

Creación de Mapas Semántico-Topológicos Para Robótica Móvil Utilizando Modelos de Aprendizaje Profundo

Building Semantic-Topological Maps for Mobile Robotics Using Deep Learning Models

Realizado por:

Jesús Moncada Ramírez

Tutorizado por:

José Raúl Ruiz Sarmiento
Javier González Jiménez

Máster en Ingeniería Mecatrónica

Málaga, julio de 2025



UNIVERSIDAD DE MÁLAGA



1

Introducción

Motivación -----

1. Introducción

Los robots móviles se usan cada vez más en tareas complejas...



Necesidad:
CAPACIDADES COGNITIVAS

Mapas semánticos (*3D scene graphs*)

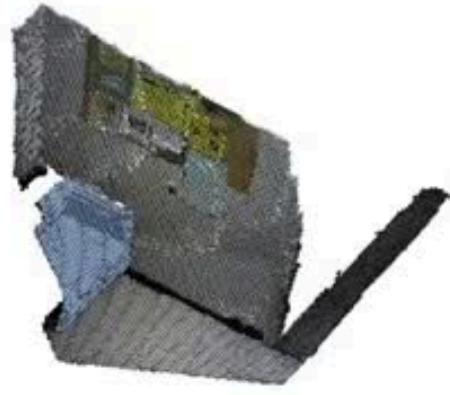
Motivación -----

1. Introducción

Mapas con información...

Geométrica

Formas y contornos de objetos,
posiciones, dimensiones, escala...



Semántica

Categorías de objetos (p.ej. “silla”,
“mesa”, etc.)

(Matez-Bandera et al. 2024)

Motivación

1. Introducción

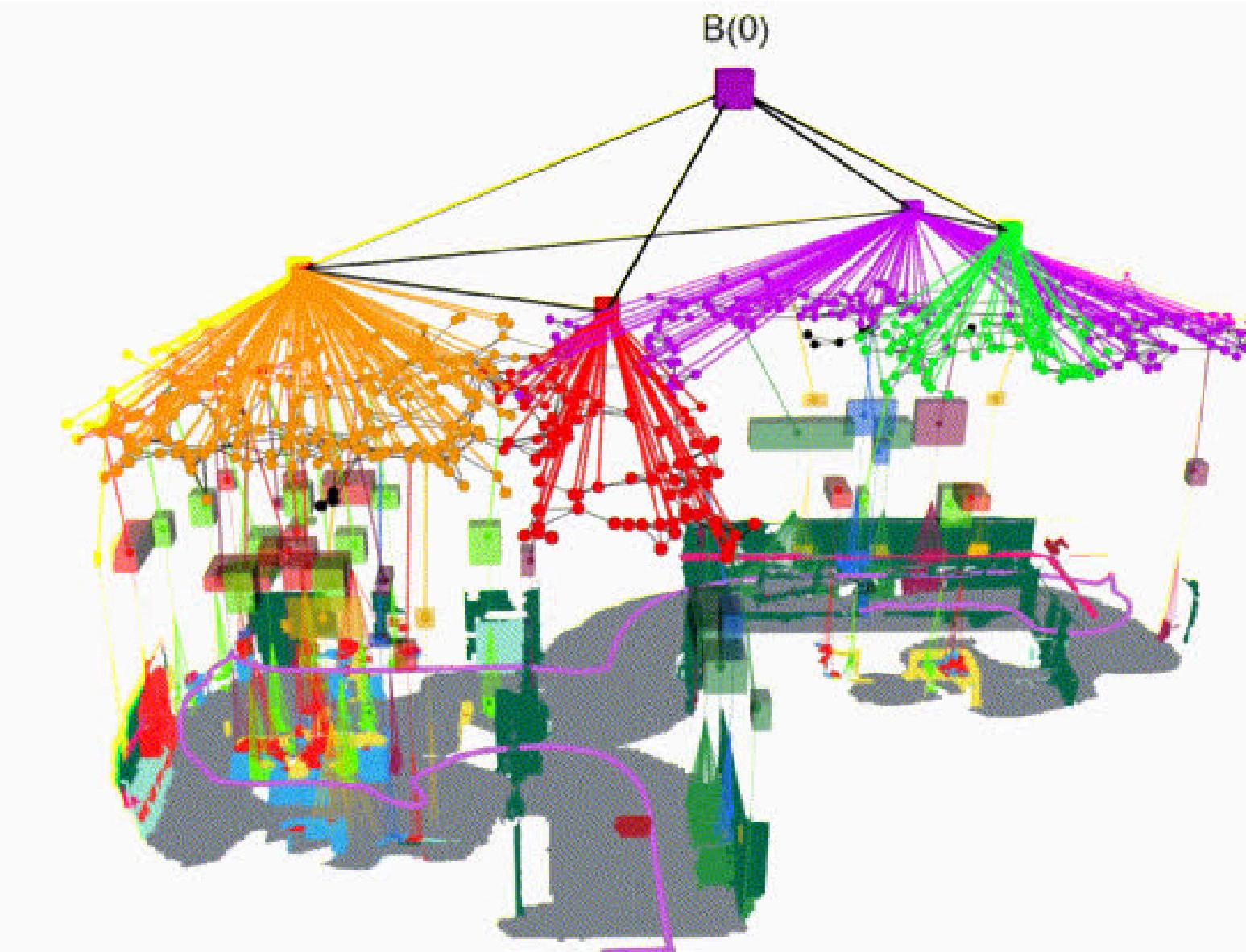
...que se pueden enriquecer con información...

Topológica

Lugares significativos del entorno

Relaciones entre objetos y lugares

Mapa **semántico-topológico**



(Hughes et al. 2022)

Motivación

1. Introducción

Mapas
semántico-topológicos
tradicionales

Solución
Modelos de gran escala
*Large Language Models (LLMs) o
Large Vision-Language Models (LVLMs)*

- (creación) ● Anotación **manual** (tiempo)
(errores)
- (creación) ● Vocabulario **cerrado** (detección)
("habitación")
- (uso) ● **Falta** validación de utilidad
- (creación) ● Proceso **automático**
- (creación) ● Vocabulario **abierto**
- (uso) ● **Validación** de utilidad

Objetivos

1. Introducción

Explorar cómo los **modelos de gran escala generativos** pueden ayudar a la creación de mapas semántico-topológicos en robótica móvil

ESTADO DEL ARTE

LUGARES

Métodos de:

Segmentación + categorización de lugares

- Vocabulario abierto
- No “habitaciones”

RELACIONES

Métodos de:
Inferencia de relaciones semánticas

- Vocabulario abierto

VALIDACIÓN

- De los **métodos** propuestos
- Del **desempeño** de un robot equipado con un **mapa semántico-topológico** y un **LLM** como motor de razonamiento

Publicaciones

1. Introducción

Uno de los métodos propuestos para segmentación + categorización de lugares



Mobile Robot Place Segmentation and Categorization Using Object Semantics

Jesus Moncada-Ramirez¹, Jose-Raul Ruiz-Sarmiento¹, Cipriano Galindo¹ and Javier Gonzalez-Jimenez¹

Abstract—A mobile robot's ability to perform complex tasks depends critically on its understanding of the environment in which it operates. This understanding—often represented as a map—must incorporate both semantic information about objects (identity, functionality, ...) and a high-level grasp of spatial organization. This latter capability, referred to as place segmentation and categorization, is essential for reasoning in diverse and dynamic spaces. However, conventional semantic mapping methods typically segment environments into predefined rooms and rely on closed-set labels, limiting adaptability and expressiveness in real-world, multi-functional settings.

In this paper, we propose a novel pipeline for indoor place segmentation and categorization that leverages rich object semantics to generate more flexible and meaningful spatial representations. Rather than adhering to rigid architectural boundaries, our approach defines places as clusters of functionally coherent objects. These clusters are formed using descriptors that integrate geometric location with semantic information, derived from context-aware functional descriptions of objects generated by Large Language Models (LLMs) and encoded as sentence embeddings. We apply density-based clustering to identify semantically meaningful places and use LLMs to assign each cluster open-set natural-language tags and detailed functional descriptions, resulting in enriched and interpretable maps.

Experimental results on the ScanNet and SceneNN datasets demonstrate that our method significantly improves segmentation consistency over geometric and basic semantic baselines while producing coherent and informative place categorizations that better reflect the functional layout of indoor environments.

I. INTRODUCTION

ACEPTADO

Evaluación del desempeño de un robot equipado con un LLM basada en...



Article

Agentic Workflows for Improving Large Language Model Reasoning in Robotic Object-Centered Planning

Jesus Moncada-Ramirez , Jose-Luis Matez-Bandera , Javier Gonzalez-Jimenez and Jose-Raul Ruiz-Sarmiento

Machine Perception and Intelligent Robotics Group (MAPIR-UMA), Malaga Institute for Mechatronics Engineering and Cyber-Physical Systems (IMECH-UMA), University of Malaga, 29071 Málaga, Spain; jemonra@uma.es (J.M.-R.); josematez@uma.es (J.-L.M.-B.); javiergonzalez@uma.es (J.G.-J.)

