determinan los usuarios de la base de datos cuando aplican la operación de añadir una entidad a la subclase; así pues, el usuario especifica individualmente la pertenencia para cada instancia; no especifica una condición que se pueda evaluar automáticamente.

Otro tipo de aprovechamiento de los conjuntos difusos en la especialización solapada o disjunta en algunas subclases, es la que define una *especialización por un tipo atributo difuso*, la cual en general, no tienen un orden en el dominio requerido (atributo difuso Tipo 3 y Tipo 4), pudiendo ser otro tipo de atributo difuso (Tipo 1 o Tipo 2). Bajo este contexto se tienen la siguiente definición:

Definición 4.15: Si un atributo difuso define una semejanza en una especialización se considera como **Tipos de atributos difusos en especializaciones.** Su notación considera una línea quebrada en forma de ángulo entre la superclase y el vínculo de la especialización, a su lado el Tipo de atributo difuso que genera la condición de la subclase, sea esta disjunta o solapada. Esta notación se muestra en la Figura 4.41.

*

Este tipo de especialización pueden tener tantas modalidades como tipos de atributos difusos (T1, T2, T3 y T4).

Cada subclase corresponde a un valor de las etiquetas definidas para ese atributo. Este caso representa que las subclases son conjuntos difusos, asociados a etiquetas lingüísticas, por lo que las subclases serán entidades difusas. Pueden definirse las subclases como entidades no difusas, pero restringirá el valor del atributo a aquellos valores que alcancen grado 1 si son comparados con las etiquetas definidas, lo cual puede ser demasiado restrictivo.

Según esto, la clasificación de una instancia e de la superclase E, se hace automáticamente según el tipo de especialización, como por ejemplo:

1. **Disjunta difusa (fd)**: Si la especialización es disjunta se asigna a la instancia e_i sólo a aquella subclase para la que existe mayor similitud entre la etiqueta asociada a la subclase y el valor del atributo difuso para la instancia e_i.

2. **Solapada difusa (fo)**: la instancia e_i se asigna a todas las subclases para las que tenga un grado de similitud mayor a 0 entre la etiqueta asociada a la subclase y el valor del atributo difuso en e_i.

3. **Disjunta no difusa (d)**: Se supone que cada instancia e_i perteneciente a E, sólo puede pertenecer a una subclase y esa pertenencia será total (con grado 1).

4. **Solapada no difusa (o)**: Se supone que cada instancia e_i perteneciente a E, puede pertenecer a varias subclases y esa pertenencia será total (con grado 1 en todas ellas).

Cada uno de los casos 1, 2, 3 y 4, se pueden utilizar para los cuatro tipos de atributos difusos lo que da un total de 16 posibilidades.

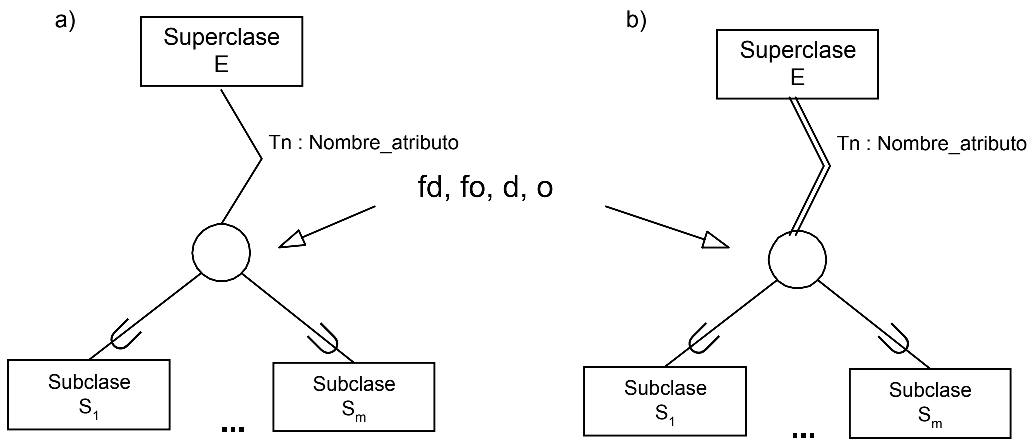


Figura 4.41: Especialización definida por un tipo de atributo difuso con n=1,2,3 y 4, y "m" número de subclases. a) Participación parcial, b) Participación total.

Lo interesante de este tipo de atributo difuso es que puede dar algún tipo de respuesta cuando ésta pudiera ser vacía, considerando que hay un grado de semejanza de lo que se desea buscar. En este caso, la definición es independiente de todas las restricciones, sean difusas o *crisp* en especializaciones disjuntas o solapadas.

El siguiente ejemplo muestra subclases definidas en una especialización disjunta no difusa, con el interés de que cada par de subclases tienen un grado de similitud entre ellas. Esta propiedad es usada para compararlos y buscar la instancia más semejante que se acomode a un requerimiento.

Ejemplo 4.15: Una superclase *EMPLEADO* con las subclases *PROFESOR*, *INVESTIGADOR*,..., I*NGENIERO*, las cuales tienen sus propios atributos, siendo en este caso una especialización disjunta. El atributo difuso `T3:Categoría`, porque si una empresa está buscando un ingeniero, esta empresa está posiblemente interesada en un investigador, dado que estos dos elementos se parecen en alguna medida dependiendo de la especialidad. Este tipo de similitud o semejanza puede ser tomado en cuenta a la hora de mostrar a una empresa toda las propiedades que pueden satisfacer sus necesidades. La representación de este modelo está en la Figura 4.42.

Como la especialización es no difusa, se supone que los empleados sólo pueden pertenecer a una categoría y esa pertenencia será total.

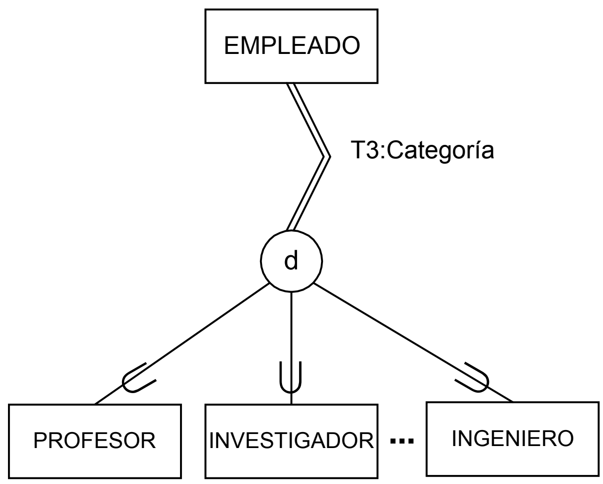


Figura 4.42: Especialización disjunta por atributo difuso `T3:Categoría` y restricción de participación total.

Por otra parte, la Figura 4.48 del Ejemplo 4.18, incluye otros dos ejemplos de especializaciones solapadas difusas para atributo Tipo 3. En el primero de ellos, se muestra una especialización de participación total de restricciones (línea doble), la cual establece que *todos* los empleados deben pertenecer a una o más categoría (profesor, investigador...) con distintos grados, considerando que *categoría* es un atributo difuso Tipo 3 (este ejemplo es distinto al mostrado en la Figura 4.42, ya que su especialización es solapada difusa "fo"). La segunda especialización es una restricción de participación difusa parcial (línea simple), considerando un cuantificador difuso "casi_todos", en el arco etiquetado representando la restricción de completitud difusa que indica que: "*Casi todos los investigadores deben pertenecer a una o más líneas de investigación*". Además, *área de investigación* es un atributo difuso Tipo 3, debido a que pueden existir similitudes entre las distintas áreas que queremos considerar para su tratamiento.

