



Indistinguibilidad y Modelado de Tamaño en Sistemas *Data-to-text*

Nicolás Marín, Gustavo Rivas-Gervilla y Daniel Sánchez,
Departamento de Ciencias de la Computación e Inteligencia Artificial
Universidad de Granada, 18071 Granada, España
Email: {nicm, griger, daniel}@decsai.ugr.es

Abstract—En este trabajo extendemos nuestra propuesta previa de modelado de tamaño en base al contexto, incluyendo el tratamiento de la indistinguibilidad en la percepción del tamaño. Ilustramos la importancia de este aspecto con una experimentación que, adicionalmente, muestra una forma de determinar experimentalmente los parámetros del método.

Index Terms—Tamaño, modelado difuso, descripción lingüística de datos, sistemas *data-to-text*

I. INTRODUCCIÓN

El lenguaje natural es una herramienta cada vez más imprescindible en el desarrollo de interfaces de comunicación hombre-máquina amigables. Prueba de ello es la importancia, cada vez más marcada, de la investigación en el ámbito de Generación de Lenguaje Natural, en general, y en el ámbito de los sistemas *data-to-text* en particular. Este tipo de sistemas tienen como objetivo elaborar información textual a partir del análisis de datos [1]. Particularmente, se pretende que dichos textos sean similares a los que escribiría un experto humano.

Uno de los problemas más importantes en este ámbito es determinar la semántica de los términos lingüísticos empleados en los textos, así como el grado de correspondencia entre dichos términos y los datos. Dado que dichos términos suelen representar conceptos difusos, es natural considerar conjuntos difusos como modelos de dicha semántica [2]–[5]. Estos conjuntos difieren de los conjuntos clásicos (o *crisp*) en que cada elemento del dominio pertenece al conjunto con un grado en el intervalo $[0, 1]$. La función que asocia a cada elemento este grado se denomina función de pertenencia. Uno de los usos más importantes de estos términos es la identificación de objetos en un conjunto, para lo cual se utilizan *expresiones de referencia*, que son sintagmas nominales formados habitualmente por una conjunción de términos que expresan propiedades de los objetos.

En [6] hemos propuesto un enfoque para la determinación de las funciones de pertenencia que modelan la semántica de conceptos difusos relativos al tamaño. Dicha propuesta considera funciones que son dependientes y condicionadas por el contexto de la aplicación donde se van a usar los términos lingüísticos.

Este trabajo ha sido financiado en parte por el Gobierno de España y el Fondo Europeo de Desarrollo Regional - FEDER a través del proyecto TIN2014-58227-P. Así como por la beca del Ministerio de Educación, Cultura y Deporte de España, FPU16/05199. Corresponding author: G. Rivas-Gervilla.

En este trabajo avanzamos un paso más en nuestra propuesta incluyendo en la misma el tratamiento de la indistinguibilidad. Este problema surge debido a que las computadoras son capaces de almacenar información sobre los objetos (sobre su tamaño en este caso) de manera más precisa de lo que nosotros somos capaces de percibir visualmente, lo cual afecta a nuestro método.

En la siguiente sección describimos la importancia del contexto en el modelado de términos relativos al tamaño y resumimos la propuesta presentada en [6]. En la sección III describimos el problema de la indistinguibilidad y cómo afecta a nuestro método, y detallamos los cambios en la misma que nos permitirán adaptarnos al grado de indistinguibilidad del tamaño de los objetos en un contexto concreto. En la sección IV se presenta un experimento con usuarios para ilustrar nuestra discusión. En esa misma sección se muestra una posible forma de determinar experimentalmente uno de los parámetros clave en nuestra propuesta. Finalmente, las dos últimas secciones contienen una breve discusión de los resultados y las conclusiones de nuestro trabajo, respectivamente.

II. MODELADO DEL TAMAÑO

A. Modelado y contexto

Como ya hemos indicado, uno de los aspectos clave en el desarrollo de sistemas *data-to-text* es la definición de la semántica de los términos lingüísticos que se van a emplear en los textos generados. En el caso particular de conceptos relativos a propiedades de los objetos, dicha semántica debe establecer una correspondencia entre características y valores de los datos asociados a dichos objetos y el cumplimiento de la propiedad. En el caso particular de propiedades difusas (como puede ser el tamaño, ya que un objeto puede ser grande en un cierto grado), este objetivo se cumple mediante la definición de las funciones de pertenencia asociadas a dichos conceptos difusos, como es bien conocido. Determinar las funciones de pertenencia es uno de los problemas clave en el uso de conjuntos difusos en éste y otros dominios de aplicación.

