Q-ANALISIS Y LOGICA DIAMOND: UN MODELO CONEXIONISTA PARA LA FORMALIZACION DE DICCIONARIOS DE SINONIMOS

BLANCA CASES
FACULTAD DE INFORMATICA. UPV/EHU
DEPTO. LENGUAJES Y SISTEMAS INFORMATICOS
APDO. 649
20080 DONOSTIA GIPUZKOA

Con el fin de desarrollar métodos para la evaluación y corrección de diccionarios de sinónimos, se aborda la representación de un diccionario de sinónimos como un complejo simplicial, en el que cada entrada léxica constituye un poliedro. Esta representación constituye la "forma física" del diccionario, soporte sobre el cual se desarrolla un flujo lógico que proviene de la lectura encadenada de sus definiciones. Para simular dicha lectura se formaliza cada entrada como una ecuación autorreferente en una lógica que incluye valores para denotar paradojas. Conseguimos caracterizar algunos errores típicos como paradojas, y tomando como restricciones las impuestas por la estructura conectiva del soporte físico, incorporarlas al sistema de ecuaciones.

REPRESENTACION DE UN DICCIONARIO DE SINONIMOS COMO COMPLEJO SIMPLICIAL

Un diccionario de sinónimos consiste básicamente en una lista de entradas de la siguiente forma:

<entrada>::= <clave> : <definición>

<clave>::= <palabra>

<definición>::=<sentido> . | <sentido>; <definición>

<sentido>::=<palabra> | <palabra> , <sentido>

Es decir, las entradas de un diccionario de sinónimos pueden adoptar alguna de las siguientes formas:

$A : A_{11}, \ldots, A_{1m1} ; \ldots ; A_{n1}, \ldots, A_{nmn}$.

o bien

$A : A_1, \ldots, A_m$.

En el primero de los casos, la palabra A tiene varios sentidos y/o acepciones, siendo cada uno de los grupos $A_{i1}, \ldots, A_{imi}$ los sinónimos de la palabra A para un significado en concreto. En el segundo caso, el sentido de A es único, siendo $A_1, \ldots, A_m$ los sinónimos de A.

Sea P el conjunto de todas las palabras que aparecen en el diccionario D. La manera más simple de representar la relación de sinonimia establecida por D es como una relación binaria $r \subseteq P \times P$ tal que:

1º p r p Para toda palabra p del diccionario D.

2º p r p' <-> existe una entrada e en D con clave p en cuya definición aparece p'.

La relación de sinonimia r definida por un diccionario no coincide en la práctica con la sinonimia "ideal" que definen los linguistas, según la cual dos palabras son sinónimas si pueden sustituírse una por otra en cualquier contexto [6]. La relación de sinonimia así definida debe formalizarse como una relación de equivalencia, es decir, debe ser reflexiva, simétrica y transitiva. Es bien conocido el trabajo de I. Warnesson, quien utiliza el "Método de Agregación de Similaridades" para encontrar la relación de equivalencia más cercana a la relación de sinonimia establecida por el diccionario sobre un subconjunto del léxico.[8]

El planteamiento que hacemos aquí es diferente. Partimos de la base de que la relación de sinonimia "ideal" no puede definirse sobre palabras aisladas sino sobre sus significados. Así pues, teniendo en cuenta que una entrada del diccionario de sinónimos define el significado de la palabra clave de dicha entrada, elegimos otro tipo de representación para la relación r definida por D. Dicha representación, debida a Atkin [1] [2] [3], contempla que cualquier relación $r \subseteq X \times Y$, $X=\{ x_1, \ldots, x_n\}$,

$Y=\{y_1, \ldots y_m\}$ se puede expresar como un complejo simplicial $K_X(Y;r)$ en un espacio

- Cada x_i de X $1 \le i \le n$ se representa como un poliedro convexo o símplice.

$x_i = <y_{i0}, \dots, y_{iji}>$ siendo los y_{ij}, vértices del poliedro, todos los elementos de Y tales que el par (xi,yij) pertenece a $r \subseteq XxY$.