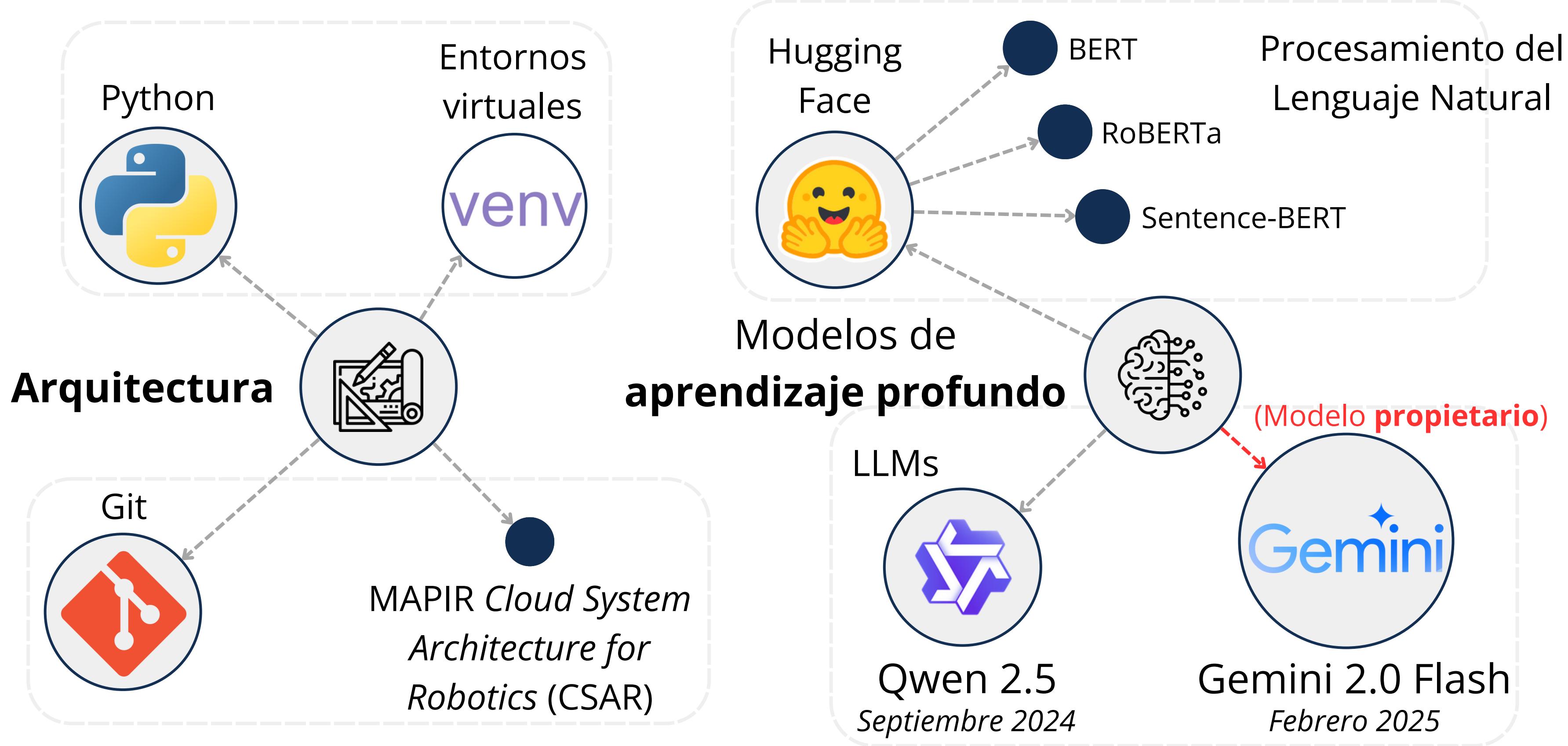
* Correspondence: jotaraul@uma.es

Abstract: Large Language Models (LLMs) provide cognitive capabilities that enable robots to interpret and reason about their workspace, especially when paired with semantically rich representations like semantic maps. However, these models are prone to generating inaccurate or invented responses, known as hallucinations, that can produce erratic robotic operation. This can be addressed by employing agentic workflows, structured processes that guide and refine the model's output to improve response quality. This work formally defines and qualitatively analyzes the impact of three agentic workflows (LLM Ensemble, Self-Reflection, and Multi-Agent Reflection) on enhancing the reasoning capabilities of an LLM guiding a robotic system to perform object-centered planning. In this context, the LLM is provided with a pre-built semantic map of the environment and a query, to which it must respond by determining the most relevant objects for the query. This response can be used in a multitude of downstream tasks. Extensive experiments were carried out employing state-of-the-art LLMs and semantic maps generated from the widely-used datasets ScanNet and SceneNN. The results show that agentic enhance object retrieval performance, especially in scenarios requiring

PUBLICADO

Tecnologías usadas

1. Introducción



2

Trabajos relacionados

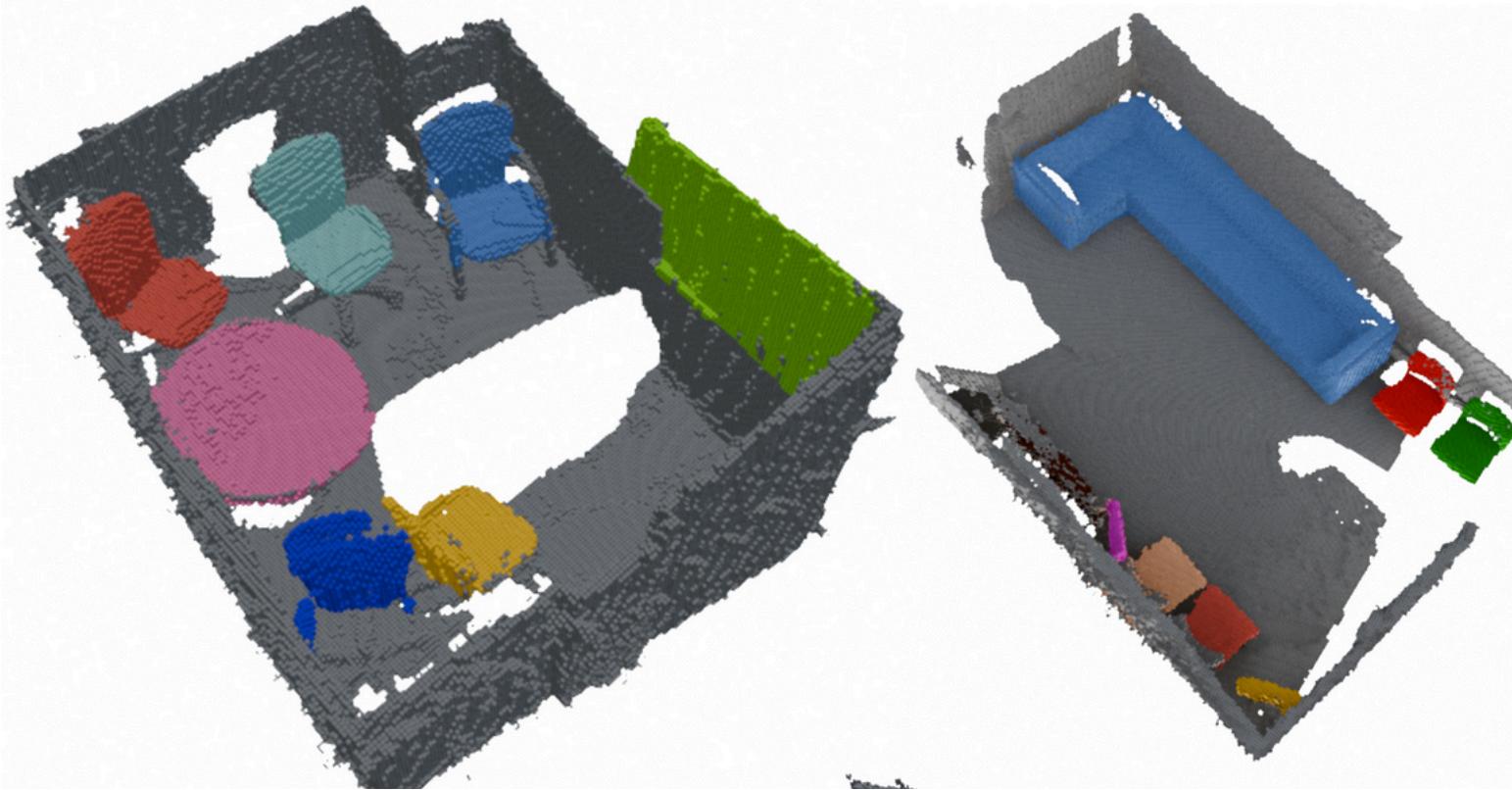
Mapeo semántico

2. Trabajos relacionados

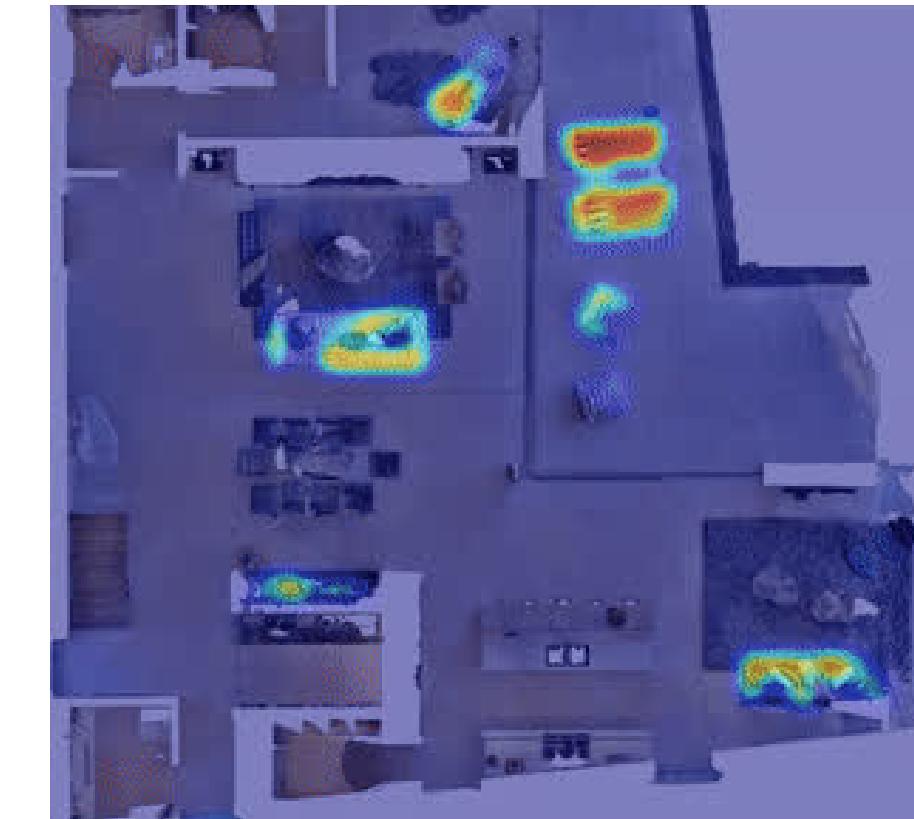
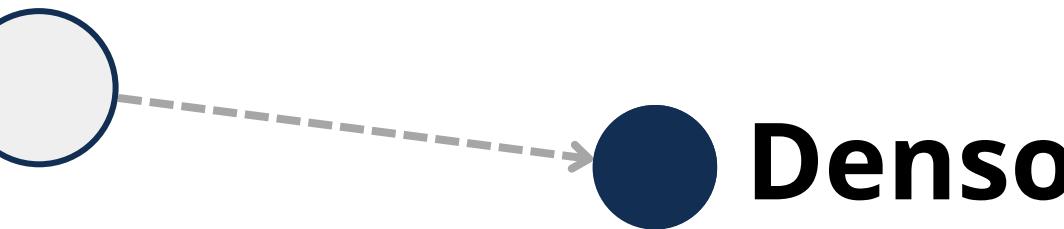
Mapeo semántico