Sin embargo, la determinación de la función de pertenencia no plantea únicamente el problema de cómo obtener la asignación de grados a elementos de un referencial. En muchas ocasiones ocurre que la función de pertenencia no es única, sino que depende del contexto de aplicación (datos concretos, usuario, objetivos, etc.). Es decir, la definición de la función de pertenencia es relativa al contexto.

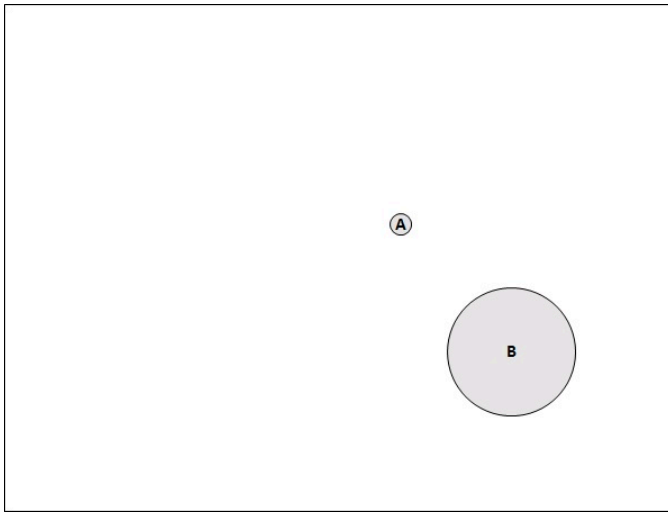


Fig. 1. Escena con objetos A y B

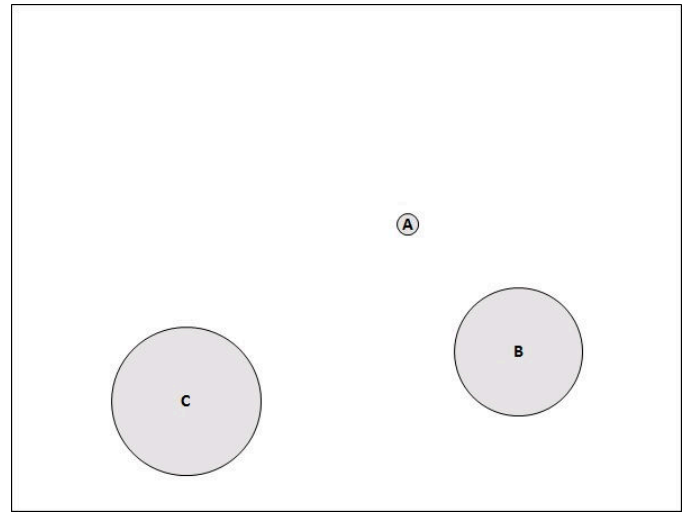


Fig. 2. Escena con objetos A, B y C

Para ilustrar esta situación en el contexto de sistemas *data-to-text*, vamos a considerar el ejemplo de las Figuras 1 y 2, extraído de [6]. En la escena de la Figura 1, para referirnos al objeto B es suficiente decir *el objeto grande*. Sin embargo, en la escena de la Figura 2, que contiene los mismos objetos de la Figura 1 más un objeto adicional, la referencia *el objeto grande* no llevaría a un usuario hipotético a señalar el objeto B, sino el C. Es decir, el contexto (en este caso, la aparición de nuevos datos, concretamente el objeto C) hace que el cumplimiento de la propiedad “tamaño” cambie, particularmente porque el rango de valores de tamaño que consideramos ha cambiado con la aparición de un nuevo tamaño (el tamaño de C). En este caso, el objeto B pasa a encajar mejor con la etiqueta *mediano*, siendo por tanto más adecuado en la Figura 2 referirse a B como *el objeto mediano* de cara a conseguir que un hipotético usuario lo identifique correctamente. Por lo tanto, queda claro que la definición de las funciones de pertenencia para las etiquetas *grande* y *mediano* es dependiente del contexto, por lo que fijar un conjunto de funciones de pertenencia independiente del contexto a priori no nos permite interactuar correctamente con el usuario en la mayoría de los casos, como demostramos en [6].