- La dimensión del poliedro $x_i = <y_{i0}, \dots, y_{iji}>$ es $dim(x_i)=j_i$, es decir, el número de vértices menos uno.

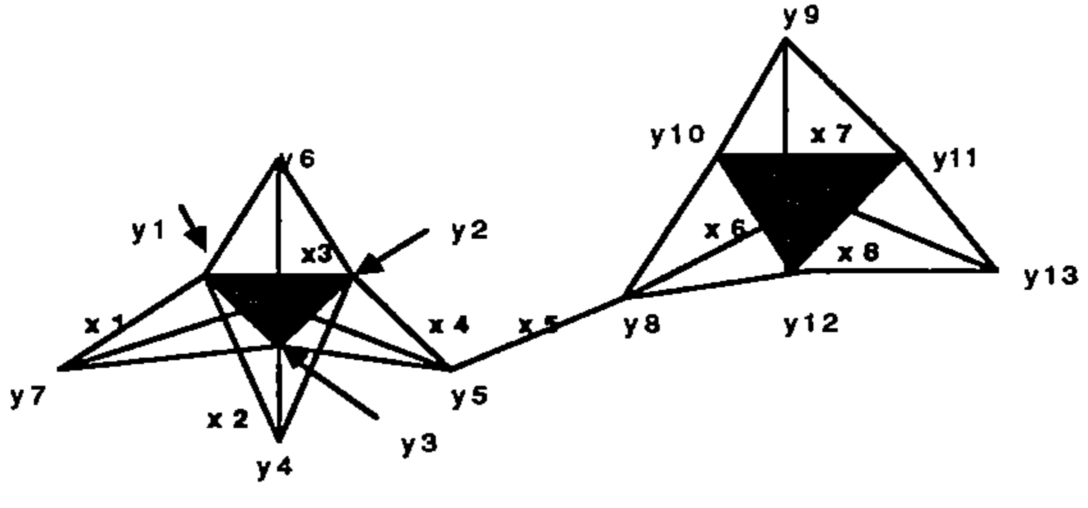
No entraremos aquí en los detalles matemáticos de esta representación. Unicamente diremos que es suficiente con que la dimensión del espacio E^N sea N=2Q+1, siendo Q la mayor de las dimensiones de los símplices de KX(Y;r) denotada por dim[KX(Y;r)], para que sea posible acomodar a todo el complejo.

La representación de r como complejo simplicial proporciona una idea gráfica de la estructura conectiva de dicha relación. Sean por ejemplo:

$X=\{ x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$

$X=\{ y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8, y_9, y_{10}, y_{11}, y_{12}, y_{13}\}$

r	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}	y_{13}
x_1	1	1	1	0	0	0	1	0	0	0	0	0	0
x_2	1	1	1	1	0	0	0	0	0	0	0	0	0
x_3	1	1	1	0	0	1	0	0	0	0	0	0	0
x_4	1	1	1	0	1	0	0	0	0	0	0	0	0
x_5	0	0	0	0	1	0	0	1	0	0	0	0	0
x_6	0	0	0	0	0	0	0	1	0	1	1	1	0
x_7	0	0	0	0	0	0	0	0	1	1	1	1	0
x_8	0	0	0	0	0	0	0'	0	0	1	1	1	1



clasificaciones de acuerdo al criterio de q-conectividad:

Sea q $1 \leq q \leq Q$ $Q = \dim [KX(Y;r)]$. Se dice que dos símplices x y x' de X están q-conectados sii existe una cadena de poliedros de X

$x = p_0, p_1, \ldots, p_k = x'$ tales que $\dim(p_i \& p_{i+1}) \geq q$. Con $(p_i \& p_{i+1})$ denotamos la cara compartida por los poliedros p_i y p_{i+1} siendo $\dim(p_i \& p_{i+1})$ el número de vértices menos uno. Si p_i y p_{i+1} no tienen vértices en común, la dimensión de la cara compartida es por defecto -1.

