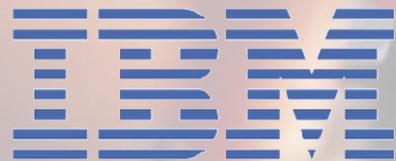


# Presione 1 si es correcto

Interpretación basada en **corpus** de  
instrucciones en **entornos virtuales**



IBM Argentina  
Grupo Silvergate



Univ. Nac. de Córdoba  
Famaf

¿Para qué queremos  
**dialogar** con  
nuestros sistemas?



# Necesitamos **diálogo** ...

*... para **cerrar brechas***

*... para **humanizar** nuestros sistemas*

*... porque **tiene sentido tenerlo***

*... porque **no sabemos a dónde nos puede llevar***

¿Qué es un  
**corpus**?



¿Qué es un  
**entorno virtual?**



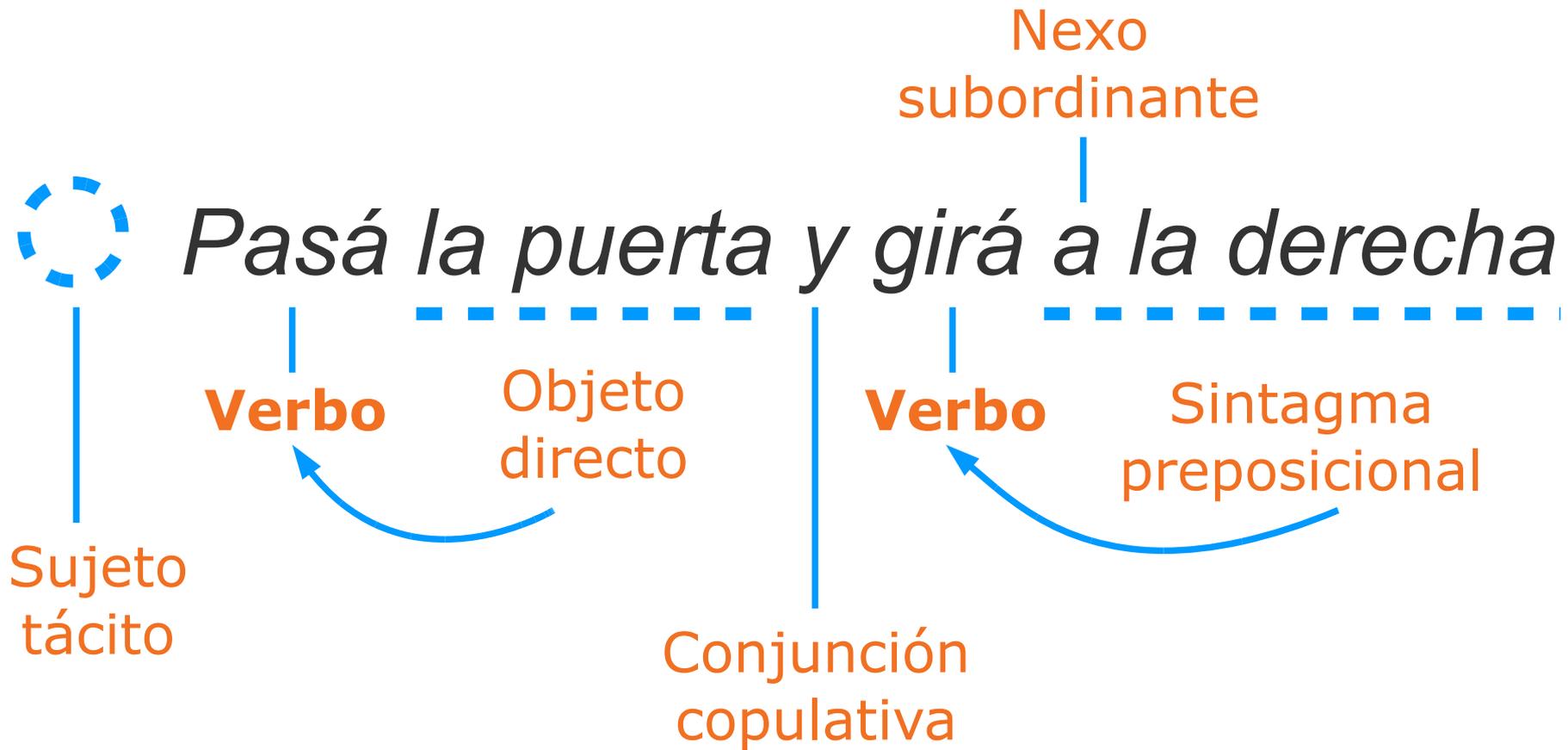


Hablemos de los  
**enfoques más comunes**  
para interpretación de  
instrucciones