Vincula info. **geometría + semántica**

A nivel de **instancia**



(Matez-Bandera et al. 2023)

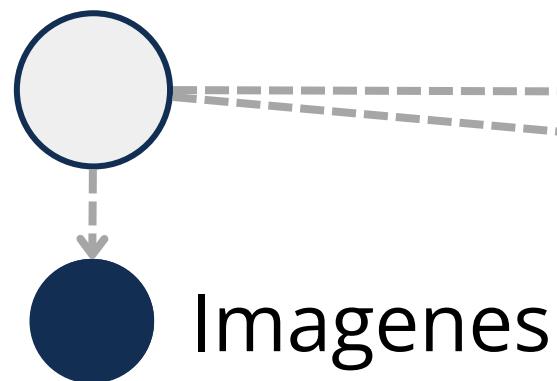


"the area of counter"
"the area of sofa"
"the area of chair"
"the area of stairs"
"the area of floor"

(Huang et al. 2023)

Segmentación y categorización de lugares

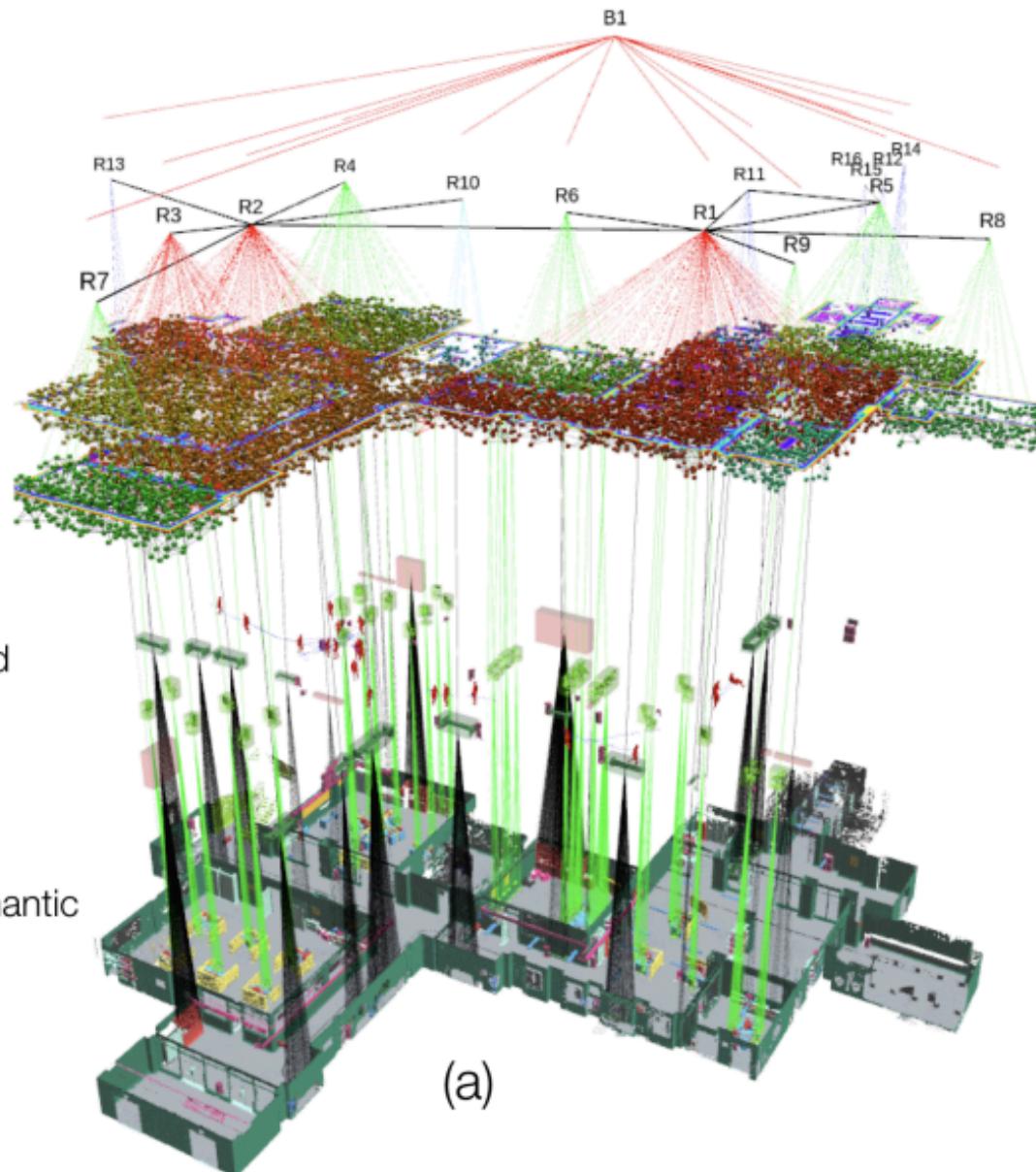
2. Trabajos relacionados



Layer 5:
Buildings
Layer 4:
Rooms
Layer 3:
Places and Structures
Layer 2:
Objects and Agents
Layer 1:
Metric-Semantic Mesh

Representaciones densas

(Rosinol et al. 2020)



Geometría (+ objetos)

Limitación

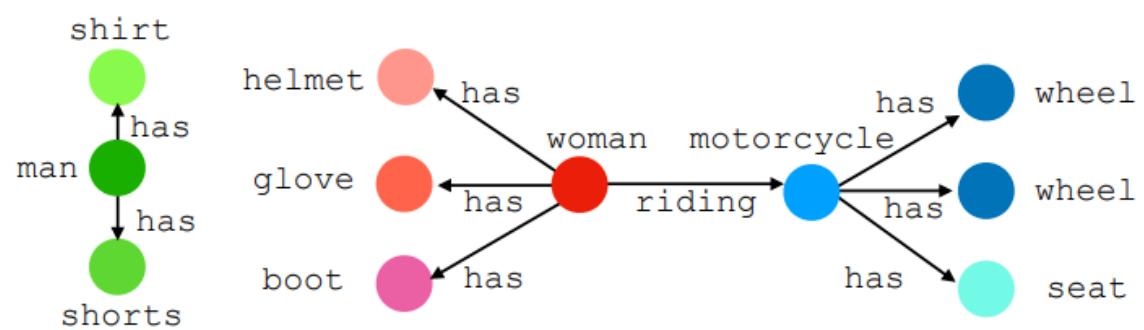
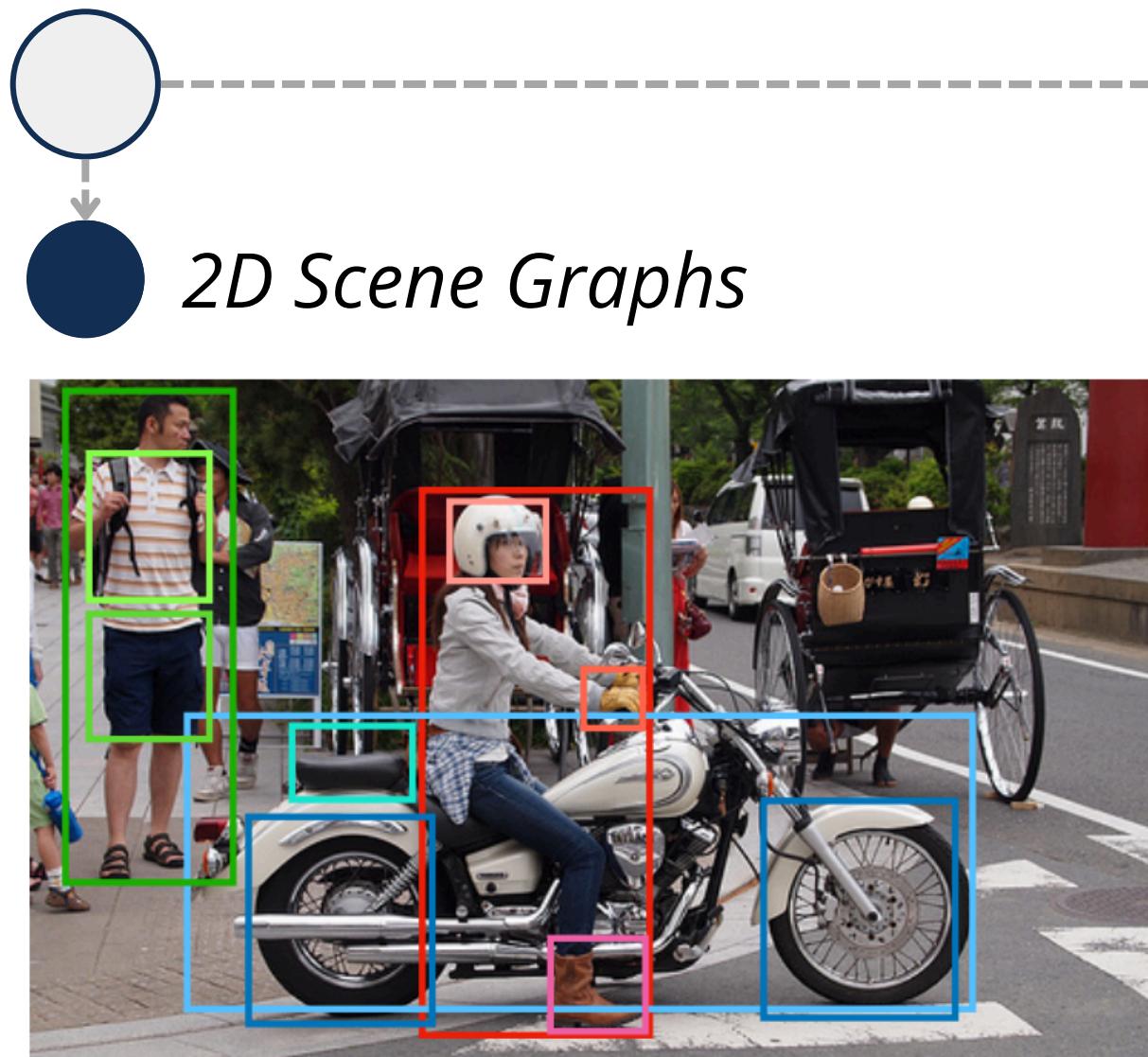
Enfoques basados en **habitaciones**

Affordances

¿Qué **acciones** puede realizar el agente en el entorno?