En el ejemplo anterior, Q=3, y el resultado del q-análisis para los distintos q-niveles es

q=3 $[x_1]\,[x_2]\,[x_3]\,[x_4] \quad [x_6]\,[x_7]\,[x_8]$

q=2 $[x_1\ x_2\ x_3\ x_4] \quad [x_6\ x_7\ x_8]$

q=1 $[x_1\ x_2\ x_3\ x_4] \quad [x_5] \quad [x_6\ x_7\ x_8]$

q=0 $[x_1\ x_2\ x_3\ x_4\ x_5\ x_6\ x_7\ x_8]$

Se llama vector de estructura de r al vector que especifica para cada q el número de clases de equivalencia que resultan del q-nivel y es en este caso $[0^1\ 1^3\ 2^2\ 3^7]$.

Si dada una relación $r \subseteq XxY$ nos fijamos en la relación

$r^t \subseteq YxX = \{(y,x) \mid (x,y)$ pertenece a $r\}$

y formamos el complejo simplicial conjugado $K^*Y(X;r^t)$, se puede obtener información interesante acerca de la naturaleza de la relación r, sobre todo si interpretamos que en un complejo simplicial KX(Y;r) X es un conjunto cuyos elementos se describen mediante los elementos del conjunto Y. En el caso que nos ocupa, en que

$r \subseteq PxP$, siendo P el conjunto de palabras del diccionario D, la comparación entre los complejos simpliciales KP(P;r) y $K^*P(P;r^t)$ permite analizar globalmente la simetría de la relación r mediante los correspondientes vectores de estructura. En el caso de los diccionarios de traducción entre dos lenguas, siendo X e Y los vocabularios a estudiar en cada una de ellas, el estudio del complejo simplicial conjugado permite reducir el análisis a poliedros definidos sobre el vocabulario de una única lengua, X.

Otro concepto que definimos en esta representación es el de excentricidad relativa entre dos símplices. Si x y x' son dos símplices en KX(Y;r), se define la excentricidad relativa de x con respecto a x' como

$ecc(x,x') = [\dim(x) - \dim(x\&x')] \div [\dim(x\&x')+1]$

que toma el valor ∞ si $\dim(x\&x') = -1$.

ecc(x,x') mide la proporción de vértices que x no comparte con x'. Si calculamos la matriz de excentricidades relativas en un complejo simplicial KX(Y;r) y realizamos sobre dicha matriz un rebanamiento para convertirla en una matriz booleana, obtenemos al realizar el q-análisis clasificaciones que ofrecen más información que la relación original.

SINONIMIA Y AUTORREFERENCIA. LOGICA DIAMOND

2º p r p' <-> existe una entrada e en D con clave p en cuya definición aparece p'

proporciona una idea gráfica del contenido del diccionario, de lo que podemos llamar su "estructura física", pero no del flujo lógico que resulta de una lectura encadenada de sus definiciones. Existe además otro hecho, y es que en una definición de la forma

$A : A_{11}, \dots , A_{1m1} ; \dots ; A_{n1}, \dots , A_{nmn}$.

el signo de puntuación ";" determina una relación de oposición entre los grupos de palabras que separa, mientras que el signo "," define como sinónimas a las palabras de un mismo grupo.

En este sentido, el hecho de que dos palabras p y p' aparezcan en una definición separadas mediante ";" mientras que aparecen en otra separadas mediante "," debe interpretarse como un error en la construcción del diccionario.

Una manera de interpretar este tipo de errores es pensar que provienen de la autorreferencia: así, mientras al estar p y p' separadas por "," en un grupo podemos decir que p y p' son sinónimas y que por lo tanto

$p=p'$

al estar separadas por ";" en otro grupo podemos decir que

$p=\neg p'$

Si damos valores booleanos a p y p' encontramos que el sistema de ecuaciones $p=p'$ / $p=\neg p'$ no tiene solución posible.

Nos inspiramos en este hecho para formalizar las definiciones de un diccionario de sinónimos como un sistema de ecuaciones autorreferentes en el que las palabras toman valores en la lógica Diamond [4], [5].