B. Una propuesta de modelado basada en contexto

En [6] aportamos una primera aproximación a la solución de este problema. La propuesta consiste en determinar las funciones de pertenencia para las etiquetas *grande*, *mediano* y *pequeño* asociadas a la propiedad “tamaño” de forma dependiente del contexto asociado a los datos, teniendo en cuenta los distintos tamaños precisos de los objetos presentes en el conjunto de datos/escena. Asumiendo que el conjunto de tamaños es $S = \{s_1, \dots, s_n\}$, $n > 1$, con $s_i \in \mathbb{R}^+$ y $0 < s_1 \leq \dots \leq s_n \forall 1 \leq i \leq n$, la propuesta puede resumirse como sigue:

- En primer lugar, determinamos los tamaños que van a ser totalmente representativos de cada una de las eti-

quetas consideradas, es decir, van a tener grado 1 de pertenencia a las mismas. De manera natural establecemos por definición $pequeño(s_1) = grande(s_n) = 1$ y $pequeño(s_n) = grande(s_1) = mediano(s_1) = mediano(s_n) = 0$. Nótese que esta definición es independiente de los valores concretos de s_1 y s_n , así como de la diferencia $|s_1 - s_n|$ entre ambos (volveremos sobre este último punto más adelante, puesto que es una de las motivaciones del presente trabajo). En el caso $n > 2$ tendremos también representantes de la propiedad *mediano*. Si n es impar hay un único representante: $s' = s'' = s_{(\lceil n/2 \rceil)}$. En caso contrario tenemos dos representantes: $s' = s_{(n/2)}$ y $s'' = s_{(n/2+1)}$. Para todo representante $s \in \{s', s''\}$ de mediano tenemos $mediano(s) = 1$ y $pequeño(s) = grande(s) = 0$ por definición.

- El grado de pertenencia a las tres etiquetas de los tamaños en S que no son representativos de ninguna de ellas se determina mediante un proceso de *clustering*, centrado en los valores representativos de cada etiqueta, y en base a las distancias de dichos tamaños con respecto a los representativos. Como método de *clustering* empleamos el método definido en [7] en base a representaciones por niveles [8]. Este método toma como punto de partida una relación difusa reflexiva y simétrica de semejanza en el conjunto de tamaños S . En el caso particular de los tamaños, esta relación se obtiene en base a las distancias a través de un parámetro d_m que define la máxima distancia a la que pueden estar dos valores para que su semejanza sea mayor que 0. Sean $s_i, s_j \in S$ y sea $d(s_i, s_j) = |s_i - s_j|$ la distancia Euclidea entre ellos. La semejanza entre s_i y s_j en base al parámetro d_m , $R^{d_m}(s_i, s_j)$, se calcula mediante la siguiente expresión:

$$R^{d_m}(s_i, s_j) = \max \left\{ 1 - \frac{d(s_i, s_j)}{d_m}, 0 \right\} \quad (1)$$



Como se demuestra en [6], para garantizar las condiciones que hemos expuesto anteriormente sobre los grados de pertenencia de s_1 , s' , s'' y s_n a las tres etiquetas de tamaño, es suficiente con elegir un d_m que satisfaga las siguientes condiciones:

- En el caso $n = 2$ se requiere $d_m \leq s_n - s_1$.
- En el caso $n > 2$ se requiere $0 < d_m \leq \min(d(s_1, s'), d(s'', s_n))$. Esta condición implica asimismo $d_m \leq s_n - s_1$ en este caso.

A partir de una relación de semejanza difusa como ésta, el proceso de *clustering* basado en representación por niveles se basa en la realización de un *clustering crisp* en los diferentes niveles $\Lambda(R^{d_m}) = \{R^{d_m}(s_i, s_j) \mid s_i, s_j \in S\} \setminus \{0\}$. Este conjunto es finito, $\Lambda(R^{d_m}) = \{\alpha_1, \dots, \alpha_k\}$ con $k \geq 1$ y $1 = \alpha_1 > \dots > \alpha_{k+1} = 0$. Para cada $\alpha_k \in \Lambda(R^{d_m})$ puede obtenerse un *clustering crisp* basado en recubrimientos de S de la siguiente forma:

- 1) Calcular el α_k -corte de R^{d_m} , que es una relación *crisp* reflexiva y simétrica.
- 2) Calcular el *clustering* de S a nivel α_k como el conjunto de cliques maximales del α_k -corte de R^{d_m} . Sea $C_{\alpha_k} \subset \{0, 1\}^S$ el conjunto de *clusters* obtenidos, cada uno de ellos asociado biunívocamente a un clique maximal. Nótese que estos *clusters* pueden solaparse, formando un recubrimiento de S .