4.4.5 Agregación Difusa

De Miguel et al. (1999) dicen que una agregación permite representar tipos de entidades compuestas que se obtienen por la unión de otras más simples. Al tipo compuesto se refiere como el *todo*, mientras que los componentes son las *partes*. Estos autores consideran dos tipos de agregación: compuesto/componente y miembro/colección. Para la agregación difusa que se quiere representar aquí, se utiliza la primera de ellas, la cual es definida como una abstracción que permite representar que un *todo* se obtiene por la unión de diversas *partes*, que pueden ser tipos de objetos distintos y que desempeñan diferentes roles en la agregación. La representación gráfica de la agregación de compuesto/componente, es un rombo junto a la entidad que forma el todo, y con líneas que unen las distintas entidades que forman las partes. Otro caso es la abstracción de agregación presentada en el apartado 3.1, donde, el compuesto/componente está formado por entidad/atributo.

Definición 4.16: Se define una **agregación difusa** de dos formas:

1. **Agregación de tipos de entidades**, que permite representar que un *todo* se obtiene por la unión de las diversas *partes* que pueden ser tipos de objetos distintos, que pertenecen al todo con un *cierto grado*, y que desempeñan diferentes roles en la agregación. Siguiendo la notación de De Miguel et al. (1999), la representación grafica de la agregación difusa de entidades es un rombo, que aquí escribimos con línea discontinua, unido a la entidad que forma el todo, y con líneas discontinuas se unen a las distintas entidades que forman las partes. Junto a dichas líneas se define el grado que la asocia a la agregación y su significado. Véase Figura 4.43 a).

2. **Agregación de atributos**, que permite representar una entidad y un conjunto de atributos que representan sus partes componentes. La representación gráfica de la agregación difusa de atributos es, un círculo de línea discontinua, junto a esté, el grado difuso que el atributo tiene para la entidad a la que pertenece (note que el grado difuso puede ser de cualquiera de los tratados en el apartado 4.1.2) . Véase Figura 4.43 b).

*

La representación de agregación de entidad clásica incluye la notación (min,max), pues en la representación de la agregación de entidad difusa, esta puede ser incorporada en el mismo sentido como fue explicada en los apartados anteriores, sólo habría que incluir junto a la línea discontinua la notación (Q_{min},Q_{max}).

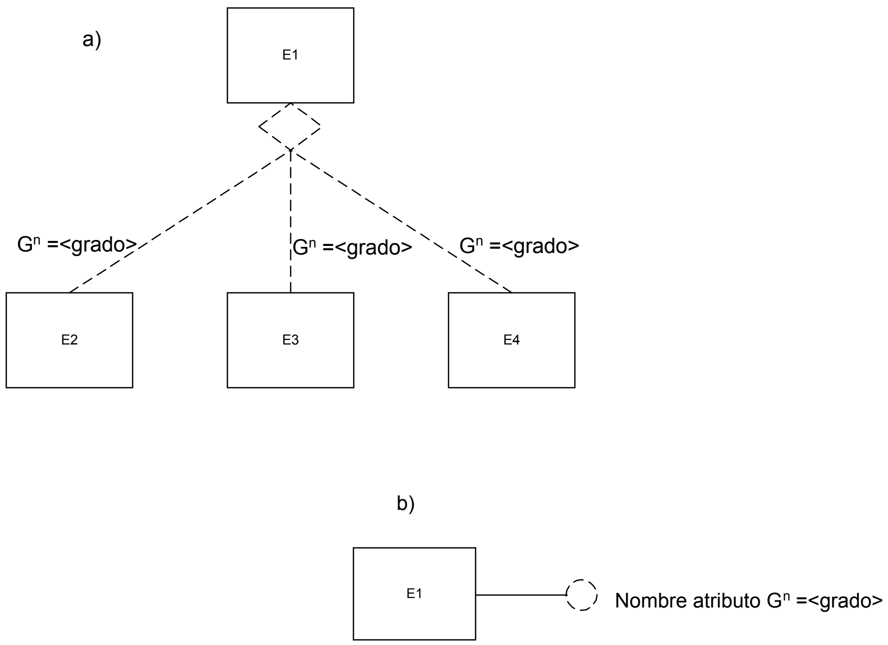


Figura 4.43: Notación de agregación en FuzzyEER. a) Agregación difusa de entidades, b) Agregación difusa de atributos.

Considérese que en la Definición 4.16, no son excluyentes la agregación clásica y la agregación difusa, ya que normalmente algunas de las entidades de la agregación compuesto/componente pueden representar ambas características. Lo mismo ocurre en la agregación de atributos, por lo general, sólo algunos atributos serán de agregación difusa a la entidad.

Ejemplo 4.16: Un *COCHE* es la unión de diversos componentes, por ejemplo Chasis, Interior y radio. También, *COCHE* se compone de la agregación de atributos {`Identificación`, `Color`,`Año`,`Dueño`}. Consideraremos que el atributo `Radio` tiene un grado de importancia 0.5, y el atributo `Interior` un grado de importancia 0.8 para la entidad *COCHE*. La Figura 4.44 a) muestra la gráfica de este ejemplo. Por otro lado tendremos que, el atributo `Identificación` tiene un grado de importancia 1, y el atributo `Color` un grado de importancia 0.6, el atributo `Año` un grado de importancia 0.8, por último el atributo `Dueño` tiene un grado de importancia 0.7, todos ellos para la entidad *COCHE*. La Figura 4.44 b) muestra su representación.

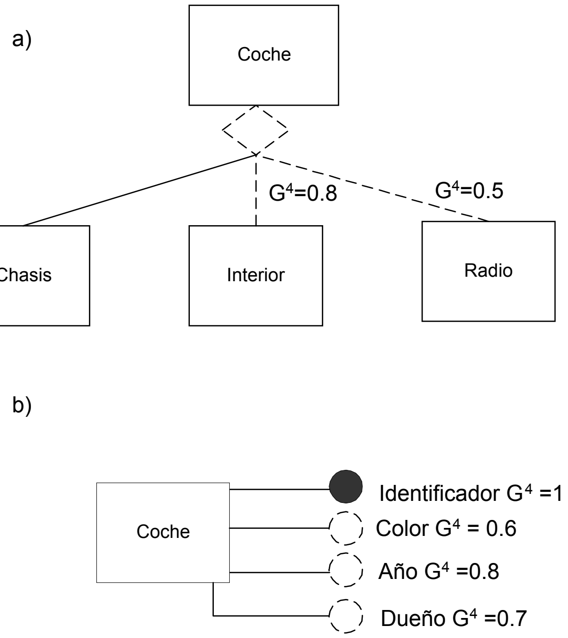


Figura 4.44: Ejemplo de agregaciones difusas. a) Agregación de entidades, b) Agregación de atributos.