# Enfoques **simbólicos**

*Pasá la puerta y girá a la derecha*

# Enfoques **simbólicos**



# Enfoques **simbólicos**

*Frená antes de llegar al botón  
junto al cuadro y apretalo*

*Dásela*

*Callate la boca y respondeme*

*Toma el ratón*

# Enfoques **estadísticos**



como hago |

como hago **para olvidarte**  
como hago **mi biografia en facebook**  
como hago **la ene en mac**  
como hago **para que me llame**

Google Search I'm Feeling Lucky

Requieren **extensa anotación manual** en la fase de entrenamiento

# Nuestro enfoque

Basado en corpus

Anotación automática

Tolerante a errores  
(*ortografía, gramática, etc*)



Parte 1

Nuestras **bases**

¿Qué es el

**give**  
challenge

?



Move forward 2 steps, and then press the blue button.



Walk Turn

↑	↓	→	←	Esc quit	H What?
---	---	---	---	-------------	------------

# Datos del **corpus**

**6** mundos virtuales

**100** juegos

**5580** instrucciones

**14:26hs.** de interacción

Dividido en dos grandes ramas



En el **Corpus Cm**,  
ambos participantes son  
seleccionados al azar

En el **Corpus Cs**, las instrucciones son dadas por una sola persona



# Pregunta:

¿Cómo hacemos  
para trabajar con  
datos **continuos**?



# Respuesta:



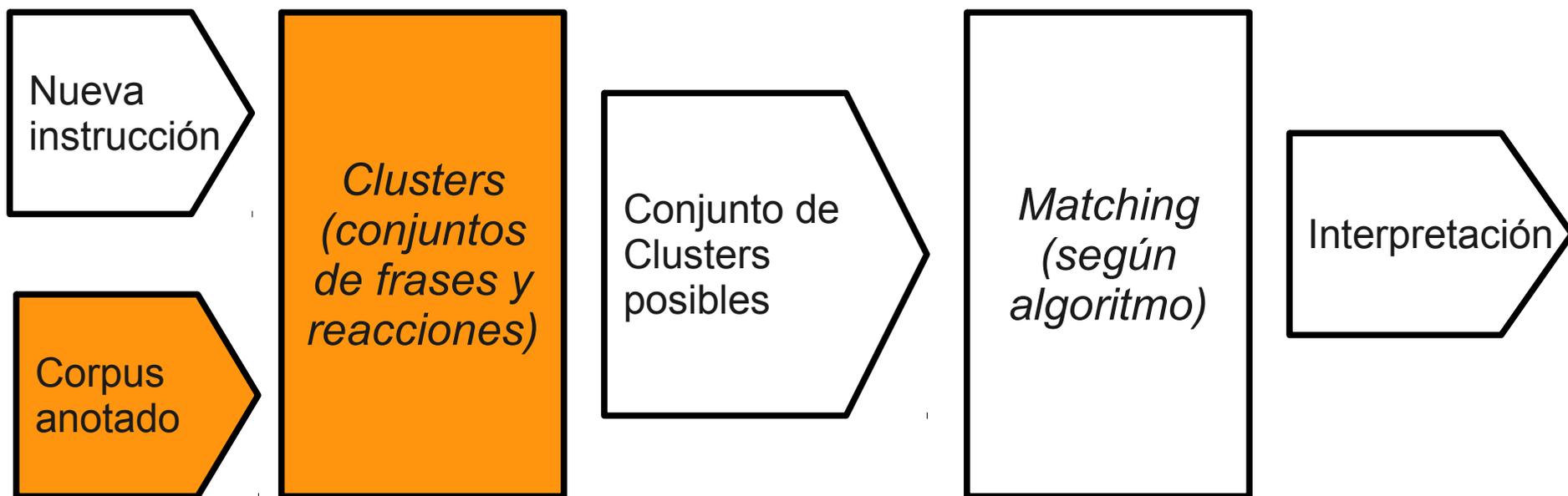
No lo hacemos.  
Los discretizamos  
con un **planner**