Como consecuencia de las restricciones impuestas sobre d_m se da $R^{d_m}(s_1, s_n) = 0$ para todo d_m que las cumpla, y por tanto $|C_{\alpha_k}| \geq 2$, dado que s_1 y s_n estarán siempre en distintos *clusters*. Además es fácil demostrar que existe un solo *cluster* $C_{\alpha_k}^p \in C_{\alpha_k}$, verificando $s_1 \in C_{\alpha_k}^p$, y un solo *cluster* $C_{\alpha_k}^g \in C_{\alpha_k}$, cumpliéndose $s_n \in C_{\alpha_k}^g$, siendo como ya hemos comentado $C_{\alpha_k}^p \neq C_{\alpha_k}^g$. Además, en el caso $n > 2$ definimos:

$$C_{\alpha_k}^m = \bigcup \{C_{\alpha_k}^i \mid C_{\alpha_k}^i \cap \{s', s''\} \neq \emptyset\} \quad (2)$$

Es decir, $C_{\alpha_k}^m$ es la unión de todos los *clusters* que contienen al menos a uno de los representantes de la categoría *mediano*. Nótese que $C_{\alpha_k}^m$ puede ser vacío, algo que ocurre en todos los niveles particularmente cuando $n = 2$.

A partir del *clustering* en cada nivel, los *clusters* difusos que modelan las funciones de pertenencia para las tres propiedades *pequeño*, *mediano* y *grande* se obtienen, siguiendo las ideas expuestas en [7], [8], como sigue: $\forall s_i \in S$,

$$\text{pequeño}(s_i) = \sum_{\alpha_k \mid s_i \in C_{\alpha_k}^p} (\alpha_k - \alpha_{k+1}) \quad (3)$$

$$\text{mediano}(s_i) = \sum_{\alpha_k \mid s_i \in C_{\alpha_k}^m} (\alpha_k - \alpha_{k+1}) \quad (4)$$

$$\text{grande}(s_i) = \sum_{\alpha_k \mid s_i \in C_{\alpha_k}^g} (\alpha_k - \alpha_{k+1}) \quad (5)$$

α_k	$C_{\alpha_k}^p$	$C_{\alpha_k}^m$	$C_{\alpha_k}^g$
1	{10}	{25}	{40}
2/3	{10}	{20, 25, 30}	{40}
1/3	{10, 20}	{20, 25, 30}	{30, 40}

TABLE I
Clusters crisp POR NIVELES PARA $d_m = 15$.

Como ejemplo, sea $S = \{10, 20, 25, 30, 40\}$ ($n = 5$). En este contexto tenemos que $s_1 = 10$ es el tamaño representativo para *pequeño*, $s_5 = 40$ para *grande*, y $s_3 = 25$ es el (único) representante para *mediano*. Sea $d_m = 15$, que es el máximo valor posible en este ejemplo. Dado que el conjunto de distancias entre tamaños es $\{0, 5, 10, 15, 20, 30\}$ tenemos que $\Lambda(R^{15}) = \{1, 2/3, 1/3\}$. La Tabla I muestra los *clusters crisp* en cada nivel. En base a esta información y las Ecuaciones (3)-(5), se tiene finalmente:

$$\begin{aligned} \text{pequeño} &= 1/10 + (1/3)/20 \\ \text{mediano} &= (2/3)/20 + 1/25 + (2/3)/30 \\ \text{grande} &= (1/3)/30 + 1/40 \end{aligned}$$

III. INDISTINGUIBILIDAD Y MODELADO

A. Influencia de la indistinguibilidad

Uno de los grandes problemas en la interacción con datos a través de las computadoras es lo que se denomina *hueco semántico*¹. La principal forma en que se manifiesta este problema es que, mientras que los ordenadores almacenan datos descritos mediante valores de variables en dominios que suelen ser muy precisos y contener muchos valores distintos (por ejemplo, mediante números reales), los seres humanos solemos utilizar información descrita en base a conceptos expresados de forma lingüística. Hemos podido ver distintos ejemplos claros de esta diferencia en las secciones anteriores: mientras que para el ordenador el tamaño de un objeto se describe mediante un valor real numérico, los seres humanos utilizamos términos lingüísticos como *grande*, etc. Cada uno de estos términos se corresponde con un grupo o intervalo de valores precisos, grupo o intervalo que suele tener una frontera imprecisa, y por tanto una forma adecuada de modelar dicha correspondencia es mediante el uso de conjuntos difusos, como hemos visto y es bien conocido en nuestro ámbito.