Dicha lógica incluye además de los valores lógicos v (cierto) y f (falso) otros dos valores i y j ordenados en un diamante:

j

v f

i

Los valores lógicos i, j se definen como aquellos que satisfacen ecuaciones del estilo $x=\neg x$, que no pueden satisfacerse en el conjunto {v,f}.

Los valores lógicos de Diamond se construyen como pares de valores lógicos sobre el conjunto {v,f} mediante el operador PERO (abreviado "|").

v es v|v

f es f|f

i es v|f

complejos, construídos como pares de números reales.

El operador PERO se extiende al diamante de la siguiente forma

(a|b) | (c|d)=a|d, y por lo tanto su tabla es la siguiente:

|	v	f	i	j
v	v	i	i	v
f	j	f	f	j
i	v	i	i	v
j	j	f	f	j

Sobre los valores de Diamond podemos definir dos negaciones:

La primera, "¬", es extensión de la negación de la lógica proposicional y recibe el nombre de negación armónica

¬	v	f	i	j
	f	v	i	j

verifica $\neg\neg x = x$, $\neg(a|b)=(\neg b)|(\neg a)$

La negación no armónica se define como

~	v	f	i	j
	v	f	j	i

verifica $\sim\sim x= x$, $\sim(a|b)=(\sim b)|(\sim a)$

Los conectivos lógicos de la lógica proposicional se extienden a los valores Diamond de la siguiente manera:

(a|b) v (c|d)= (a v c) | (b v d)

(a|b) & (c|d)= (a & c) | (b &d) siendo sus tablas

v	v	f	i	j
v	v	v	v	v
f	v	f	i	j
i	v	i	i	v
j	v	j	v	j

&	v	f	i	j
v	v	f	i	j
f	f	f	f	f
i	i	f	i	f
j	j	f	f	j

Se definen también los operadores "+" y "x" , que calculan el máximo y mínimo respectivamente entre dos valores lógicos siguiendo el orden definido sobre el diamante:

x	v	f	i	j
v	v	i	i	v
f	i	f	i	f
i	i	i	i	i
j	v	f	i	j

+	v	f	i	j
v	v	j	v	j
f	j	f	f	j
i	v	f	i	j
j	j	j	j	j

La relación de orden determinada por el diamante puede definirse a partir de los operadores "+" y "x" de la siguiente manera:

x>y <-> x=x + y

x<y <-> x=x x y

<variable>=<expresión>

Así, partiendo de una valoración inicial de las variables, se van calculando valoraciones sucesivas como resultado de evaluar simultáneamente los miembros derechos de las ecuaciones, hasta alcanzar un punto fijo.

FORMALIZACION DE UN DICCIONARIO DE SINONIMOS COMO SISTEMA DE ECUACIONES AUTORREFERENTE.

Una entrada del diccionario

$A : A_{11}, \dots, A_{1m1} ; \dots ; A_{n1}, \dots, A_{nmn} .$

se interpreta como una ecuación autorreferente que tiene la siguiente forma:

$$A = (OP1\ \ (OP4\ (FUN\ a\ a)\ \ A)\ \ (OP2_{k=1}^{n}\ (OP3_{l=1}^{m_k}\ (OP4\ (FUN\ a_{kl}\ \ a)\ \ A_{kl}))))$$

Es decir:

* Cada signo de puntuación se interpreta como un operador lógico

$: = OP1 \quad ; = OP2 \quad , = OP3$

* Cada símbolo **A** o $\mathbf{A_{kl}}$ representa una palabra de la entrada que actúa como variable en la ecuación autorreferente.

* **a** y $\mathbf{a_{kl}}$ representan los símplices asociados a **A** y $\mathbf{A_{kl}}$ respectivamente en la representación del diccionario como KP(P;r), siendo (FUN a_{kl} a) una expresión que produce un valor lógico diamond a partir de las relaciones numéricas que se pueden encontrar entre dichos poliedros (las dimensiones de cada uno y de su cara compartida, excentricidad relativa, etc.)