## Parte 2

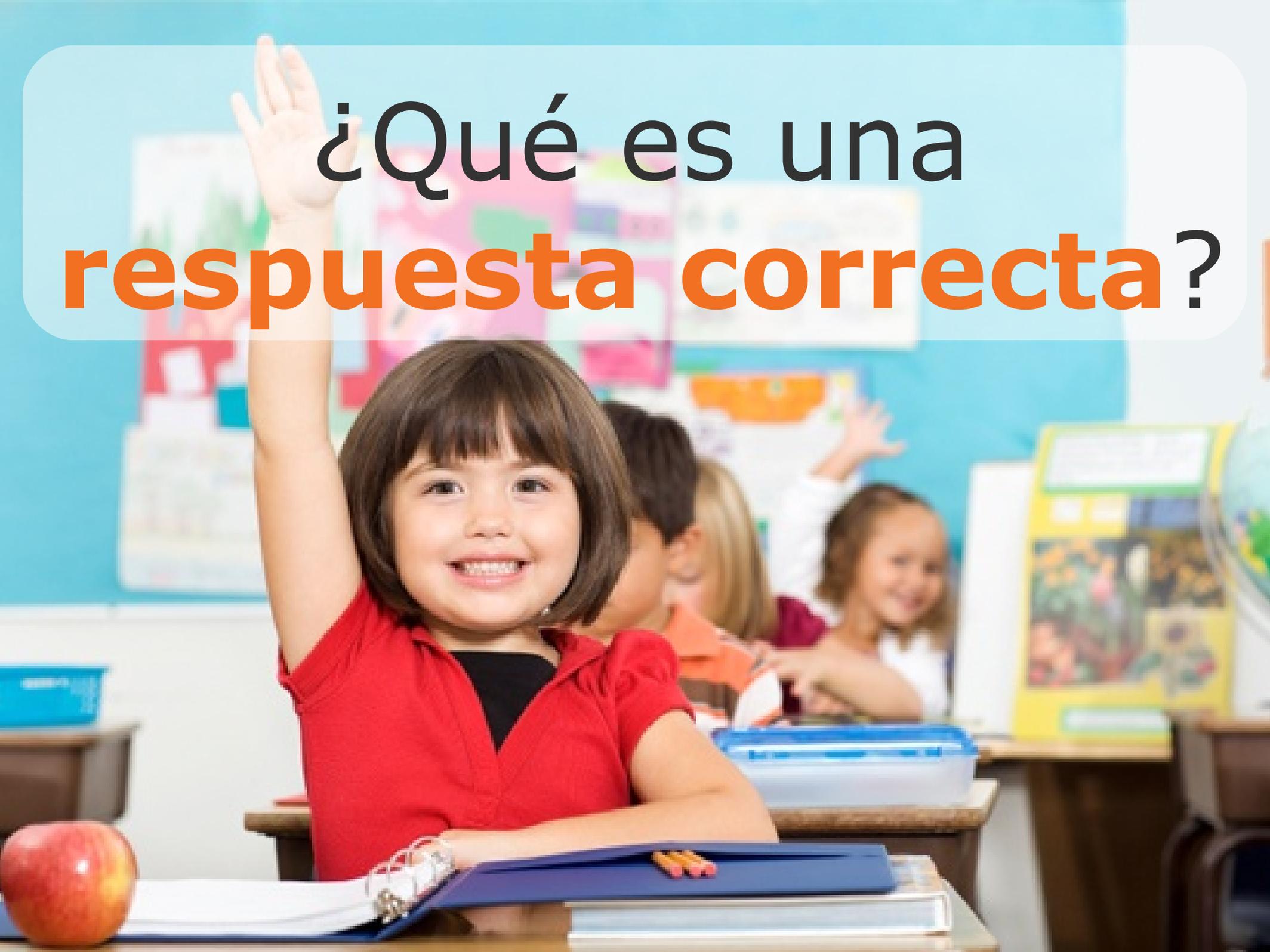
En busca de  
una **solución**



# Nuestra **arquitectura**



¿Qué es una  
**respuesta correcta?**



# Anotación **automática**

**Presiona el botón rojo en la habitación más lejana**

*[entra en la habitación]*

*[presiona el botón rojo]*

*[gira a la derecha]*

**Verde a junto la lámpara en habitación a tu derecha**

*[va a la habitación a su derecha]*

*[presiona el botón verde]*

# Anotación basada en **comportamiento**

**Presiona el botón rojo en la habitación más lejana**

*[entra en la habitación]*

*[presiona el botón rojo]*

*[gira a la derecha]*

**Verde a junto la lámpara en habitación a tu derecha**

*[va a la habitación a su derecha]*