Sin embargo, existe un problema adicional asociado al *hueco semántico* que recibe menos atención en la literatura, y que tiene que ver con la distinguibilidad. Este problema es típico en, aunque no exclusivo de, aplicaciones que involucran la percepción humana, como lo es el problema del tamaño que nos ocupa en este trabajo, y puede expresarse de manera informal como sigue: el ser humano no es capaz de percibir diferencias en el valor de propiedades de objetos al mismo nivel que una computadora.

Podemos ilustrar el problema que supone la indistinguibilidad para nuestro método de modelado sensible al contexto con

¹Del inglés *semantic gap*.

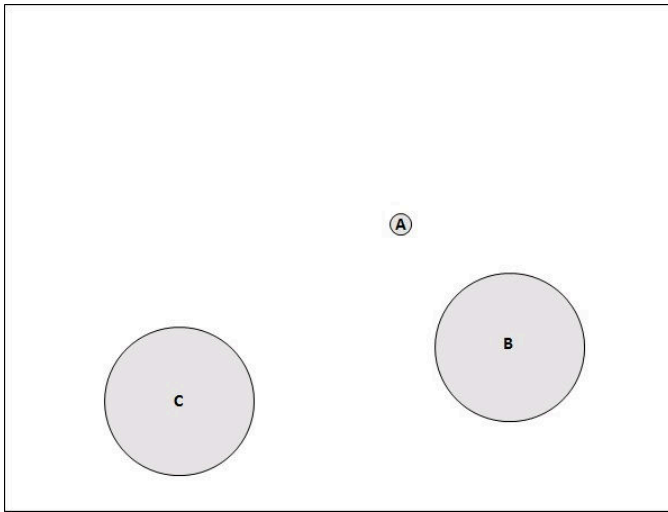


Fig. 3. Modificación de la escena de la Figura 2: tamaño del objeto B distinto pero indistinguible del tamaño del objeto C

un ejemplo. Consideremos la escena de la Figura 3, que se ha obtenido a partir de la escena de la Figura 2 aumentando el tamaño del objeto B hasta ser indistinguible (pero no igual) del tamaño del objeto C. En este caso, dado que para una computadora los tamaños de C y B son distintos, nuestro método diría que el objeto B es *mediano* con grado 1, mientras que los objetos A y C son medianos con grado 0. En base a este modelado sensible al contexto, un algoritmo de generación de expresiones de referencia propondría *el objeto mediano* como una forma de identificar claramente al objeto B. Sin embargo, al ser los tamaños de B y C indistinguibles visualmente para el usuario, éste no tendría claro a cuál de los dos objetos (B ó C) estamos haciendo referencia con dicha expresión (dado que A es claramente el pequeño en este caso).

Para solventar este problema es necesario modificar nuestra propuesta para tener en cuenta la distinguibilidad de objetos. Para ello vamos a considerar las siguientes hipótesis:

- Es posible determinar un umbral de distancia d_s tal que los tamaños s_i y s_j son distinguibles si, y solo si, $d(s_i, s_j) > d_s$.
- El umbral d_s puede ser sensible a distintos aspectos del contexto, tales como distancia entre los objetos, etc. Más adelante prestaremos atención a este problema.

En la siguiente sección detallamos la modificación que proponemos a nuestro método de modelado sensible al contexto para resolver el problema de la indistinguibilidad de objetos.

B. Método propuesto

Nuestra propuesta consta de los siguientes pasos:

- 1) Determinación del umbral de distinguibilidad d_s adecuado para el modelado en el contexto concreto de aplicación. Es posible obtener d_s de distintas formas. En la Sección IV veremos un sencillo ejemplo de obtención de d_s de forma experimental bajo ciertos supuestos. El



Fig. 4. Ejemplo de pregunta del cuestionario.

estudio y propuesta de otras técnicas se plantea como trabajo futuro.