* OP4 es un operador que actúa sobre una palabra A_{kl} y el valor lógico que representa la conexión que existe entre Akl y la clave de la entrada A, expresada mediante (FUN a_{kl} a) como queda dicho.

Cuando la definición tiene la forma

$A : A_1, \dots, A_m.$

la ecuación correspondiente sería

$$A = (OP1\ \ (OP4\ (FUN\ a\ a)\ \ A)\ \ (OP3_{l=1}^{m}\ (OP4\ (FUN\ a_l\ \ a)\ \ A_l)))$$

y si A es una palabra de P que no aparece como clave

$A=(OP1\ (OP4\ (FUN\ a\ a)\ \ A)$

El diccionario se convierte entonces en un sistema de ecuaciones autorreferente donde las

- El valor j se interpreta como contradicción
- Los valores v y f se interpretan como indicadores de pertenencia a clases diferentes C_1 y C_2 : así, todas las palabras de C1 pueden recibir valores i o v, pero no f. Las palabras de C2 pueden recibir valores i o f, pero no v. El valor j indica la pertenencia a las dos clases.

Definiremos ahora los operadores OP_j. Supondremos que

$(OP4\ (FUN\ a\ a_{kl})\ A_{kl}) = A_{kl}$, ya que el valor lógico (FUN a a_{kl}) tiene una interpretación diferente de la que hemos dado para las variables.

1.-El signo de puntuación ":" corresponderá al operador "+", máximo entre valores lógicos. Si la definición del diccionario es

A : <definición>, la ecuación autorreferente tendrá la forma

A = A + <expresión>

y por lo tanto los sucesivos valores que adopte A constituirán una sucesión creciente, lo que asegura la convergencia del sistema.

2.-El signo "," se identifica también con el operador "+". Así, si en un grupo de palabras separadas por "," aparecen a la vez los valores v y f (que indican pertenencia a clases diferentes), automáticamente se registra una contradicción. Si los valores para el grupo están en el conjunto {i,v} o en el conjunto {i,f} el valor resultante será respectivamente v o f. Es de hacer notar que si en la entrada es de la forma

$A : A_1, \ldots, A_m.$

la respectiva ecuación será

$A : A + A_1 + \ldots + A_m.$

Esto implica que:

*si A tiene el valor i, el nuevo valor que adquiera será el transmitido por $A_1 + \ldots + A_m$.

*si A tiene un valor en {v,f} contradictorio con el de $A_1 + \ldots + A_m$, dicha contradicción se registra transmitiéndose a A el valor j.

3.-El signo de puntuación ";" requiere una definición especial, ya que separa grupos que no son sinónimos entre sí, estableciéndose entre ellos una relación de oposición.

Si $A : A_{11}, \ldots, A_{1m1} ; \ldots ; A_{n1}, \ldots, A_{nmn}$. es una entrada del diccionario, es de esperar que para cada par $A_{kl}\ A_{k'l'}$, $k \neq k'$ que nunca pertenezcan a una misma clase. Por lo tanto, jamás podrán tener simultáneamente los valores v o f, debiendo detectarse dicha determinar cuál sentido elegir).

Así pues, identificamos el signo ";" con el operador n-ario $n \geq 2$

$$**(x_1, \ldots, x_n) = \sum_{1 \leq k,l \leq n \;\; k \neq l} (x_k + \neg x_l) \times (\neg x_k + k_l)$$

Es decir,

$**(x_1, \ldots, x_n) =$ j si alguno de los x_k es j o si existen k, l $1 \leq k,l \leq n$ $k \neq l$ tales que $x_k = xl = v$ ó $xk = xl = f$. En otro caso, i.

La ecuación que representa a una entrada del diccionario indica cómo se transmiten a la clave de dicha entrada los valores lógicos relativos a las palabras que la forman. El objetivo es, a partir de una valoración inicial de las variables conseguir otra valoración que constituya un punto fijo para el sistema de ecuaciones, como resultado de sucesivas iteraciones a partir de la valoración inicial. De acuerdo con la interpretación anterior, dicho punto fijo tendrá el siguiente significado:

* Todas las palabras que reciban el valor v en el punto fijo serán sinónimas de las que recibieron el mismo valor inicialmente. Este conjunto incluirá a las palabras que tenían el valor v de partida más las claves de las entradas del diccionario que permiten llegar a dichas palabras a través de una cadena de definiciones. Nótese aquí que la relación de sinonimia se extiende al cierre transitivo de la relación r definida anteriormente y que representábamos como KP(P;r).