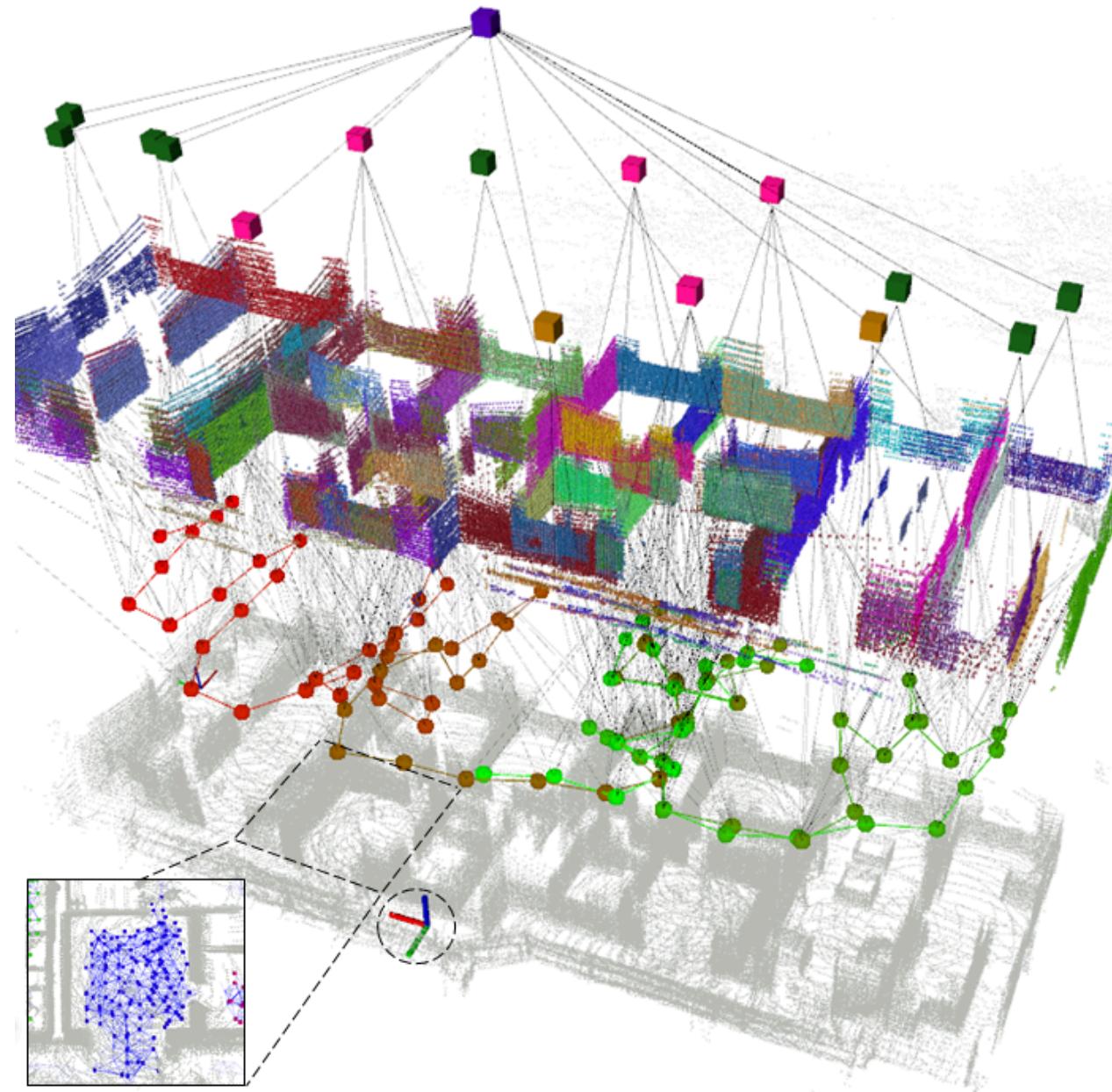
Inferencia de **relaciones** semánticas

2. Trabajos relacionados



(Zellers et al. 2018)

3D Scene Graphs



(Bavle et al. 2022)

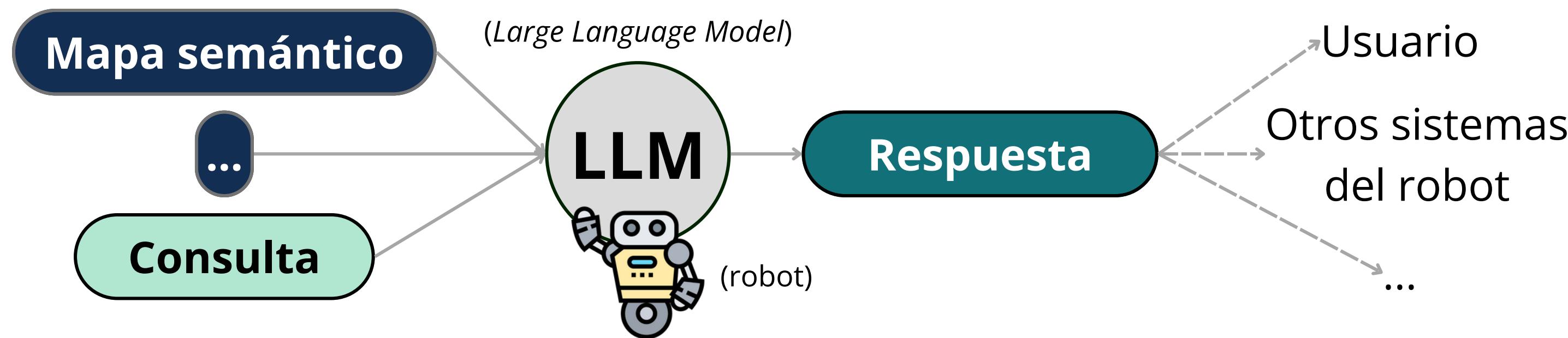
Limitación

Vocabulario **cerrado**
Dominios **específicos**

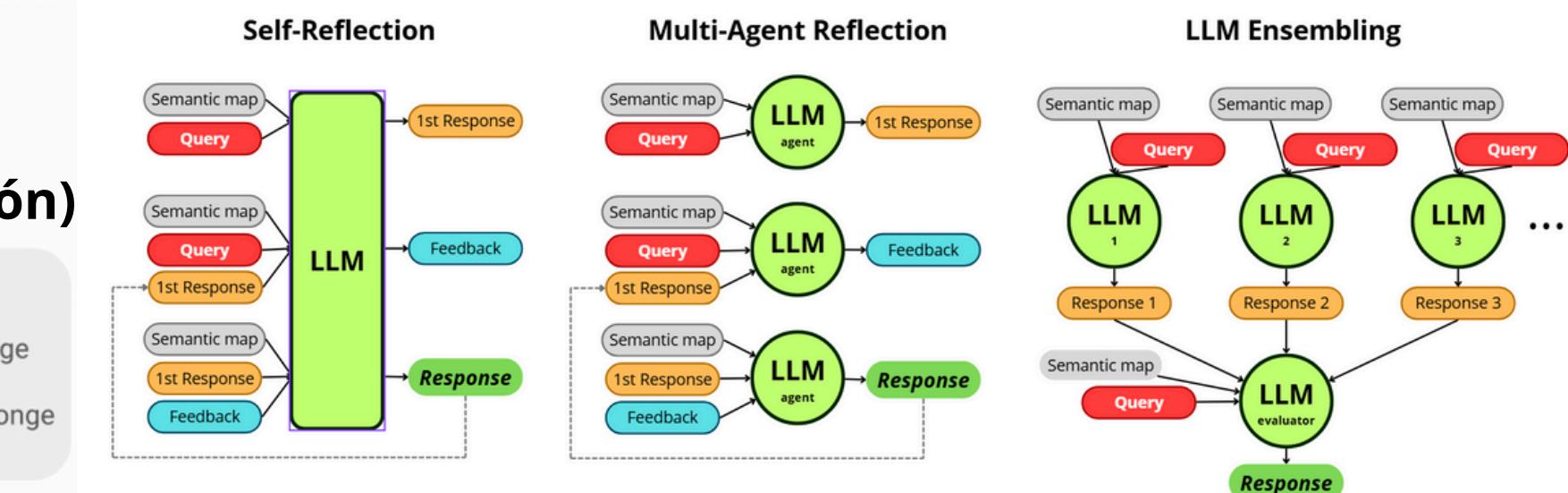
LLMs
Affordances

Explotación de mapas semánticos con LLMs

2. Trabajos relacionados



(Ahn et al. 2022)



(Moncada et al. 2025)