- 2) Determinación del umbral de similitud d_m para el contexto concreto de aplicación, teniendo en cuenta el valor de d_s en dicho contexto. Es importante determinar d_m después de d_s dado que el valor de este último afecta a las condiciones que se imponen sobre d_m . En primer lugar, por la semántica de ambos umbrales, es evidente que debe cumplirse $d_s < d_m$, dado que dos tamaños son indistinguibles cuando la distancia entre los mismos está por debajo de d_s , mientras que son absolutamente diferentes cuando la distancia está por encima de d_m . Pero además, el valor de d_s afecta a la forma en que se obtiene d_m , dado que, como hemos visto, este último viene restringido por condiciones que tienen que ver con la distancia entre objetos representativos de cada una de las etiquetas para los cuales, tras la discusión y el ejemplo que hemos visto en la sección anterior, debemos exigir que sean distinguibles. Más concretamente:

- Si s_1 y s_n son indistinguibles entonces no podemos modelar tamaños de forma sensible al contexto.
- Supongamos s_1 y s_n distinguibles y sea $S_m \subset S$ el subconjunto de tamaños que son distinguibles tanto de s_1 como de s_n (obviamente $s_1, s_n \notin S_m$). Entonces:
 - Si $S_m = \emptyset$ solo podemos modelar dos tamaños, *pequeño* y *grande*, siendo sus representantes s_1 y s_n , respectivamente.
 - Si $S_m \neq \emptyset$ entonces podemos modelar también *mediano*, siendo su(s) representante(s) el valor o valores centrales de S_m (dependiendo de que S_m tenga un número impar o par de valores).

Una vez determinados los representantes de las etiquetas se seleccionará un valor de d_m adecuado que cumpla con los requisitos ya expuestos anteriormente: $d_m \leq s_n - s_1$ para dos etiquetas, y $0 < d_m \leq$



$\min(d(s_1, s'), d(s'', s_n))$ para las tres.

- 3) Aplicar nuestra técnica de *clustering* para determinar las funciones de pertenencia para las tres etiquetas. Para ello utilizaremos una relación de semejanza que introduce una modificación con respecto a la de la Ecuación (1), dado que queremos que cuando la distancia entre dos tamaños sea menor o igual a d_s , su semejanza sea 1. La nueva relación de semejanza viene dada por

$$R^{d_m}(s_i, s_j) = \begin{cases} 1 & d(s_i, s_j) \leq d_s \\ 1 - \frac{d(s_i, s_j)}{d_m} & d_s < d(s_i, s_j) < d_m \\ 0 & d_m \leq d(s_i, s_j) \end{cases} \quad (6)$$

Es fácil demostrar que cuando $d_s = 0$ (lo cual implica que todo par de tamaños distintos son distinguibles), nuestra propuesta se reduce a la propuesta expuesta en la Sección II-B.

IV. EJEMPLO DE APLICACIÓN

Para ilustrar la influencia de la indistinguibilidad y su dependencia del contexto se ha realizado una sencilla experimentación. Concretamente, se ha elaborado una encuesta con diferentes escenas en las que cada escena consta de dos objetos de distinto tamaño a una distancia determinada. Al usuario se le pregunta que identifique el más pequeño de los dos, permitiéndole elegir entre tres respuestas posibles: *el que está a la derecha, el que está a la izquierda y no sabría decirte*. La figura 4 muestra un ejemplo de escena de la encuesta (aparece marcada la respuesta correcta).

Para elaborar el conjunto de escenas de la encuesta, se ha trabajado con tres granos de tamaño $C < B < A$ (elegidos en los extremos y centro del referencial), con cinco niveles de distancia (0, 0.3, 0.5, 0.7 y 1), también relativos, y configurando ejemplos en el que el ratio de parecido en tamaño va desde 0.95 hasta 0.995. La experimentación se ha realizado en la misma computadora con seis usuarios distintos a los que se les presenta el mismo conjunto de escenas con orden aleatorio.

El objetivo de la encuesta es señalar la posible dependencia del contexto de la indistinguibilidad, comprobando la variabilidad de respuestas entre usuarios, tamaños y distancias, e ilustrar una forma experimental para fijar el umbral d_s para una configuración de contexto dada.

A. Resultados

La tabla II muestra, para cada pareja distancia/grano de tamaño, los umbrales de ratio entre tamaños a partir de los cuales se ha obtenido una respuesta distinta de la correcta. Se resalta en negrita el valor mínimo obtenido en cada columna. Las distancias se muestran en valor relativo, representando el 1 la máxima distancia utilizada y el 0 la mínima distancia utilizada.