Nótese que el ejemplo representado por la Figura 4.44 a) es también tratado por Ma et al. (2001), en nuestra opinión, la notación propuesta en el modelo FuzzyEER es más representativa que la de este autor. También, el tipo de ejemplo de la Figura 4.44 b) es tratado en Chen (1998), pero estos autores no usa el concepto agregación, sino que sólo como atributos que componen la entidad, la notación del modelo FuzzyEER queda más representativa de este caso.

4.4.6 Grado difuso en las Especializaciones

En el mismo sentido como fueron incorporados los grados difusos en la agregación en el apartado 4.4.4, se puede utilizar como grados difusos en las especializaciones, es así como tenemos:

Definición 4.17: Se define un **grado difuso en las especializaciones difusas** de dos formas:

1. **Grado de especialización en sus subclases**, esto es dar un grado a cada subclase. La representación gráfica es G^n=<grado> junto a la línea que une a cada subclase con el vínculo de la especialización.

2. **Grado de la especialización**, esto es dar un mismo grado a todas las subclases de esa especialización. La representación gráfica es G^n=<grado> junto al vínculo de la especialización.

Para ambos casos se debe considerar los tipos de grados expuestos en el apartado 3.2.2 como G^n, En la notación G^n=<grado>, <grado> indicará el valor del grado que esta representando en un intervalo entre [0,1]. La Figura 4.45 muestra para estos casos la notación en el modelo FuzzyEER.

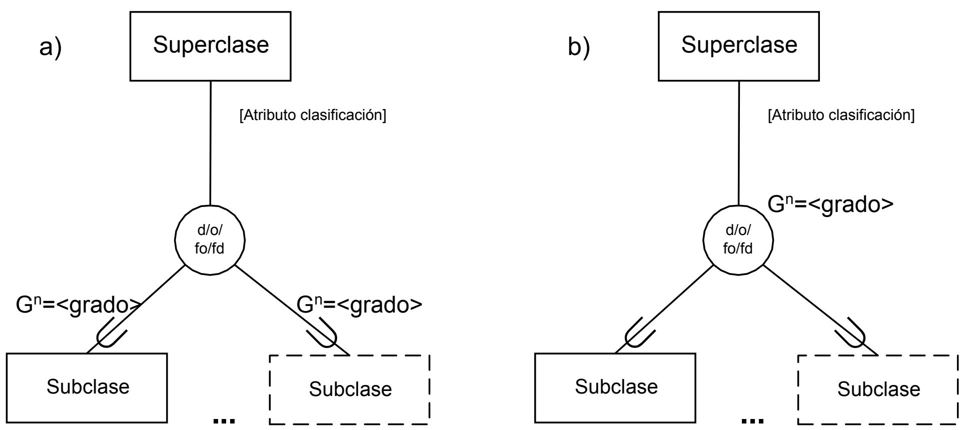


Figura 4.45: Notación FuzzyEER. a) Grado en cada subclase de la especialización, y b) Grado en la especialización.

Considérese que este tipo de concepto y representación (algún tipo de grado) se puede usar para todas las jerarquías de especialización, generalización y agregación de igual forma como fueron definidas en esta tesis.

4.5 Restricciones Difusas sobre los Tipos de Unión o Intersección FuzzyEER

Todos los vínculos superclases/subclases anteriormente vistos tienen una sola superclase. Sin embargo, en algunos casos es necesario modelar un solo vínculo superclase/subclase con más de una superclase, donde las superclases representan diferentes tipos de entidades. Este considera dos posibilidades, la subclase es una *categoría* (Unión) *o subclase compartida* (Intersección) (Elmasri y Navathe, 2002).

En la representación del modelo FuzzyEER se conservan todas las restricciones difusas propuestas hasta el momento con el uso de cuantificadores difusos. A continuación son definidas las *categorías* y las *subclases compartidas*.

4.5.1 Restricciones Difusas sobre los Tipos de Unión en Categorías: Participación y Completitud en FuzzyEER

Una categoría tiene dos o más superclases que pueden representar distintos tipos de entidades, mientras que otro vínculo superclase/subclase siempre tiene una sola superclase. Hay que considerar que, una categoría es un subconjunto de la *unión* de sus superclases. A su vez, una categoría puede ser parcial o total, se identifican con una línea sencilla o una doble respectivamente que conecta la categoría al vínculo. Cada miembro de la subclase categoría debe ser un miembro de una de las superclases si es total. En las categorías parciales se da el caso de que instancias de superclases no pertenezcan a la categoría, porque las categorías son un subconjunto de todas las uniones de las superclases.

Las jerarquías en este caso, son representadas por el símbolo unión dentro del círculo del vínculo que une a la subclase compartida y las superclases con una línea o dos según corresponda (parcial o total).

En este tipo de jerarquías es posible aplicar restricciones difusas de dos maneras: restricción de participación difusa en una o más superclases, y restricción de completitud difusa en una categoría de un tipo de unión.

Definición 4.18: Sea C una categoría (o subclase) de un tipo de unión con n superclases E_i, con i=1,2,...,n, se define **una restricción de participación difusa en una o más superclases**, y se representa con un arco que cruza las líneas que unen las superclases seleccionadas con el círculo de la unión o vínculo con la subclase. El arco debe estar etiquetado con uno o más cuantificadores difusos, o bien, con la notación (min,max).

Las superclases seleccionadas E_j, $\forall j \in$ J con J $\subseteq \{1,2,...,n\}$ y $j<i$, son aquellas a las que le afecta la restricción (unión). Denotaremos como ∇ la unión de todas esas superclases seleccionadas existiendo dos casos:

$$\nabla = \cup_{j\in \mathrm{J}}\, E_j$$

1. Si el arco está etiquetado con un cuantificador Q, esta restricción establece que:

$$Q(\Phi) \geq \gamma \tag{4.30}$$

donde γ es el umbral mínimo establecido para Q y:

$$\Phi = \begin{cases} a & \text{si } Q \text{ es absoluto} \\ a/b & \text{si } Q \text{ es relativo} \end{cases} \tag{4.31}$$

Siendo a, el número de instancias de ∇ que pertenecen a C, y b es el número total de instancias de ∇.

2. Si el arco está etiquetado dos cuantificadores difusos usando la notación (min, max), de la forma ($Q_{\min}$ y $Q_{\max}$), entonces esta restricción establece que:

$$\lambda_{\min} \leq \Phi_{\min} \quad \wedge \quad \lambda_{\max} \geq \varphi_{\max} \tag{4.32}$$

donde:

$$\begin{aligned} \lambda_{\min} &= \min\{\alpha : \alpha = (Q_{\min})^{-1}(\gamma_{\min})\} \\ \lambda_{\max} &= \max\{\beta : \beta = (Q_{\max})^{-1}(\gamma_{\max})\} \end{aligned} \tag{4.33}$$

Siendo, $\gamma_{\min}$ y $\gamma_{\max}$, los umbrales aplicados a $Q_{\min}$ y $Q_{\max}$ respectivamente.

$$\Phi_{\min} = \begin{cases} a & \text{si } Q_{\min} \text{ es absoluto} \\ a/b & \text{si } Q_{\min} \text{ es relativo} \end{cases}$$

$$\Phi_{\max} = \begin{cases} a & \text{si } Q_{\max} \text{ es absoluto} \\ a/b & \text{si } Q_{\max} \text{ es relativo} \end{cases} \tag{4.34}$$

Siendo a y b los mismos valores definidos en el caso anterior.