*[presiona el botón verde]*

# Anotación basada en **visibilidad**

**Presiona el botón rojo en la habitación  
más lejana**

*[entra en la habitación]*

*[presiona el botón rojo]*

*[gira a la derecha]*

**Verde a junto la lámpara en habitación  
a tu derecha**

*[va a la habitación a su derecha]*

*[presiona el botón verde]*

Está claro que el  
método no es **perfecto**

**Presiona el botón verde**

*[se para junto al botón equivocado]*

**No**

*[se mueve junto al correcto]*

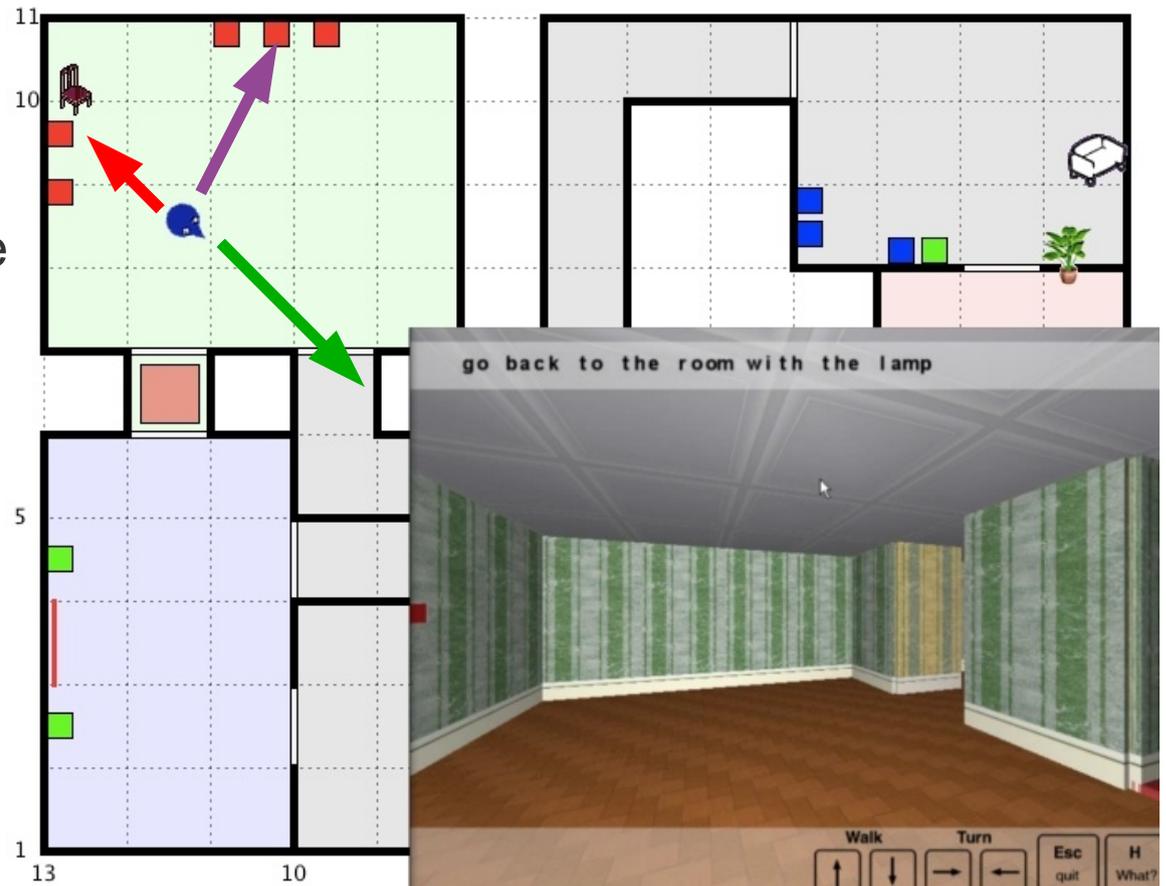
*[presiona el botón correcto]*

Ahora que tenemos **reacciones canónicas**, podemos clasificar instrucciones:

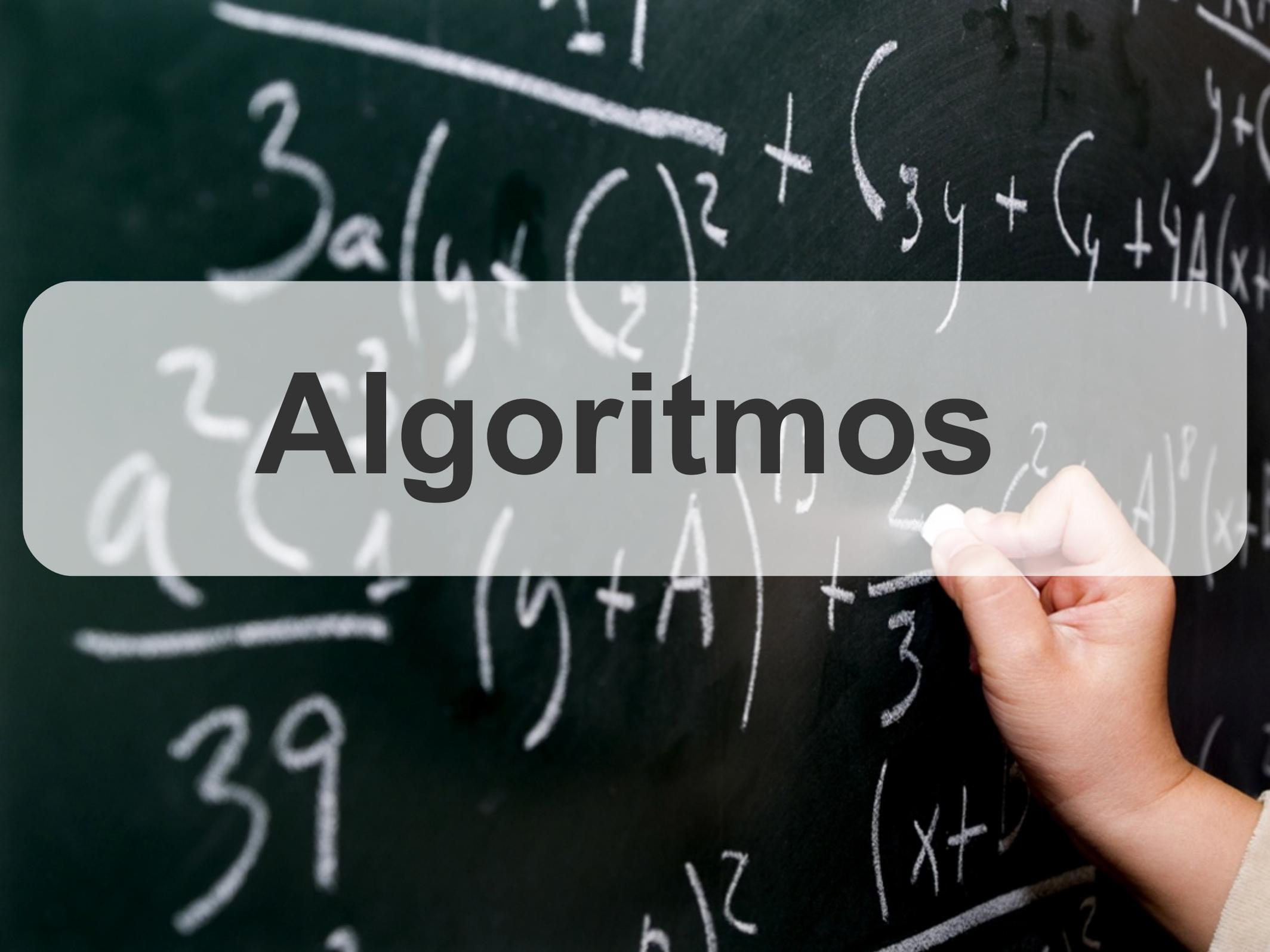
Si dos instrucciones tienen **la misma reacción** en el **mismo lugar**, pertenecen al mismo **grupo** o **cluster**

# Gráficamente, luce **así**

- Red left of chair  
Press the one behind u
- Press middle button  
In group of 3, press middle one  
Counting right, 2nd one
- To the room with lamp  
Go back to the big room  
Go back to the hallway  
Go back out the room  
Down the passage  
Go back out  
Yes  
Out the way you came in  
Exit the way you entered  
Go through the opening on the left with the yellow wall paper



# Algoritmos

A hand is shown writing on a chalkboard. The board is covered with various mathematical expressions and equations, including algebraic formulas like  $3a(y+z)^2 + (3y + (y + 4)A(x + ...))$  and numbers like 39. The word "Algoritmos" is prominently displayed in a large, bold, black font within a semi-transparent white rounded rectangle in the center of the image.

# Índices de **semejanza**

**Jaccard**

$$\frac{|A \cap B|}{|A \cup B|}$$

**Overlap**

$$\frac{|A \cap B|}{\min(|A|, |B|)}$$

# Índices de **distancia**

## **Levenshtein**

(sobre letras y sobre frases)

**Ejemplo:** Campo → Ciudad: 6

(Campo → Ciampo → Ciuampo → Ciudampo  
→ Ciudadpo → Ciudadp → Ciudad)

# Otros índices

## **BLEU**

Chequeo de traducciones automáticas  
(Machine Translation)

## **Support Vector Machines**

Clasificación lineal binaria  
(Machine Learning)

A large white cruise ship is shown listing heavily to its left in the ocean. The ship's upper decks and funnel are visible, and the water is dark blue. The scene is set against a clear sky.

**¿Y si pudiéramos  
corregir errores?**

# Reintentos

Si la predicción es **incorrecta**, permitimos volver a intentarlo

(no más de 3 veces)

## Parte 3

¿Y cuáles fueron  
los **resultados**?



# Resultados **generales**

**70%** de efectividad con el  
mejor algoritmo  
(y eso sin corregir errores)

**Chen & Mooney** (2011) lograron un 56%  
**Gorniak** (2007) logró un 72%  
con anotación manual

Los **humanos** consiguen usualmente un 69%

***Pregunta:***

¿Cuál es el **mejor**  
algoritmo?