La figura 5 ilustra los datos de la tabla de manera gráfica. Como se puede observar, en la experimentación realizada, conforme descende el tamaño, en general, baja también el umbral, es decir, el usuario necesita una mayor distancia en

TABLE II
UMBRALES AGREGANDO CON EL MÍNIMO

Distancia (rel.)	Tamaño		
	A	B	C
1	98	98,5	95
0,7	98,5	98,5	96
0,5	98	98	95
0,3	98	98,5	97
0	98,5	98	96

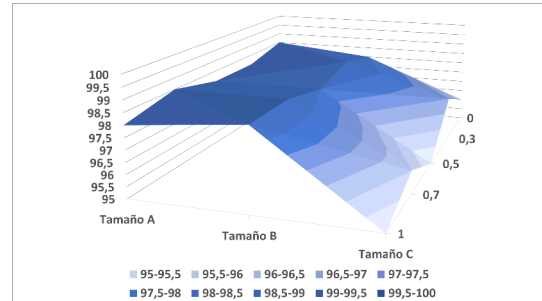


Fig. 5. Umbrales por tamaño y distancia, agregando con el mínimo.

tamaño relativo para acertar. No se aprecia dicha tendencia en el caso de la distancia. Los mismos datos, agregando los valores de umbral obtenidos en cada usuario con una media en lugar de con el mínimo, se muestran en la tabla III.

La figura 6 ilustra los datos de la tabla de manera gráfica. Se observan similares tendencias, pero menos acusadas, dado que la media limita el efecto de los usuarios con menor precisión en la percepción del tamaño.

Por último, se muestran los datos de dos usuarios a título de ejemplo. La tabla IV y la figura 7, muestran los datos de un usuario con alta precisión en la percepción del tamaño, mientras que la tabla V y la figura 8 muestran los datos de un usuario con menor precisión.

V. DISCUSIÓN

A la vista de la anterior experimentación, se pone de manifiesto que la *indistinguibilidad* es un elemento variable altamente dependiente del contexto. Por tanto, es importante que el sistema de generación de expresiones de referencia sea capaz de ajustar su funcionamiento a la capacidad de distinguibilidad propia del contexto en el que se ejecuta.

La experimentación que hemos desarrollado aquí, aunque simple, sirve para ilustrar que el ajuste de la indistinguibilidad es complejo. Nótese que la experimentación realizada utiliza solo tres variables para caracterizar el contexto, a saber, el *usuario*, el *tamaño* del objeto y la *distancia* entre ellos. El resto de variables se fijan: el *tipo* de objeto, el *color*, la *posición* en la pantalla, el *dispositivo*, ... por enumerar solo algunas variables de una lista que puede ser muy extensa dependiendo del sistema y del tipo de escenas con las que trabaje. Por citar solo un ejemplo, piénsese que se varía el color de los objetos: es bien conocido por todos que el color influye en la percepción del tamaño de los objetos.

TABLE III
 UMBRALES AGREGANDO CON LA MEDIA

Distancia (rel.)	Tamaño		
	A	B	C
1	98,7	99,1	98,0
0,7	99,2	99,3	97,8
0,5	99,2	99,0	97,8
0,3	98,7	99,3	98,5
0	99,2	99,0	98,1

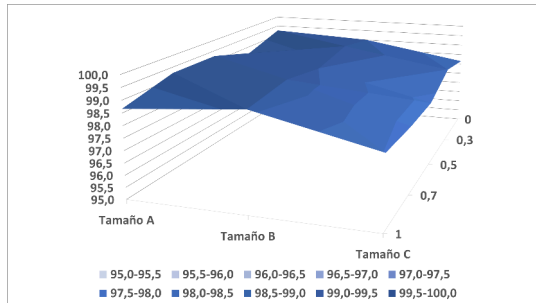


Fig. 6. Umbrales por tamaño y distancia, agregando con la media.

Por tanto, un sistema de generación de expresiones de referencia será mejor cuanto mayor sea el ajuste al contexto que tiene en cuenta para producir sus expresiones.