*

Por ejemplo, con un cuantificador difuso *casi todos* de una superclase, esta restricción puede expresar que: "*casi todos los elementos de la superclase pertenecen a la categoría"*. Otra opción es unir dos o más superclases con un arco indicando que la unión de las instancias de esas superclases son restringidas por un cuantificador difuso. Esta restricción permite también incorporar la notación (min,max) indicando el número mínimo y máximo que pertenecen a la categoría (usando cuantificadores difusos absoluto o relativos), la notación (min,max) difusa fue definida en el apartado 4.4.2.

Definición 4.19: Una **restricción de completitud difusa en una categoría de un tipo de unión** se define con un arco etiquetado que cruza la línea que une la categoría con el vínculo o círculo con el símbolo unión, y que está etiquetado con un o dos cuantificadores, cuando se usa la notación difusa (min,max). Este tipo de restricción es equivalente a una restricción de participación difusa (Definición 4.20) que abarque todas las superclases: J={1,2,...,n}.

*

Otra opción, es unir dos o más superclases con un arco indicando que la unión de las instancias de esas superclases están restringidas por un cuantificador difuso, en cuanto a su participación en la categoría. Esta restricción permite también la notación difusa (min,max), indicando el número mínimo y máximo que pertenece a la categoría (usando cuantificadores difusos absolutos o relativos).

Por ejemplo, con un cuantificador "casi_todos" en la categoría de una superclase, es una restricción que expresa qué: "*Casi todos los elementos de la unión de todas las superclases pertenecen a la categoría*".

La Figura 4.46 muestra las notaciones de la Definición 4.18 (cuantificador1 y notación (min,max)), y Definición 4.19 (cuantificador2). En el vínculo que une la especialización se debe colocar, en este caso, el símbolo unión.

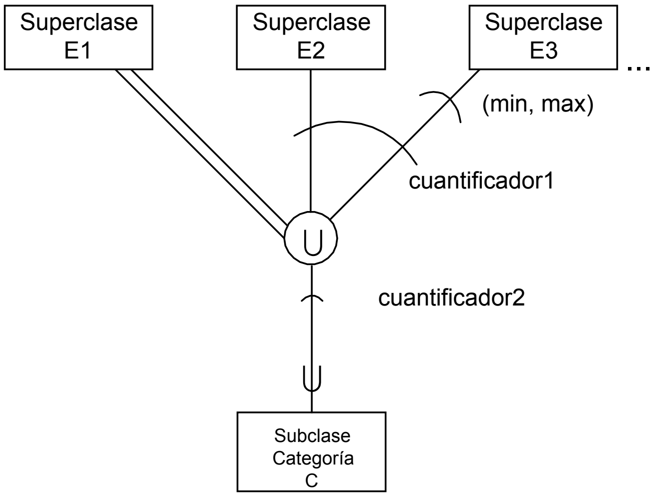


Figura 4.46: Restricciones de participación difusa (cuantificador2) y completitud difusas (cuantificador1) y (min,max) en categorías.

Una subclase compartida o tipo de intersección es una subclase con varias superclases (cada miembro de la subclase debe ser un miembro de todas las superclases), es decir, las subclases es un subconjunto de las intersecciones de todas las superclases. Una subclase compartida es representada en uniones de todas sus superclases por una línea simple con el símbolo de inclusión.

Ejemplo 4.17: Se tienen cuatro superclases: *AUTOMOVIL*, *CAMIÓN*, *MOTOCICLETAS* y *BICICLETAS*, cada vehículo puede pertenecer a la subclase categoría *VEHÍCULOS REGISTRADO* por el vínculo unión. La Figura 4.47 muestra este modelo con algunas restricciones de participación difusa, tales como: "*Casi todos los vehículos y todos los camiones deben ser vehículos registrados*". Además, el modelo permite una restricción (min,max) difusa para bicicletas teniendo "0" como mínimo y como máximo de "aproximadamente_5" bicicletas que pueden ser vehículos registrados. El arco etiquetado con el cuantificador difuso "más_de_6", indica que más de seis motocicletas o bicicletas pueden ser registradas.

Por otra parte, la restricción difusa de completitud establece que "aproximadamente_la_mitad" de los vehículos existentes deben ser vehículos registrados.

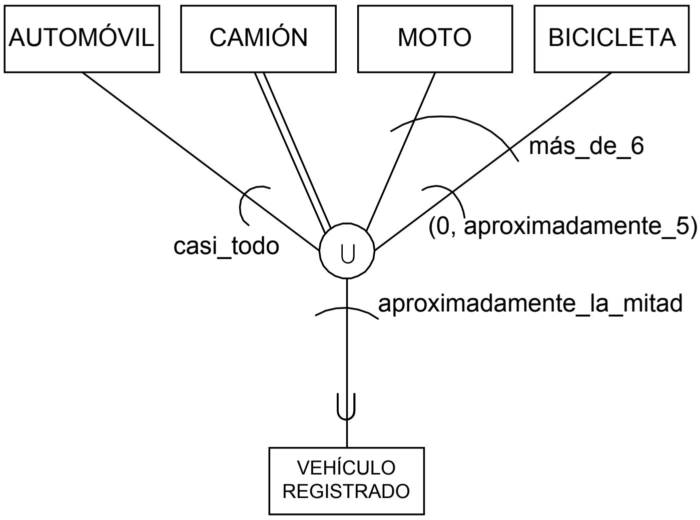


Figura 4.47: Ejemplo de restricción difusa de tipo de categoría o unión.

Otra representación utiliza el símbolo de intersección dentro de un círculo que une a las superclase, y estas superclases están unidas a este círculo por una línea, por otro lado. Las subclases están unidas al círculo usando una línea simple con el símbolo de inclusión (Elmasri y Navathe, 2002).

4.5.2 Restricciones Difusas sobre los Tipos de Intersección en Subclases Compartidas: Participación y Completitud

Tal como en los tipos de unión, en este tipo de especialización es posible aplicar restricciones difusas de dos maneras.

Definición 4.20: Sea S una subclase compartida de un tipo intersección con n superclases E_i, con i=1,2,...,n. Se define **una restricción de participación difusa en una o más superclases**, y se representa con un arco que cruza las líneas que unen las superclases seleccionadas con el círculo de la intersección o vínculo con la subclase. El arco debe estar etiquetado con un cuantificador difuso, o dos con la notación difusa (min,max).

Las superclases seleccionadas E_j, $\forall j \in$ J, con J $\subseteq \{1,2,...,n\}$ y $j<i$, son aquellas a las que le afecta la restricción: . Denotaremos como Δ a la intersección de todas esas superclases, la cual definimos de dos formas:

$$\Delta = \cap_{j \in J} E_j$$

1. Si el arco está etiquetado con un cuantificador Q, esta restricción establece que:

$$Q(\Phi) \geq \gamma \tag{4.35}$$

donde γ es el umbral mínimo establecido para Q y:

$$\Phi = \begin{cases} a & \text{si } Q \text{ es absoluto} \\ a/b & \text{si } Q \text{ es relativo} \end{cases} \tag{4.36}$$

Siendo a es el número de instancias de Δ, que pertenecen a S, y b es el número total de instancias de Δ.