* Lo dicho para el valor v será exactamente igual para el valor f. Sólo hay que tener en cuenta que v y f indican pertenencia a clases de sinónimos diferentes.

* Las palabras que reciban el valor j son las que sirven de nudo para llegar a una u otra clase mediante la lectura encadenada de definiciones, o bien las que son claves de entradas con varios significados y la lectura encadenada atraviesa dos o más de entre ellos, interpretándose así como contradicciones.

* Las que reciben el valor i son aquellas que no intervienen en el proceso o las que formando parte de una cadena consistente tienen varios sentidos diferentes.

Estamos ahora en condiciones de ilustrar la interpretación anterior mediante un ejemplo.

Sean

A: B, C, D; E, F.	A = A+ B + C + D; E + F.
B: A, C, D.	B = B + A + C + D.
C: A, B, I.	C = C+ A + B + I.
E: F, G, H.	E = E+ F + G + H.
I: E, G.	I = I + E + G.

que asigna v a la clase C_1 y f a la C_2.

A	B	C	D	E	F	G	H	I	
i	v	v	v	f	f	i	i	i	valoración inicial
i	v	v	v	f	f	i	i	f	el valor f se transmite a I
i	v	j	v	f	f	i	i	f	el elemento C acusa una contradicción que viene de I=f
j	j	j	v	f	f	i	i	f	la contradicción se transmite a A y B por C
j	j	j	v	f	f	i	i	f	obtenemos el punto fijo

La corrección pertinente sería, por lo tanto

A: B, C, D; E, F. A = A+ B + C + D; E + F.

B: A, C, D. B = B + A + C + D.

C: A, B. C = C+ A + B . Eliminar I de la definición

E: F, G, H. E = E+ F + G + H.

I: E, G. I = I + E + G.

De esta manera, ninguna lectura encadenada de las definiciones llevaría a mezclar las dos clases iniciales.

UN MODELO CONEXIONISTA PARA LA REPRESENTACION DE DICCIONARIOS DE SINONIMOS.

Hemos supuesto en el apartado anterior que (OP4 (FUN a a_{kl}) A_{kl})= A_{kl}. Esto puede conseguirse si hacemos (FUN a a_{kl})=j constante y OP4=x.

Cuando para una entrada A : $A_{11}, \dots, A_{1m1}$; ... ; $A_{n1}, \dots, A_{nmn}$. contemplamos el valor (FUN a a_{kl}), nos estamos refiriendo a un valor lógico que expresa algo acerca de la estructura conectiva de la relación de sinonimia interpretada como complejo simplicial y que representaba la "estructura física" del diccionario. Dicha estructura física puede limitar el flujo de lectura encadenada de las definiciones. Hemos supuesto en el apartado anterior que (OP4 (FUN a a_{kl}) A_{kl})= A_{kl}. Esto puede conseguirse si hacemos (FUN a a_{kl})=j constante y OP4=x.

En un caso más general, observemos el siguiente ejemplo

A: B, C, D, E.

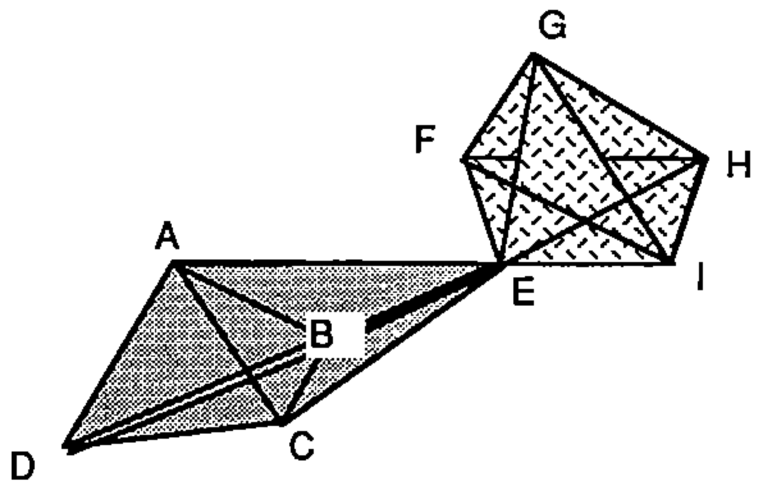
B: A, C, D.