Codificación del mapa

Limitación

(LLM) Alucinaciones

(LLM) Ventanas de contexto

3

Segmentación y categorización de lugares

Formulación del problema

3. Lugares

Mapa semántico

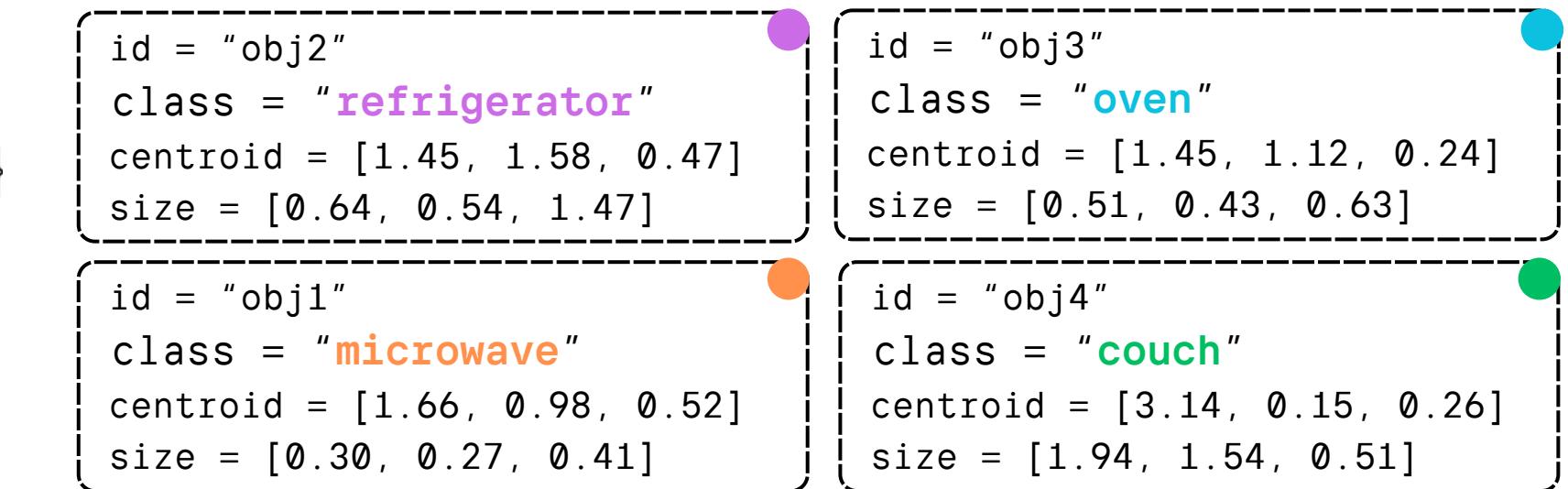
$$\mathbf{m} = \{\mathbf{o}_1, \dots, \mathbf{o}_N \mid \mathbf{o}_i = (c_i, d_i, l_i, \mathbf{p}_i^{\text{object}}), \forall i \in \{1, \dots, N\}\}$$

Centroide

Tamaño

Clase semántica

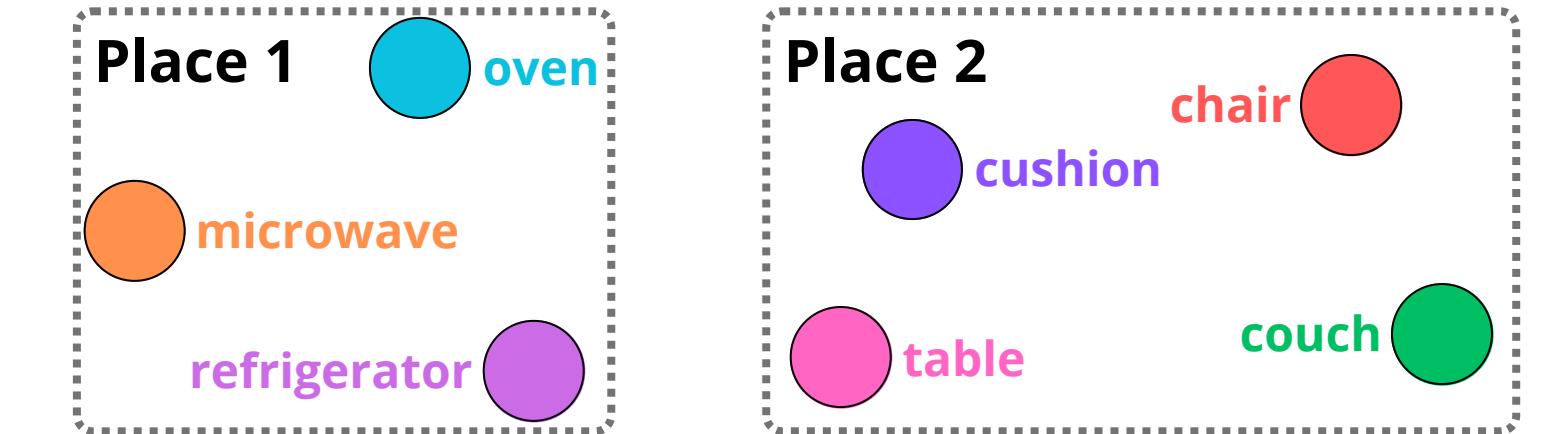
Propiedades



Segmentación de lugares

$$\mathcal{P} = \psi(\mathbf{m}) = \{P_k \mid k \in K\}$$

donde cada lugar $P_k = \{\mathbf{o}_j \mid j \in J_k \subseteq \{1, \dots, N\}\}$