2. Si el arco está etiquetado dos cuantificadores difusos usando la notación (min, max), de la forma ($Q_{\min}$ y $Q_{\max}$), entonces esta restricción establece que:

$$\lambda_{\min} \leq \Phi_{\min} \quad \wedge \quad \lambda_{\max} \geq \varphi_{\max} \tag{4.37}$$

donde:

$$\lambda_{\min} = \min\{\alpha : \alpha = (Q_{\min})^{-1}(\gamma_{\min})\} \tag{4.38}$$

$$\lambda_{\max} = \max\{\beta : \beta = (Q_{\max})^{-1}(\gamma_{\max})\}$$

Siendo, $\gamma_{\min}$ y $\gamma_{\max}$ son los umbrales aplicados a $Q_{\min}$ y $Q_{\max}$ respectivamente.

$$\Phi_{\min} = \begin{cases} a & \text{si } Q_{\min} \text{ es absoluto} \\ a/b & \text{si } Q_{\min} \text{ es relativo} \end{cases}$$
$$\Phi_{\max} = \begin{cases} a & \text{si } Q_{\max} \text{ es absoluto} \\ a/b & \text{si } Q_{\max} \text{ es relativo} \end{cases} \tag{4.39}$$

Siendo a y b los mismos valores definidos en el caso anterior.

*

Por ejemplo, con una superclase usando el cuantificador “*casi_todo*” en la siguiente restricción: “*Casi todos los elementos de la superclases pertenecen a la categoría* (o subclase compartida)”. O bien, con dos o más superclases se unen con un arco indicando que el cuantificador tiene efecto sobre la intersección de las superclases involucradas. Al igual que en los otros casos, se permite el uso de la restricción de la notación difusa (min,max), que indica, el número mínimo y máximo de instancias que pertenecen a la subclase compartida (ya sean cuantificadores difusos absolutos o relativos). Generalmente la restricción de participación no es usada, porque una restricción en una superclase (o en varias superclases) depende de la relación de sus instancias con las otras superclases (recuerde que la subclase es un subconjunto de la intersección). Esto se aclarará con el Ejemplo 4.18.

Definición 4.21: Una **restricción de completitud difusa en una subclase compartida** (tipo de intersección) se define con un arco etiquetado que cruza la línea que une la subclase compartida con el vínculo, y asociado a este el símbolo intersección, y que está etiquetado con uno dos cuantificadores, cuando se usa la notación difusa (min,max). Este tipo de restricción es equivalente a una restricción de participación difusa (Definición 4.20) que abarque todas las superclases: J={1,2,...,*n*}.

*

La Figura 4.48 muestra las notaciones de las Definiciones 4.20 y 4.21. Obsérvese en el vínculo que une la especialización, el símbolo intersección.

Por ejemplo, usando el cuantificador *casi todos* en una superclase compartida representa la siguiente restricción: “*Casi todos los elementos de esa superclase pertenecen a la subclase compartida*”. Esta restricción permite el uso de la notación (min, max), indicando el número mínimo y el máximo de instancias de la intersección de todas las superclase que pertenecen a la subclase compartida. Obsérvese que todas las restricciones son siempre referidas a la intersección de todas las superclases.

Ejemplo 4.18: Consideremos una entidad *EMPLEADO* con los atributos (`sueldo_extra`, `número_de_premios`, `motivación`, etc.). Se puede dar el caso de que un miembro de esta subclase deba ser, a la vez, ingeniero, jefe y empleado permanente. La Figura 4.48 muestra este modelo con la siguiente restricción de participación: “*Casi todos los jefes y empleados permanentes deben ser empleados especiales*”. Es interesante notar cómo esta restricción fuerza a que casi todos los *jefes* y *empleados permanentes* sean *ingenieros* a la vez. Esto es así debido a que todos los empleados especiales pertenecen a la subclase ingeniero. Por otra parte, la restricción difusa de completitud establece que la mitad de los empleados que son ingenieros, jefes y empleados permanentes, deben ser además empleados especiales.

Obsérvese que una restricción de participación difusa que abarque todas las superclases es, de hecho, una restricción difusa de completitud (tanto de tipo unión como de tipo de intersección).

En el Ejemplo 4.15 se explicó la especialización total por tipo de atributo difuso `T3:Categoría`, así como también la restricción de especialización parcial por tipo de atributo difuso `T3:Area_investigación`.

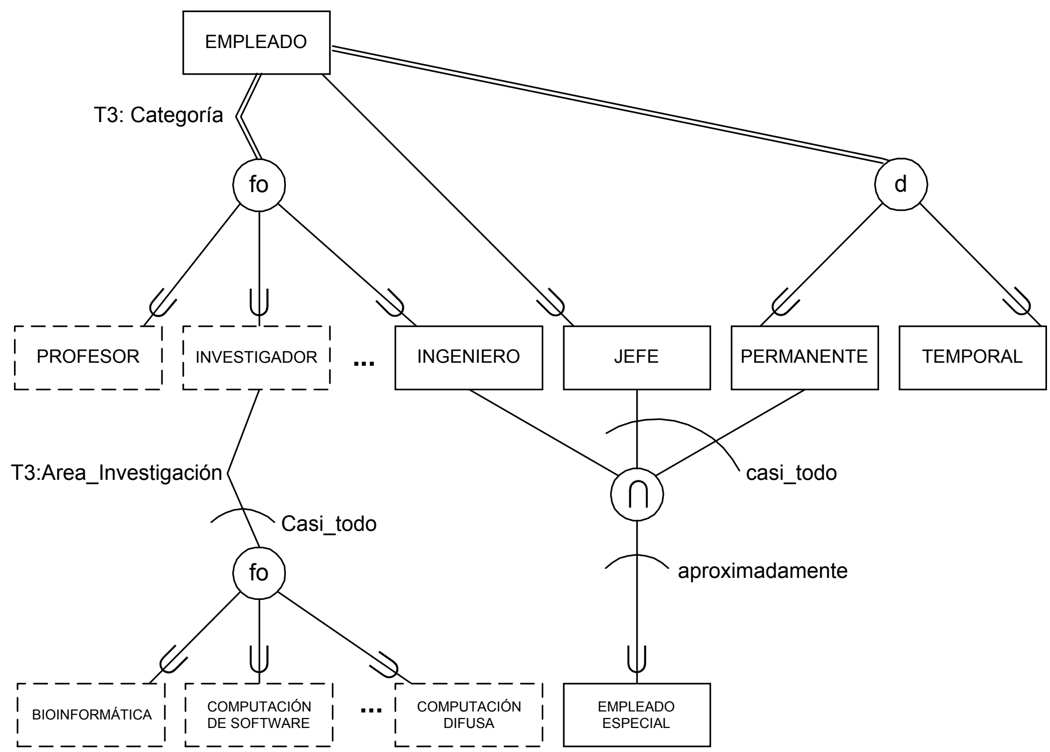


Figura 4.48: Ejemplo de atributos difusos definidos en una especialización solapada y restricciones difusas en una subclases compartidas.

Finalmente hasta aquí, hemos extendido el modelo ER/EER a un modelo difuso **FuzzyEER** con un total de **21 definiciones** y **18 nuevas notaciones**, las cuales son resumidas en el apartado 4.7.1 y comparadas con otros trabajos en el apartado 4.7.2.