VI. CONCLUSIONES

Una de las contribuciones más reconocidas de la Teoría de Conjuntos Difusos en el ámbito de los sistemas *data-to-text* es el modelado de propiedades difusas. Este trabajo contribuye a facilitar dicha tarea proponiendo una metodología para determinar de forma automática la semántica de propiedades difusas en base al contexto y teniendo en cuenta la distinguibilidad de las propiedades de los objetos, que es uno de los problemas más arduos que surgen en este tipo de sistemas.

REFERENCES

- [1] E. Reiter and R. Dale, *Building Natural Language Generation Systems*. Cambridge, UK: Cambridge University Press, 2000.
- [2] N. Marín and D. Sánchez, "On generating linguistic descriptions of time series," *Fuzzy Sets and Systems*, vol. 285, pp. 6–30, 2016.
- [3] A. Ramos-Soto, A. Bugarín, and S. Barro, "On the role of linguistic descriptions of data in the building of natural language generation systems," *Fuzzy Sets and Systems*, vol. 285, pp. 31–51, 2016.
- [4] K. van Deemter, *Not Exactly: In Praise of Vagueness*. New York, NY, USA: Oxford University Press, Inc., 2012.
- [5] A. Gatt, N. Marín, F. Portet, and D. Sánchez, "The role of graduality for referring expression generation in visual scenes," in *Proceedings IPMU 2016, Part I, CCIS 610*, J. Carvalho, M.-J. Lesot, U. Kaymak, S. Vieira, B. Bouchon-Meunier, and R. R. Yager, Eds. Springer, 2016, pp. 191–203.
- [6] N. Marín, G. Rivas-Gervilla, and D. Sánchez, "An approximation to context-aware size modeling for referring expression generation," in *IEEE International Conference on Fuzzy Systems, FUZZ-IEEE*, 2018.
- [7] D. Dubois and D. Sanchez, "Fuzzy Clustering based on Coverings," in *Towards Advanced Data Analysis by Combining Soft Computing and Statistics*, ser. Studies in Fuzziness and Soft Computing, C. Borgelt, M. A. Gil, J. M. C. Sousa, and M. Verleysen, Eds. Springer, 2013, vol. 285, pp. 319–330.
- [8] D. Sánchez, M. Delgado, M. Vila, and J. Chamorro-Martínez, "On a non-nested level-based representation of fuzziness," *Fuzzy Sets and Systems*, vol. 192, pp. 159–175, 2012.

 TABLE IV
 UMBRALES PARA USUARIO CON ALTA PRECISIÓN EN LA PERCEPCIÓN DEL TAMAÑO

Distancia (rel.)	Tamaño		
	A	B	C
1	98,7	99,1	98,0
0,7	99,2	99,3	97,8
0,5	99,2	99,0	97,8
0,3	98,7	99,3	98,5
0	99,2	99,0	98,1

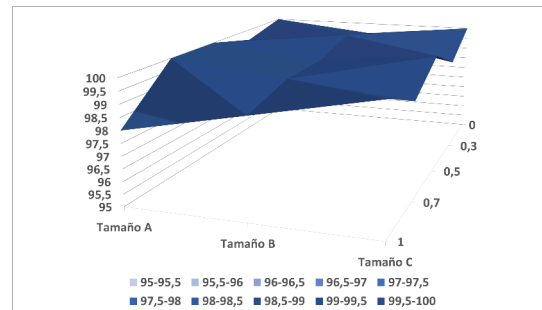


Fig. 7. Umbrales para usuario con alta precisión en la percepción del tamaño.

 TABLE V
 UMBRALES PARA USUARIO CON MENOR PRECISIÓN EN LA PERCEPCIÓN DEL TAMAÑO

Distancia (rel.)	Tamaño		
	A	B	C
1	98,7	99,1	98,0
0,7	99,2	99,3	97,8
0,5	99,2	99,0	97,8
0,3	98,7	99,3	98,5
0	99,2	99,0	98,1

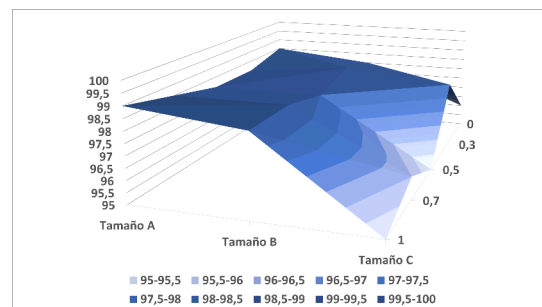


Fig. 8. Umbrales para usuario con menor precisión en la percepción del tamaño.