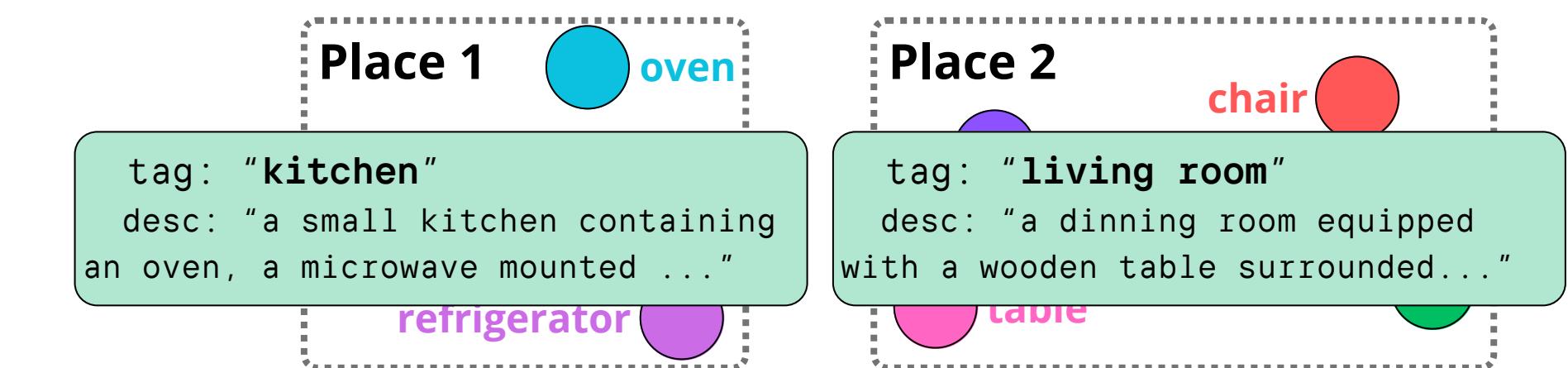
Categorización de lugares

$$\phi(P_k) = \mathbf{p}_k^{\text{place}}$$

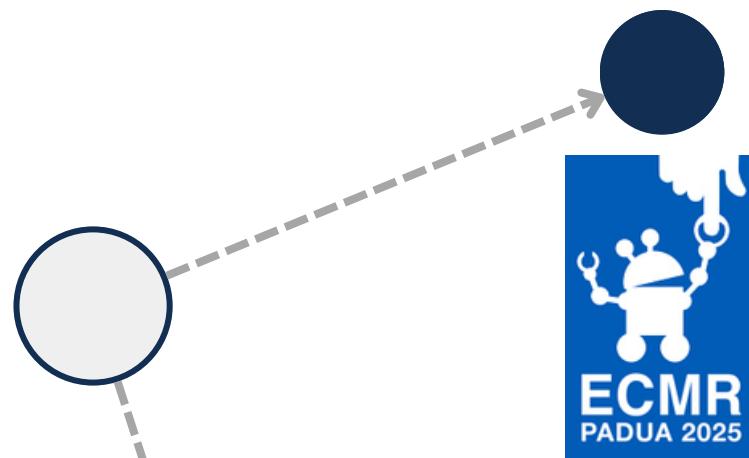
$$t_k \in \mathbf{p}_k^{\text{place}}$$

$$d_k \in \mathbf{p}_k^{\text{place}}$$

Etiqueta corta
Descripción



Métodos ----- 3. Lugares



Método 1

Segmentación

Descriptores geométrico-semánticos

Clustering

Refinamiento

Categorización

LLM

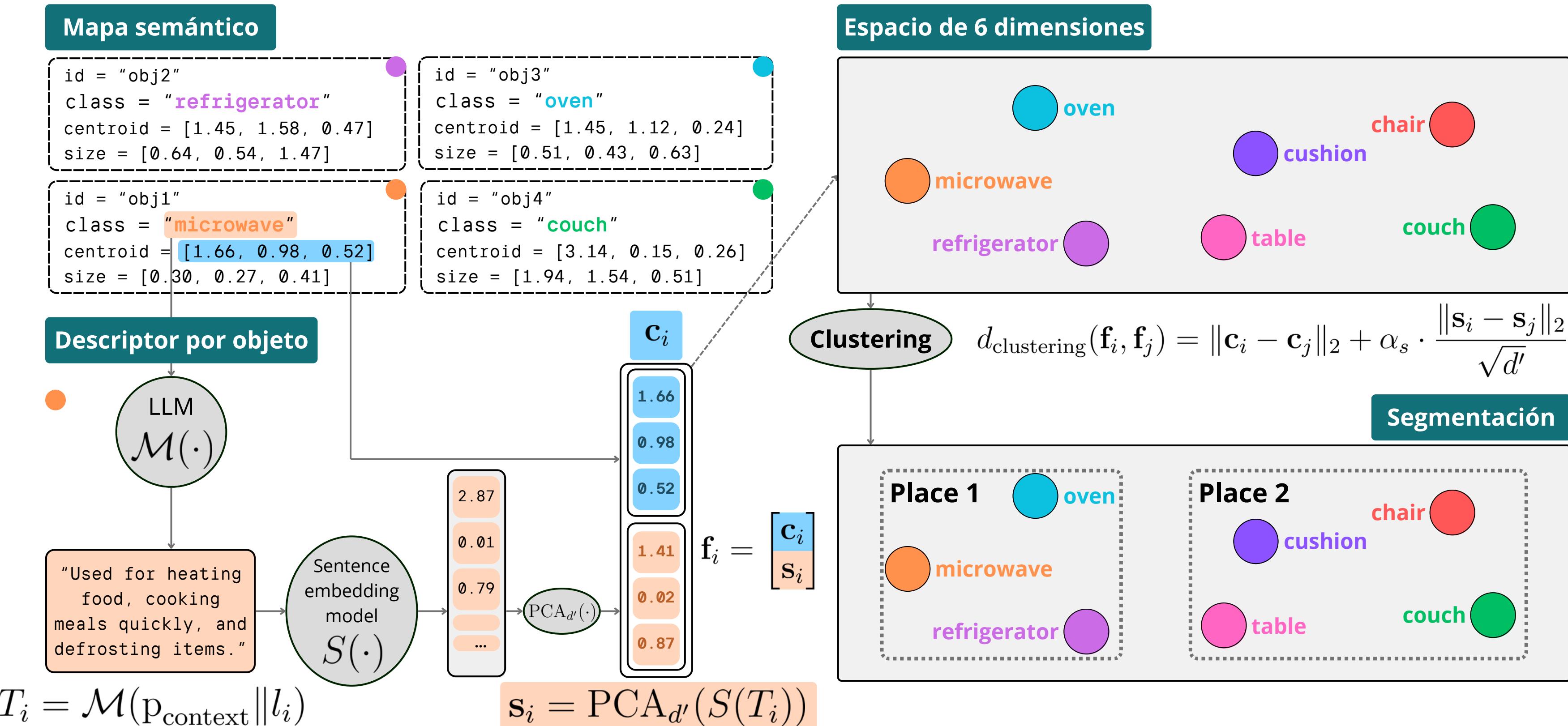
Método 2

Segmentación

Categorización

LLM

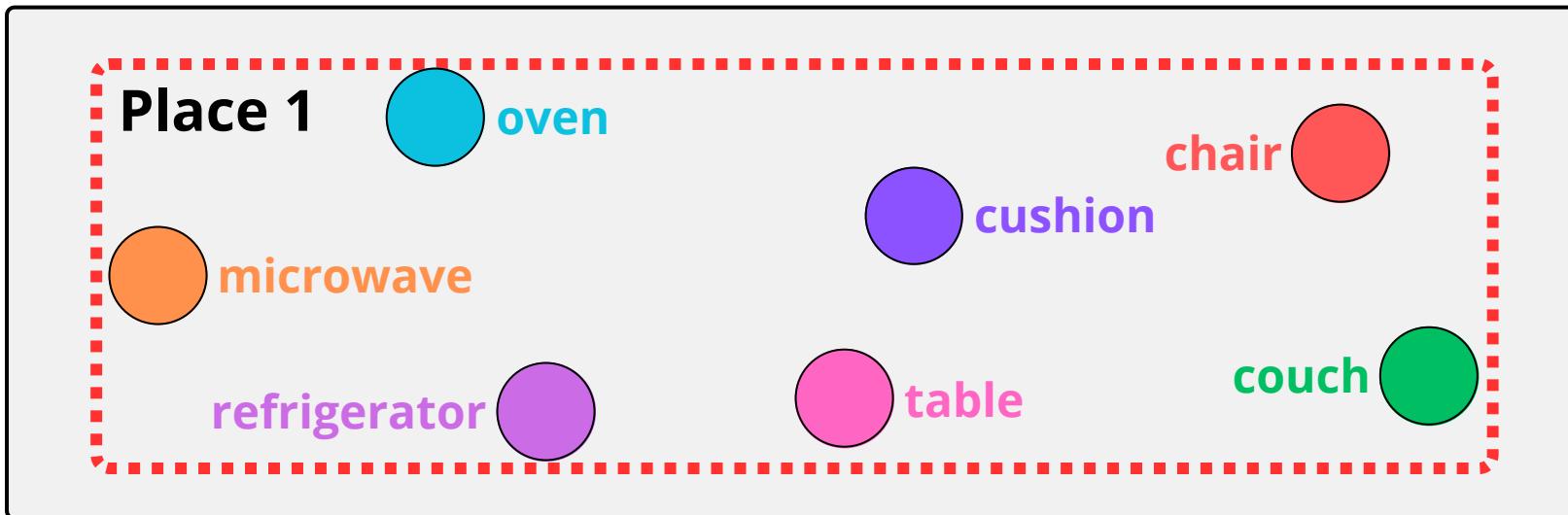
Método 1 | Clustering ----- 3. Lugares



Método 1 | Refinamiento

3. Lugares

Sub-segmentación



División

$$\sigma_k = \lambda \cdot \text{Var}_{\text{sem}}(P_k) + (1 - \lambda) \cdot \text{Dist}_{\text{avg}}(P_k)$$

Variabilidad semántica

$$\text{Var}_{\text{sem}}(P_k) = \frac{1}{n} \sum_{i=1}^n \|\mathbf{s}_i - \bar{\mathbf{s}}\|_2^2$$

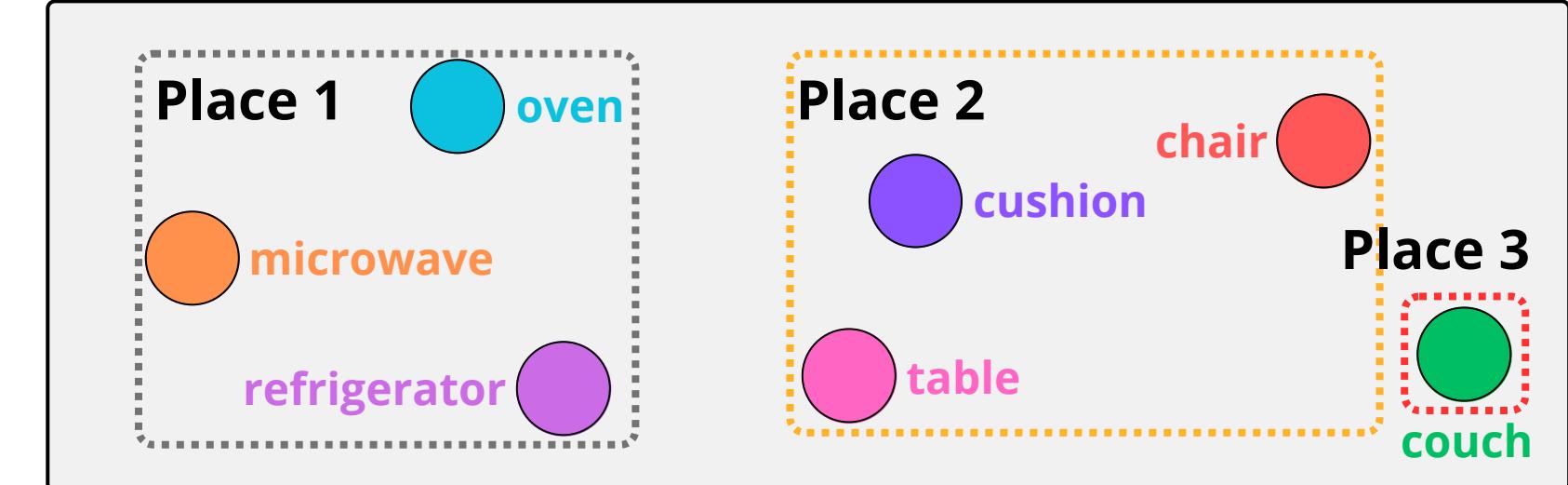
Dispersión geométrica

$$\text{Dist}_{\text{avg}}(P_k) = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j=i+1}^n \|\mathbf{c}_i - \mathbf{c}_j\|_2$$