4.6 Notaciones Alternativas Utilizadas en Publicaciones de la Tesis

Al inicio de esta tesis, se publicaron algunos trabajos con notaciones que luego fueron dejándose de lado por otras más representativas, las que presentamos a continuación.

Para el caso de los **atributos difusos** se utilizó la notación de una línea quebrada y sobre ella el tipo de atributo difuso marcado a la vez con un círculo estrellado (Urrutia y Galindo, 2001). Posteriormente, se desecho la línea quebrada, ya que, bastaba sólo con el círculo estrellado para identificar el tipo de atributo difuso, siendo incorporado al nombre del atributo difuso anteponiéndole el tipo de atributo (T1, T2, T3 y T4) separado por dos puntos. Las dos notaciones son mostradas, en la Figura, 4.49 a) la notación actual y en la Figura 4.49 b) la notación inicial.

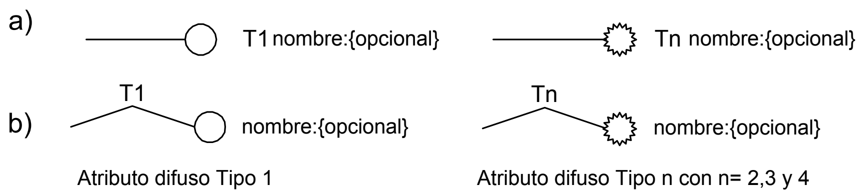


Figura 4.49: Dos notaciones para tipos de atributos difusos.

Para el caso es la **restricción de participación difusa**, en efecto al inicio fue propuesta una notación de una *línea zigzag* junto a ella uno o más cuantificadores que unía la entidad con la interrelación que relacionaba, tal como lo muestra la Figura 4.50 (Galindo et al., 2001). Posteriormente, esta fue cambiada por una línea simple interceptada por un arco junto al cuantificador, ya que, esta notación simplificaba su uso cuando una interrelación tenia un grado mayor a dos, y a su vez, unidas por una restricción de un mismo cuantificador.

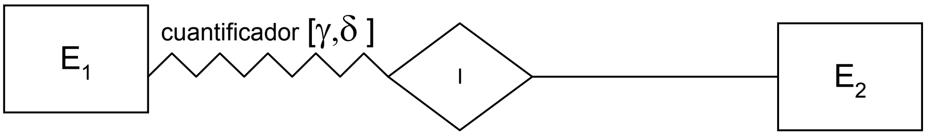


Figura 4.50: Restricción de participación difusa por línea zigzag.

Para el caso de la **restricción de participación difusa en la especialización**, en efecto, al inicio fue propuesta una *línea zigzag* unía la superclase al vínculo de especialización junto con el cuantificador (Galindo et al., 2001). Posteriormente se propuso *un arco* sobre una línea simple y junto a este el cuantificador. La razón de este cambio, es que la notación se hace compleja cuando un cuantificador relaciona más de una subclase o superclase, esto lo muestra la Figura 4.51.

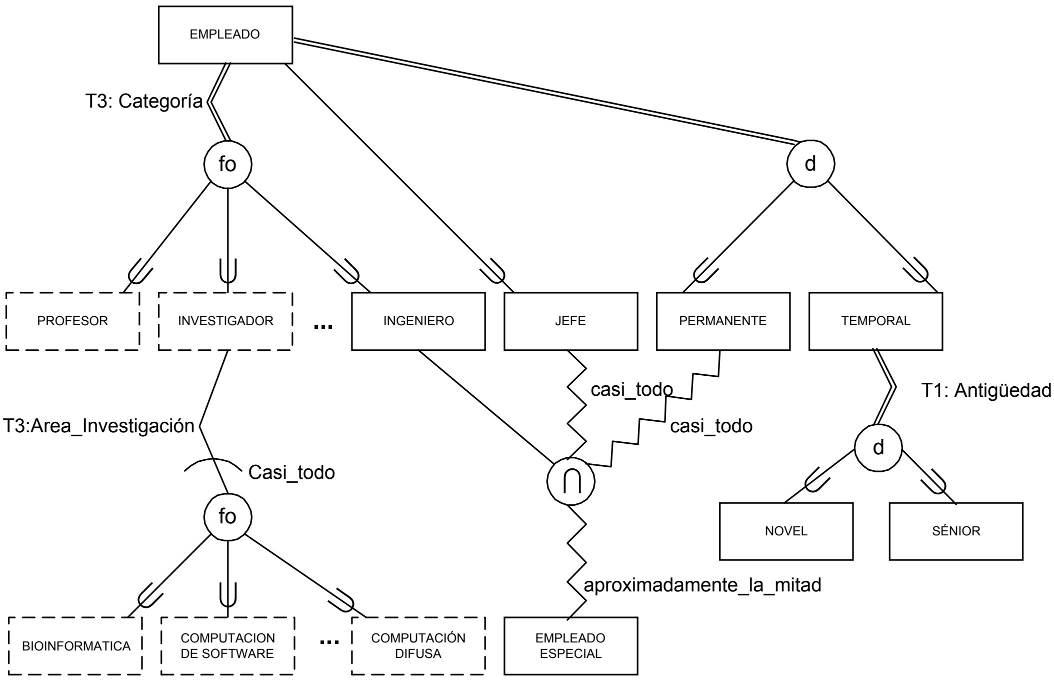


Figura 4.51: Ejemplo de participación difusa usando línea zigzag y cuantificador en jerarquías.

En modelos reales, distintas restricciones difusas sobre la misma especialización deben ser mezcladas con cuidado. La Figura 4.51 muestra el mismo ejemplo de la Figura 4.48, pero utilizando la notación de línea zigzag para el caso de la participación y completitud difusa, lo que lo hace más compleja cuando un cuantificador restringe a más de una superclase o subclase (ejemplo: subclase *EMPLEADO ESPECIAL* en la unión de las superclases *JEFE* y *PERMANENTE*). Además, de agregar una especialización difusa por un tipo de atributo difuso "`T1:Antigüedad`", que da un grado de semejanza entre la superclase *EMPLEADOS* y las subclases *TEMPORAL* con antigüedad de Novel o Senior. Si comparamos ambos esquemas (Figura 4.48 y 4.51) se ve que la notación con arcos para la participación y completitud difusa, es más representativa que la notación de los arcos asociados a un cuantificador difuso o notación (min,max), ya que la restricción expresada con la línea quebrada tiende a quedar confusa en el esquema FuzzyEER.

Por último, para el caso de **Tipos de atributos difusos en Interrelación difusa**, se propuso un rombo con línea cortada, opcionalmente se coloca, en la parte inferior del rombo un atributo difuso T1, T2, T3 o T4, según sea el caso, entre paréntesis el número máximo de valores asociados al atributo difuso, así lo muestra la Figura 4.52 con una interrelación difusa por el atributo difuso `T3:Barrio` donde se consideran tres barrios como máximo para las entidades *INMUEBLE* y *BARRIO*. Posteriormente se propuso las dos notaciones mostradas en la Figura 4.15: un atributo difuso se identifica con un círculo estrellado, y un grado difuso con un círculo con línea cortada, ambos casos, unidos al rombo con línea cortada que simboliza la interrelación difusa.