# Resultados **generales**

Los algoritmos de semejanza (**Jaccard** y **Overlap**) obtuvieron los mejores resultados

**BLEU** y **Levenshtein** (*por palabras*)  
obtuvieron resultados decentes

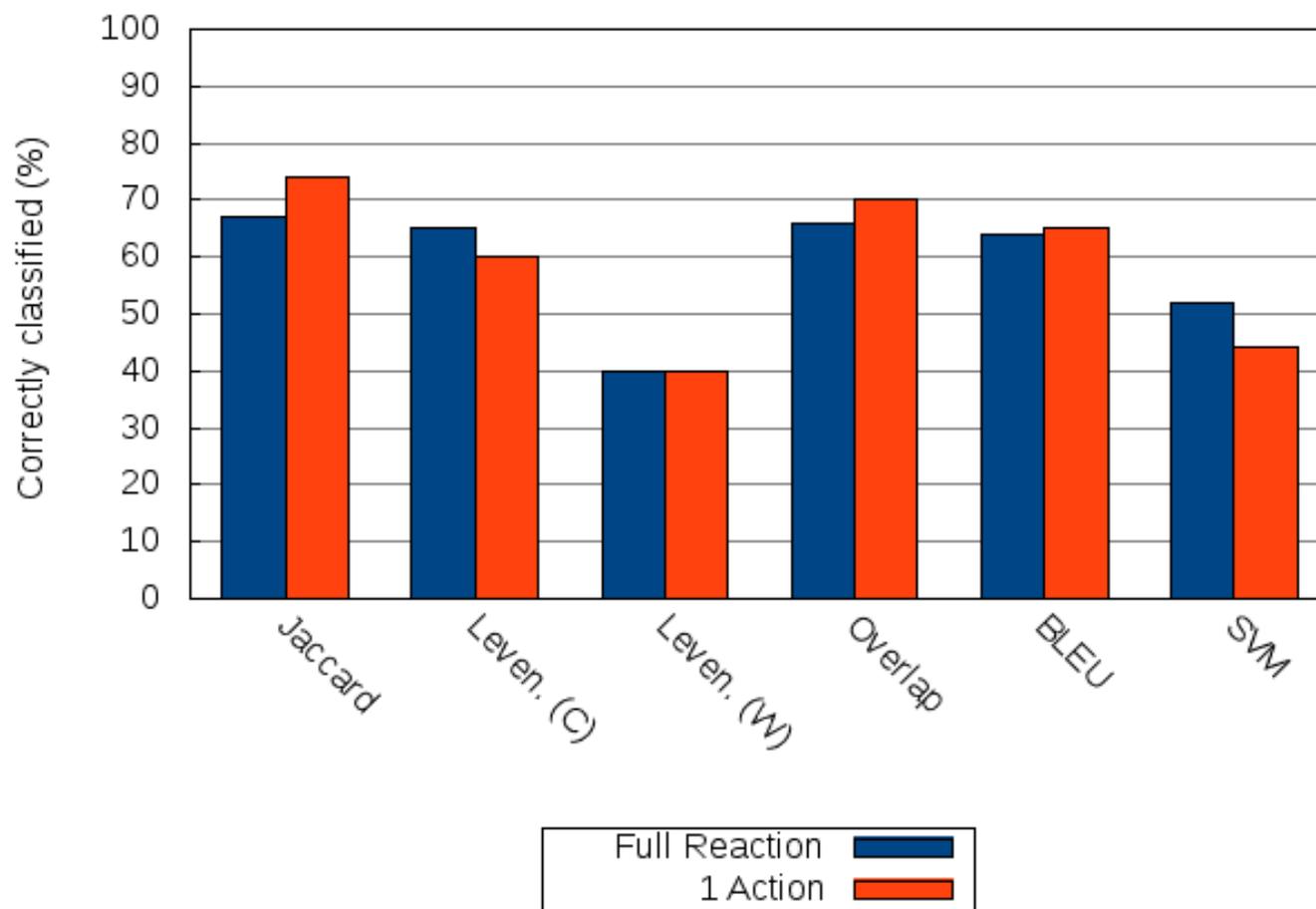
**SVM** no obtuvo buenos resultados  
(hipótesis: training desbalanceado)

**Levenshtein** (*por letras*) no es  
mucho mejor que random

# ***Pregunta:***

¿Qué es mejor, clasificar  
por **visibilidad** o por  
**comportamiento**?