Clustering

$$P_k \rightarrow \{P_k^{(1)}, \dots, P_k^{(m)}\}$$

Sobre-segmentación



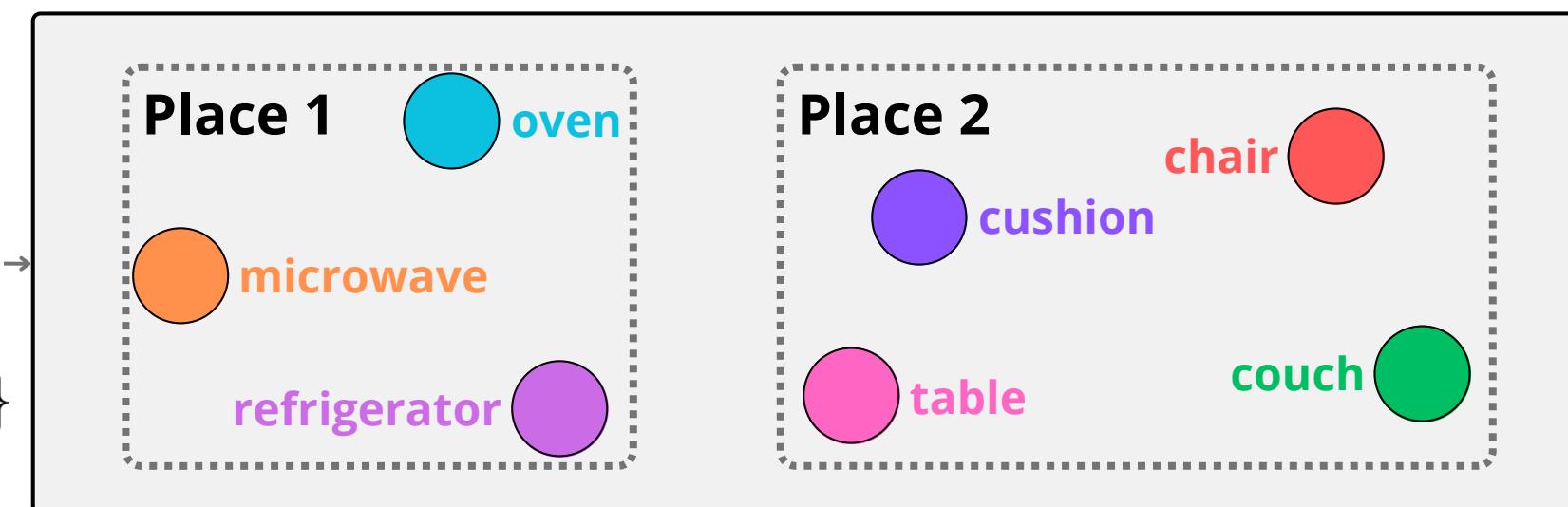
Unión

$$d_g(P_k, P_l) = \|\bar{\mathbf{c}}_k - \bar{\mathbf{c}}_l\|_2$$

Espacialmente próximos

Similitud semántica

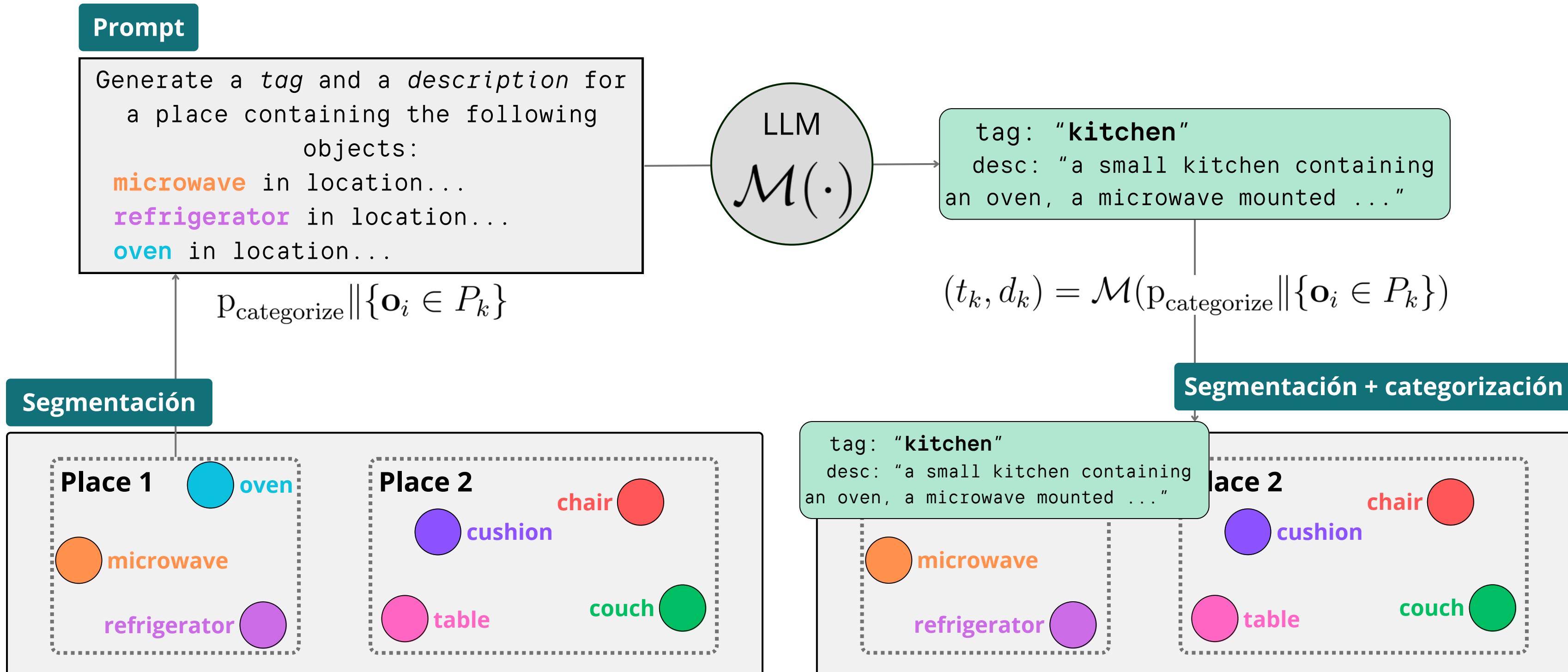
$$\text{sim}_{\text{cos}}(P_k, P_l) = \frac{\bar{\mathbf{s}}_k \cdot \bar{\mathbf{s}}_l}{\|\bar{\mathbf{s}}_k\|_2 \|\bar{\mathbf{s}}_l\|_2}$$



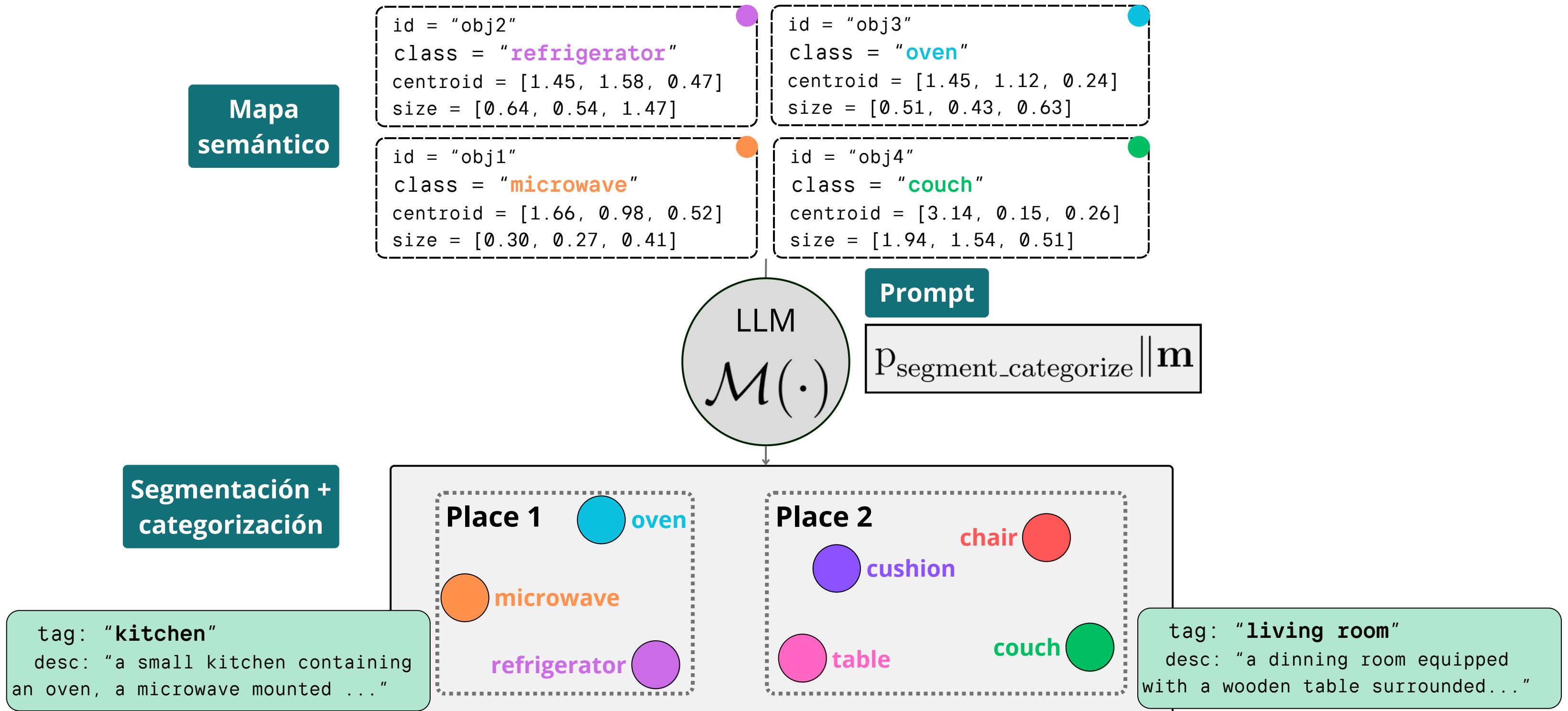
Unión

$$\{P_j^{(1)}, P_j^{(2)}\} \rightarrow P_k$$

Método 1 | Categorización ----- 3. Lugares



Método 2 | LLMs ----- 3. Lugares



$$(\mathcal{P}, \mathcal{P}_{\text{desc}}) = \mathcal{M}(p_{\text{segment_categorize}} \parallel \mathbf{m})$$

4

Inferencia de relaciones semánticas

Formulación del problema

4. Relaciones

Mapa semántico

$$\mathbf{m} = \{\mathbf{o}_1, \dots, \mathbf{o}_N \mid \mathbf{o}_i = (c_i, d_i, l_i, p_i^{\text{object}}), \forall i \in \{1, \dots, N\}\}$$

Centroide ← Tamaño ← Clase semántica → Propiedades

Inferencia de relaciones

$$\mathcal{R} = \varphi(\mathbf{m}) = \{\mathbf{r}_{ij}^{(k)} \mid i, j \in \{1, \dots, N\}, i \neq j, k \in \mathcal{K}_{ij}\}$$

donde cada relación $\mathbf{r}_{ij}^{(k)} = (\mathbf{o}_i, t_{ij}^{(k)}, p_{ij}^{(k)}, \mathbf{o}_j)$