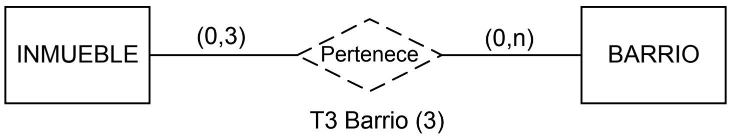


Figura 4.52: Interrelaciones difusas por atributos difuso `T3 Barrio` (3).

4.7 Discusión del Capítulo

Se ha presentado una extensión del modelo de datos FuzzyEER que flexibiliza las restricciones del modelo ER/EER, considerando cuantificadores difusos tanto relativos como absolutos. El modelo propuesto permite también recoger especificaciones de requerimientos para datos que presentan imprecisión, considerando los tipos de dominios subyacentes de un atributo (ordenado o no ordenado). Nuestra propuesta incorpora en gran parte las investigaciones expuestas en el capítulo 3 (Yazici y Merdan, (1996); Chen G., (1998) y Ma et al., (2001)), además de otras características, esto lo muestra la Tabla 4.5. Al respecto, consideramos que esta tesis se puede seguir expandiendo en algunas restricciones más complejas, como por ejemplo: especializaciones en la superclase como una entidad difusa, categorías difusas y subclases compartidas difusas. Cada una de ellas son casos especiales de las notaciones ya mostradas en este capítulo, así que no será complejo representarla a la hora de que una especificación de requerimientos de un sistema las solicite.

Gran parte de lo investigado y propuesto en este capítulo se encuentran en publicaciones, actas de congresos y en revistas comentadas en el apartado 7.3. Por otro lado, en el Capítulo 6, se muestra una herramienta CASE, llamada FuzzyCASE, que permite generar esquemas de sistemas de información con las propuestas del modelo de datos conceptual FuzzyEER, así como también, algunos sistemas de información modelados con FuzzyEER están el capítulo 5 y apéndice IV.

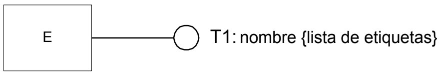
El modelo FuzzyEER se puede resumir en un total de **18 representaciones gráficas**, **21 definiciones** explicadas formalmente y **4 tipos de atributos difusos**. Acompañamos también, una tabla de comparación (véase Tabla 4.5) del modelo FuzzyEER, con las propuestas de algunos autores que nos han parecido más relevantes expuestas en el apartado 3.3. Además la Tabla 4.6 muestra una tabla de comparación de los tipos de datos difusos que presentan algunos autores con respecto a los nuestros.

4.7.1 Resumen de la Notación FuzzyEER

En este apartado presentamos un resumen de las notaciones (18 representaciones gráficas) propuestas para el modelo FuzzyEER.

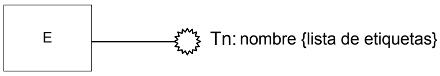
1. Valores difusos en los atributos.

a)



Atributo difuso Tipo 1

b)



Atributo difuso Tipo n con n= 2,3 y4

c)

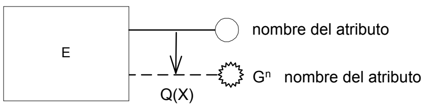
atributo difuso simple — atributo difuso derivado

d)

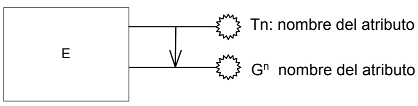
atributo difuso múltiple opcional — atributo difuso múltiple obligatorio

2. Grado en cada valor de un atributo: Grado difuso asociado a un atributo.

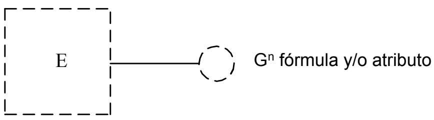
a)



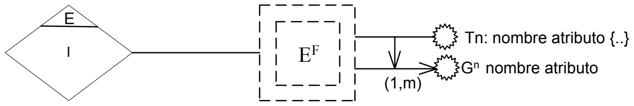
b)



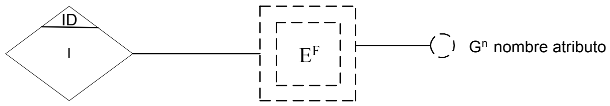
3. Entidades difusas: Grado en toda la instancia de la entidad.



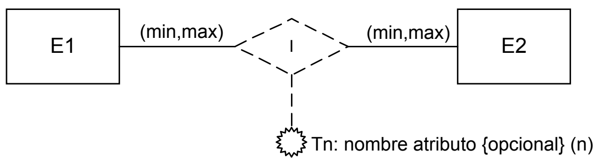
4. Entidad débil difusa.



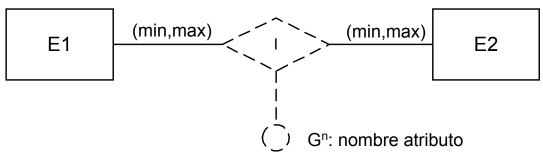
5. Entidad débil difusa definida por un grado.



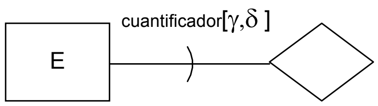
6. Interrelación difusa: Tipo de atributos difusos en la interrelación.



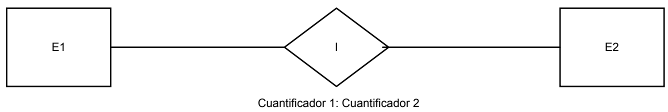
7. Interrelación difusa: Grado en la interrelación.



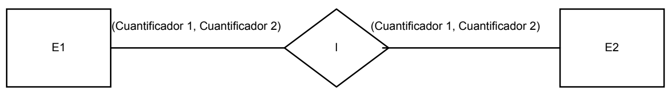
8. Participación difusa definida por cuantificadores difusos.



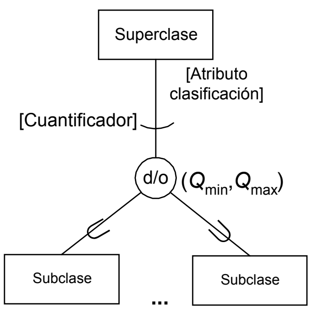
9. Razón de cardinalidad o Tipo de correspondencia con cuantificadores difusos.



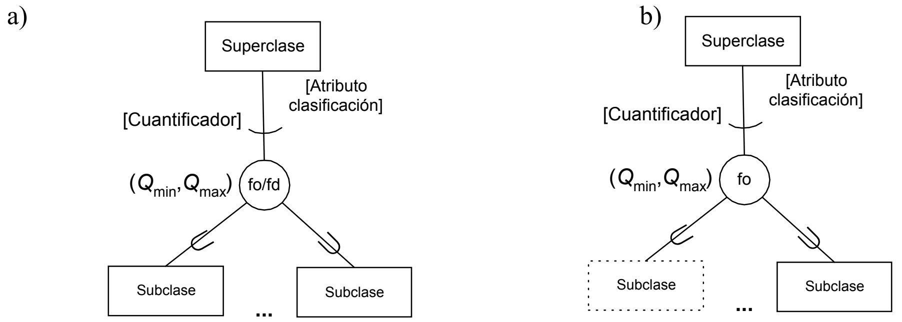
10. Notación (min,max) con cuantificadores difusos en interrelaciones.