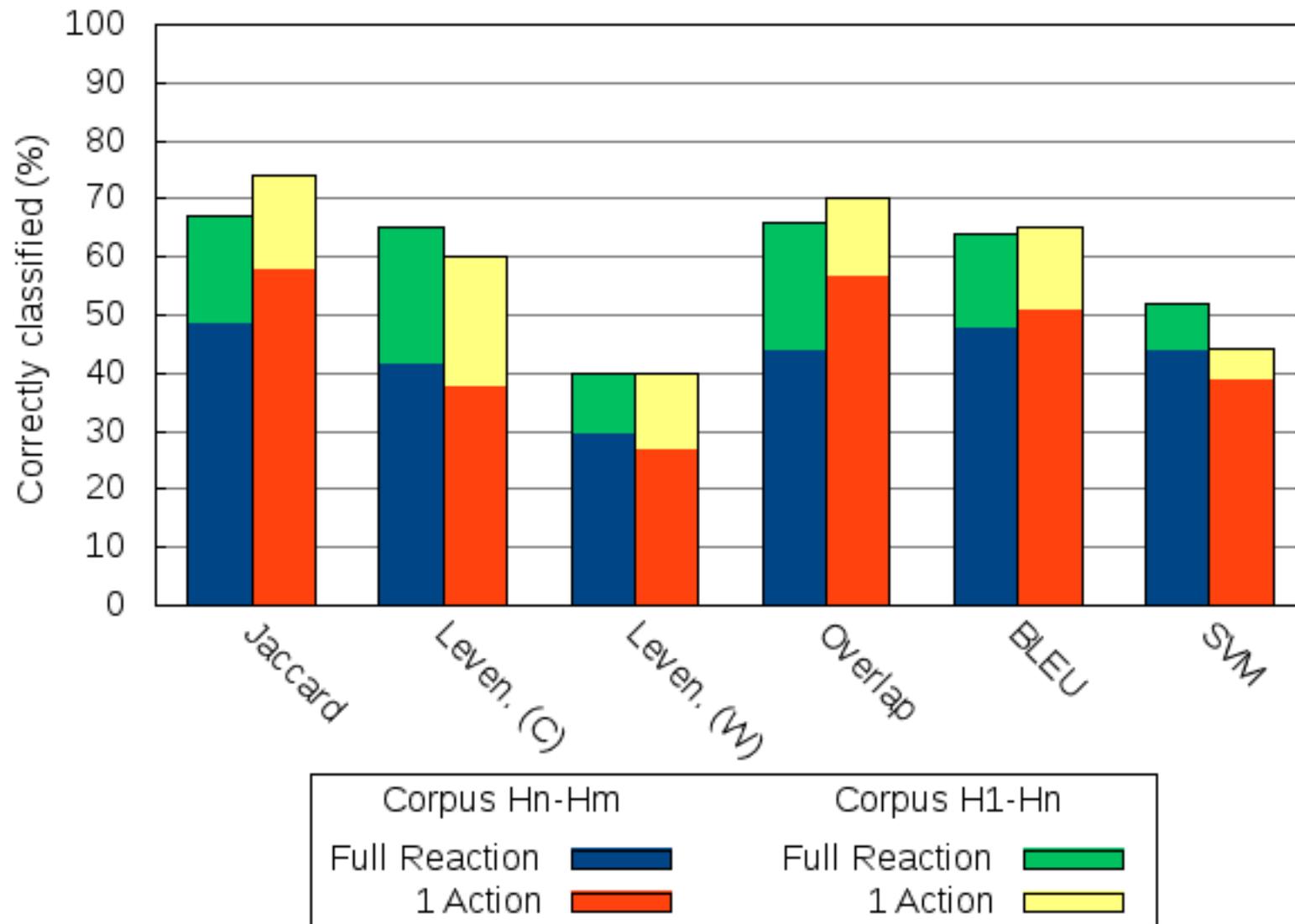
Algorithm vs Reaction clustering  
Corpus H1-Hn