C: A, B, D.

E: F, G, H, I.

G: E, F, H.

H: E, F, G.



Si interpretamos el conjunto de palabras P como complejo simplicial, y si consideramos que dos palabras pueden considerarse sinónimos sólo si los respectivos poliedros están 3-conectados, obtenemos las siguientes clases para la 3-conexión

[A B C] [E G H I]

Establecemos entonces sobre el conjunto de símplices la siguiente función:

(FUN x x')= j si dim(x&x')≥3

i si dim(x&x')<3

E identificamos el operador OP4 con x.

Por lo tanto, si (FUN a b)= j --> (FUN a b) x B=B

si (FUN a b)= i --> (FUN a b) x B=i

El valor j en una conexión significa, por lo tanto, que existe suficiente conectividad entre los poliedros a y b como para que el valor de B pueda ser transmitido a la entrada de clave A. El valor i, por el contrario, indica que la conexión entre B y la clave A no es suficiente como para permitir la transmisión del valor B. El sistema de ecuaciones asociado a las definiciones anteriores sería, por lo tanto:

A= (jxA)+(jx B)+(jxC)+(ixD)+(ixE)

B=(jxB) + (jxA) + (jxC) + (ixD)

C=(jxC) + (jxA) +(jx B) +(ix D)

D=ixD

E=(jxE) + (ix F) + (jxG) + (jxH) + (jxI)

F=ixF

A B C D E F G H I

v i i i i i f f i valoración inicial

v v v i f i f f f

v v v i f i f f f punto fijo

No se produce contradicción, como hubiera ocurrido de no establecer restricción alguna. La representación del diccionario mediante un modelo de esta índole, inspirado en la computación con Redes Neuronales, [7] abre las puertas a una enorme gama de posibilidades, entre ellas la de incorporar un sistema de aprendizaje que permita modificar la propia estructura de las entradas del diccionario para adecuarlo a fines predeterminados. Esta es la línea en la que trabajamos y en la que esperamos obtener resultados en breve.

BIBLIOGRAFIA

1 Atkin, Ronald H.
"The Metodology of Q-analisys Applied to Social Systems"
SYSTEM MEHODOLOGY IN SOCIAL SCIENCE RESEARCH
Edited by Roger Cavallo
Kluwer.Nijhoff Publishing 1982 Boston

2 Casti, John
CONNECTIVITY, COMPLEXITY AND CATASTROPHE IN LARGE-SCALE SYSTEMS
John Wiley & Sons. 1979

3 Gould, P.
"Q-analysis, or a language of structure: an introduction for social scientists, geographers and planners"
International Journal of Man Machine Studies (1980) 13, pp.169-199
1980 Academic Press (London)

4 Hellerstein, Nathaniel
"Diamond: a Logic for Paradox"
Cybernetics vol 1 Nº 1 Summer-Fall 1985
American Society for Cybernetycs. c/o Departament of Decisions Sciences
George Mason University, Fairfax, VA 22030

5 Kauffman, Louis H.
"Self-reference and Recursive Forms"
Journal of Social and Biological Structures
1987 10 pp. 53-72
Academic Press. Londres

6 Lyons, John
INTRODUCCION EN LA LINGUISTICA TEORICA
Ed. Teide Barcelona 7ª edición 1985

7 Lippmann, Richard P.
"An Introduction to Computing with Neural Nets"
IEEE ASSP MAGAZINE APRIL 1987