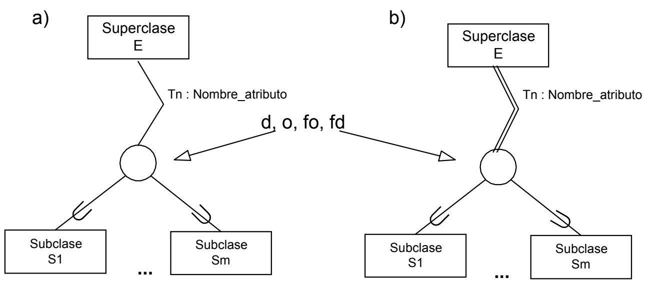
11. Restricción de completitud difusa y notación (min,max) con cuantificadores difusos.



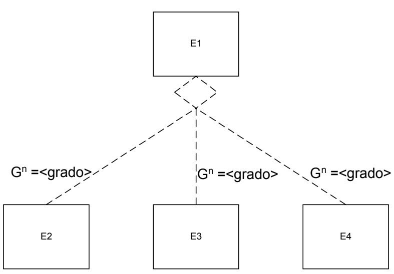
12. Restricciones difusa disjuntas y solapadas sobre especializaciones con notación (min,max). a) Sin subclases difusas y b) Con subclases difusas para especialización solapada.



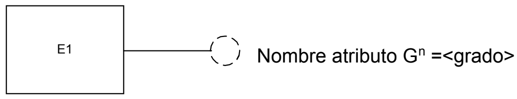
13. Atributo difuso que definen especializaciones de participación a) parcial y b) total.



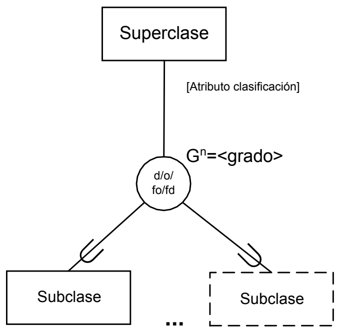
14. Agregación difusa de entidades.



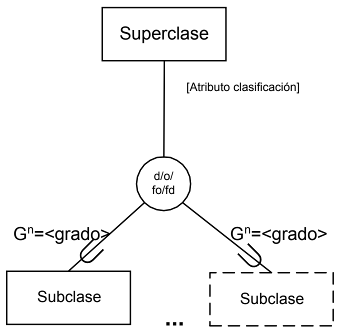
15. Agregación difusa de atributos.



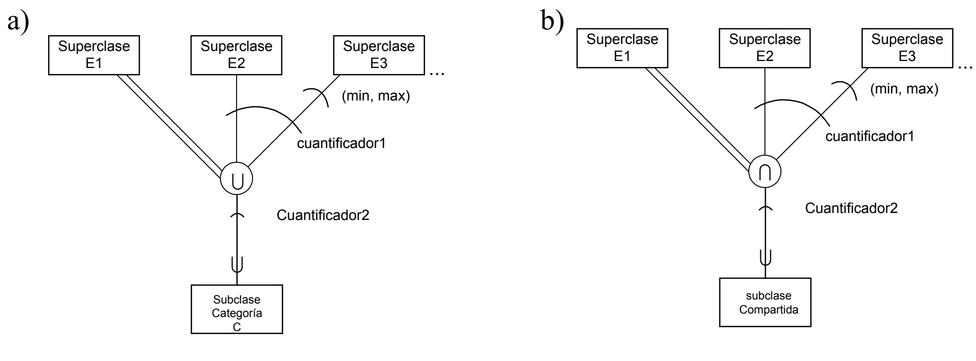
16. Grado en la especialización.



17. Grado en la subclase de especialización.



18. Restricción difusa de completitud en subcategorías. a) unión, y b) intersección.



4.7.2 Tabla de Comparación de Modelos Difusos

La Tabla 4.5 muestra una comparación del modelo FuzzyEER con otros modelos difusos de algunos autores que fueron discutidos en el apartado 3.3. Cada celda denota un "Sí" si el modelo tiene esa notación, en caso contrario la celda esta vacía. Cuando además la celda incluya un "Sí*" significa que la notación está presente en ese modelo pero con características distintas a la del modelo FuzzyEER, o bien, sus características son limitadas y más reducidas que las del modelo FuzzyEER aquí propuesto. Por lo general esta diferencia está dada por el uso de cuantificadores difusos en FuzzyEER y algún tipo de grado en los otros modelos.

Características / Modelos difusos	FEER Ma et al. (2001)	FERM Chaudhry et al. (1994)	ExIFO Yazici y Merdan (1996)	Fuzzy ER Chen (1998)	FuzzyEER Urrutia (2003)
1. Valores difusos en los atributos	Sí*	Sí*	Sí*	Sí*	Sí
2. Grado en cada valor de un atributo: Grado difuso asociado a un atributo	Sí*		Sí*	Sí*	Sí
3. Entidades difusas	Sí*		Sí*	Sí*	Sí
4. Entidad débil difusa		Sí*			Sí
5. Entidad débil por grado de atributo					Sí
6. Interrelación difusa: Tipo de atributos difusos en la interrelación		Sí*			Sí
7. Interrelación por grado de un atributo	Sí*		Sí	Sí	Sí
8. Participación difusa				Sí*	Sí
9. Razón de cardinalidad o Tipo de correspondencia				Sí*	Sí
10. Notación (min,max)					Sí
11. Restricción de completitud difusa y notación (min,max)					Sí
12. Restricciones difusa disjuntas y solapadas sobre especializaciones				Sí*	Sí
13. Atributo difuso que definen especializaciones				Sí*	Sí
14. Agregación difusa en entidades	Sí		Sí	Sí	Sí
15. Agregación difusa de atributos	Sí		Sí	Sí	Sí
16. Grado en la especialización	Sí*		Sí*	Sí	Sí
17. Grado en las subclases de especialización	Sí*			Sí	Sí
18. Restricción difusa de completitud en unión e intersección				Sí*	Sí
19. Herramienta gráfica					Sí

* Notación presente pero con características distintas o limitadas a las del modelo FuzzyEER

Tabla 4.5: Comparación de modelos difusos FEER, FERM, ExIFO, Fuzzy ER y FuzzyEER.

Es importante destacar, que estos autores no han utilizado cuantificadores difusos en las restricciones, como se ha dicho anteriormente. Otro cuadro comparativo de modelo difusos se puede establecer en lo referente al tipo de atributos que pueden utilizar. El resultado se muestra en la Tabla 4.6.

Atributos difusos / Modelos difusos	FEER Ma et al. (2001)	FERM Chaudhry et al. (1994)	ExIFO Yazici y Merdan (1996)	Fuzzy ER Chen (1998)	FuzzyEER Urrutia (2003)
Tipo 1					Sí
Tipo 2	Sí*	Sí*	Sí*	Sí*	Sí
Tipo 3	Sí*		Sí*		Sí
Tipo 4					Sí
grados G^0 o G^1o G^2o G^3 o G^4				Sí*	Sí
Otros		Sí			

* Notación presente pero con características distintas o limitadas a las del modelo FuzzyEER

Tabla 4.6: Comparación de atributos difusos en los modelos FEER, FERM, ExIFO, Fuzzy ER y FuzzyEER.

Considérese, que las distintas celdas de las Tablas 4.5 y 4.6 marcados Sí* tienen diferencias con el modelo FuzzyEER que fueron discutidas en el capítulo 3 apartado 3.3.