***Respuesta:*** No hay diferencias significativas

# ***Pregunta:***

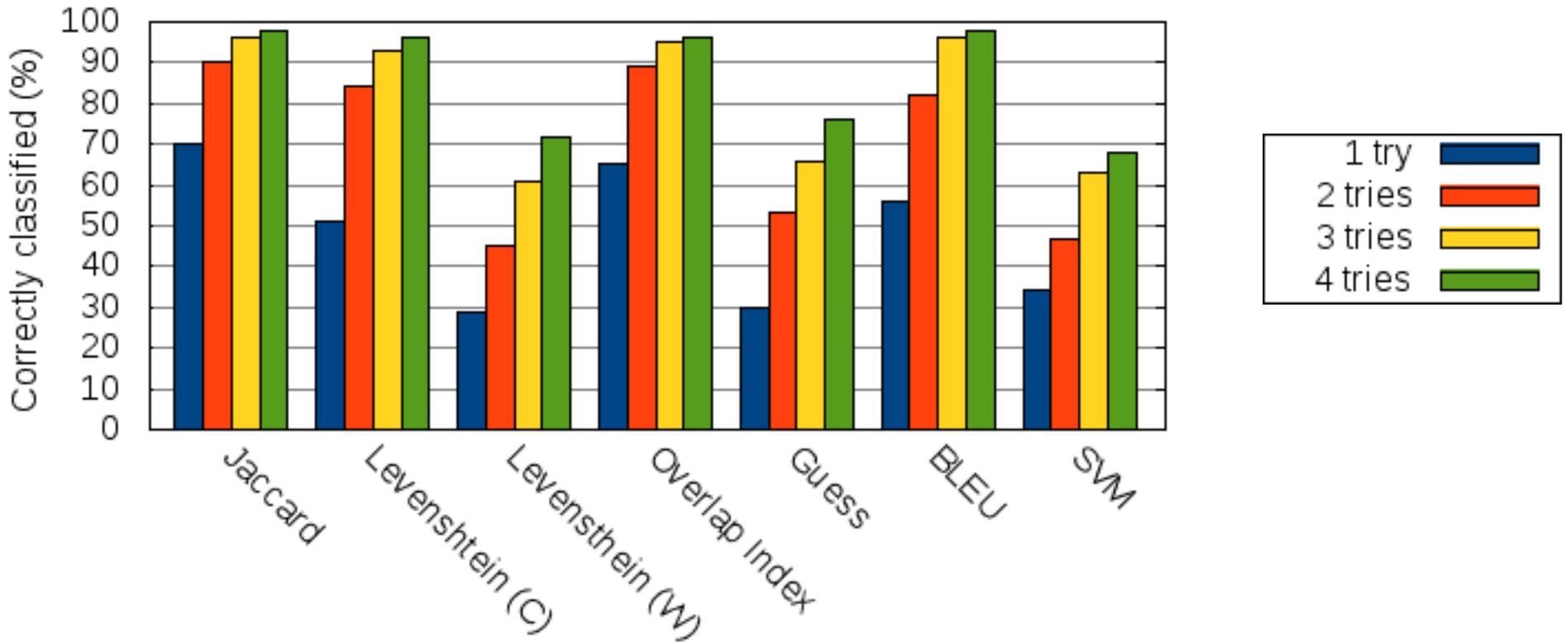
¿Es relevante que las instrucciones provengan o no de **una misma persona?**



***Respuesta:*** los datos **Cs** obtuvieron mejores resultados en **todos** los tests

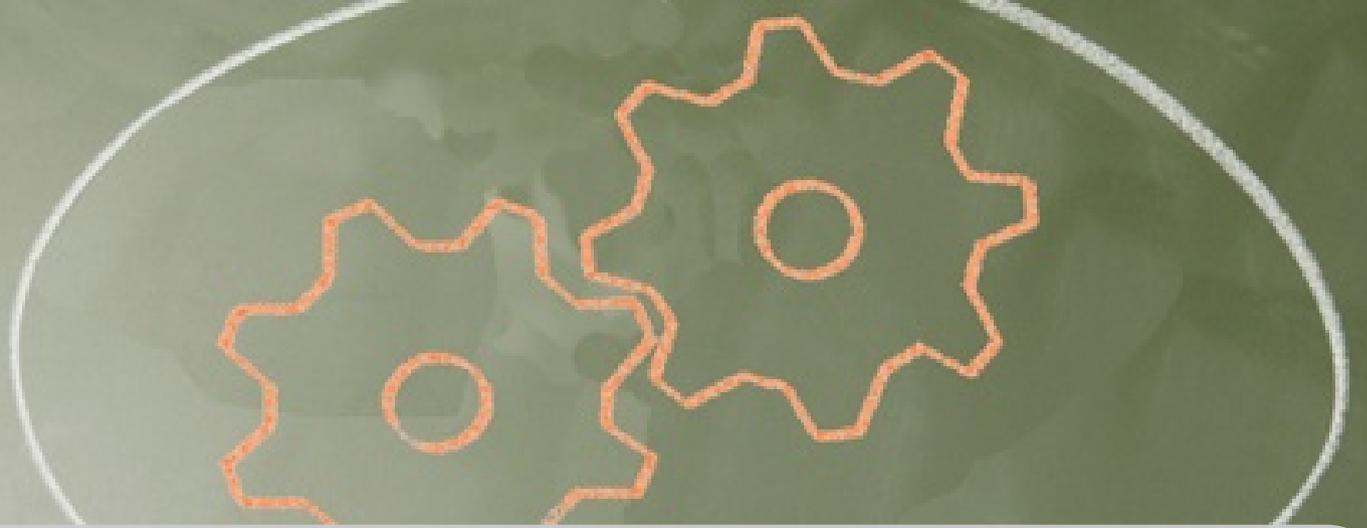
# ***Pregunta:***

¿En qué medida  
mejoran los  
resultados si permito  
**corregir errores?**



***Respuesta:*** Un **buen** algoritmo raramente necesita más de 2 intentos.

Un **mal** algoritmo mejora linealmente.



**Ahora es su turno:**

**¿Preguntas?**



# Gracias por su atención

Tessa **Lau**

tessalau@us.ibm.com

Julián **Cerruti**

jceruti@ar.ibm.com

Luciana **Benotti**

luciana.benotti@gmail.com

Martín **Villalba**

villalba@famaf.unc.edu.ar