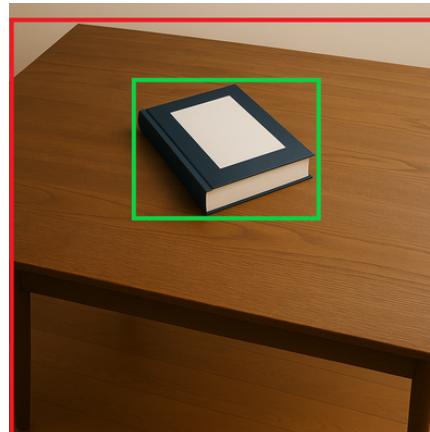
Obj. origen

Tipo

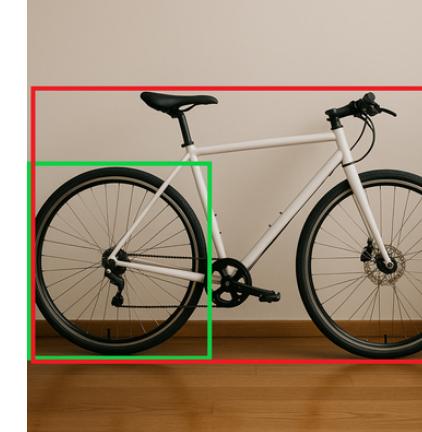
Predicado

Obj. destino

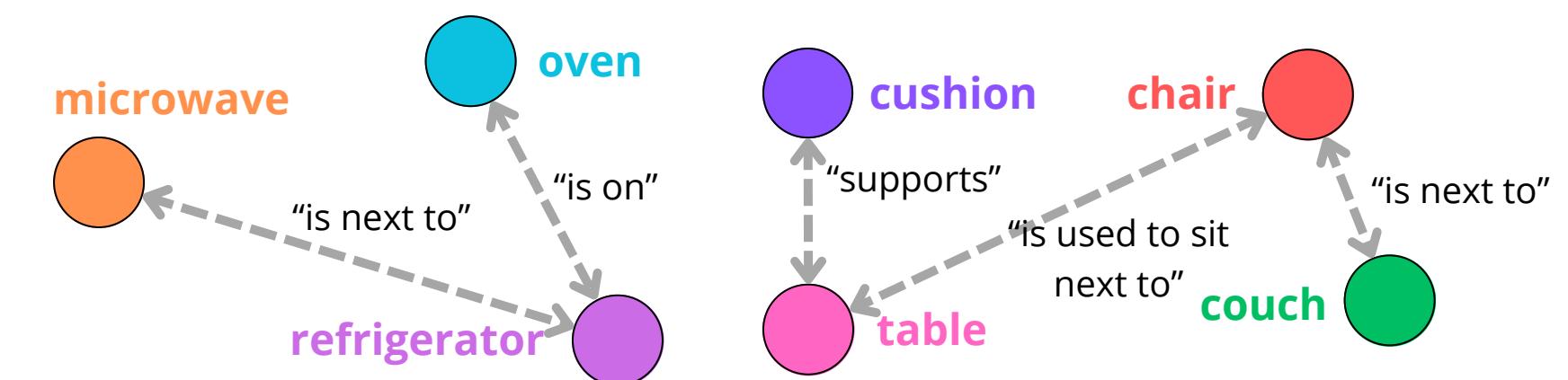
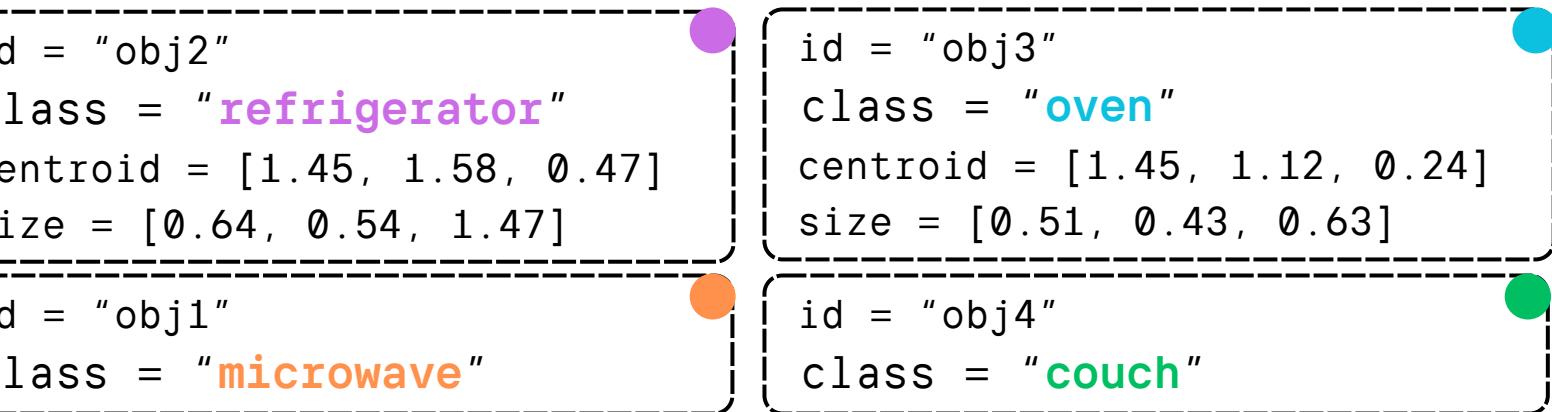
Tipos de relaciones



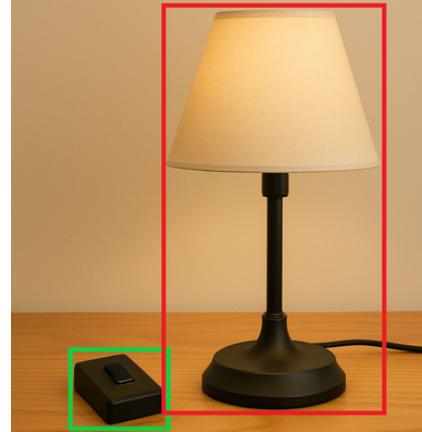
Espaciales



Estructurales



Funcionales



Causales

Métodos

4. Relaciones

(Basado en ConceptGraphs, Gu et al. 2024)

Identificación de
pares de objetos candidatos

$$\gamma_{ij} = \left\| \max \left(0, |c_i - c_j| - \frac{d_i}{2} - \frac{d_j}{2} \right) \right\|_2$$

Distancia entre AABBs



Método 1 | LLMs

4. Relaciones

Objetos candidatos

```
id = "obj2"  
class = "refrigerator"  
centroid = [1.45, 1.58, 0.47]  
size = [0.64, 0.54, 1.47]
```

```
id = "obj3"  
class = "oven"  
centroid = [1.45, 1.12, 0.24]  
size = [0.51, 0.43, 0.63]
```

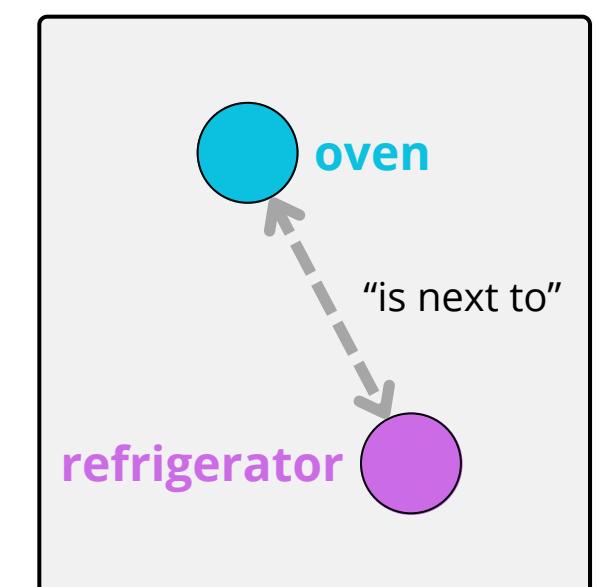
Prompt

Generate relationships between two objects:
- object 1 is a **refrigerator**, at location [...] and size [...]
- object 2 is an **oven**, at location [...] and size [...]

LLM
 $\mathcal{M}(\cdot)$

Respuesta

```
{  
    "source_id": "obj2",  
    "source": "refrigerator",  
    "target_id": "obj3",  
    "target": "oven",  
    "type": "spacial",  
    "predicate": "is next to"  
}
```



$$p_{\text{relationship_llm}} \| l_i, l_j, c_i, c_j, d_i, d_j$$

$$\{\mathbf{r}_{ij}^{(k)}\}_{k \in \mathcal{K}_{ij}} \cup \{\mathbf{r}_{ji}^{(k)}\}_{k \in \mathcal{K}_{ji}} = \mathcal{M}(p_{\text{relationship_llm}} \| l_i, l_j, c_i, c_j, d_i, d_j)$$

Método 2 | LVLMs ----- 4. Relaciones

Objetos candidatos

```

id = "obj2"
class = "refrigerator"
centroid = [1.45, 1.58, 0.47]
size = [0.64, 0.54, 1.47]

```

```

id = "obj3"
class = "oven"
centroid = [1.45, 1.12, 0.24]
size = [0.51, 0.43, 0.63]

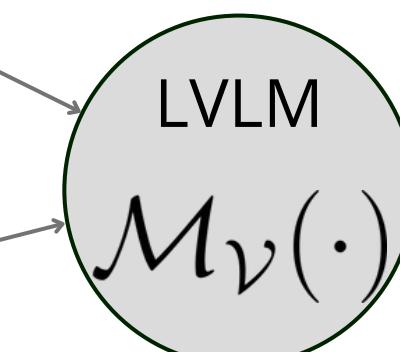
```

Prompt

Generate relationships between two objects:

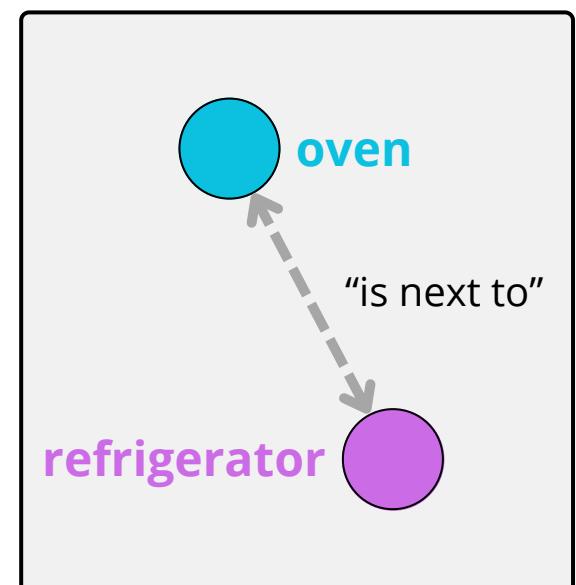
- object 1 is a **refrigerator**, at location [...] and size [...]
- object 2 is an **oven**, at location [...] and size [...]

Imagen óptima


 $p_{\text{relationship_lvlm}} \| l_i, l_j, c_i, c_j, d_i, d_j$


Respuesta

```
{
  "source_id": "obj2",
  "source": "refrigerator",
  "target_id": "obj3",
  "target": "oven",
  "type": "spacial",
  "predicate": "is next to"
}
```



$$\{\mathbf{r}_{ij}^{(k)}\}_{k \in \mathcal{K}_{ij}} \cup \{\mathbf{r}_{ji}^{(k)}\}_{k \in \mathcal{K}_{ji}} = \mathcal{M}_{\mathcal{V}}(p_{\text{relationship_lvlm}} \| l_i, l_j, c_i, c_j, d_i, d_j \| I^*)$$

$$I^* = \arg \max_{I \in \mathcal{I}} \left[\text{AreaProyectada}(\mathbf{o}_i, I) + \text{AreaProyectada}(\mathbf{o}_j, I) \right]$$

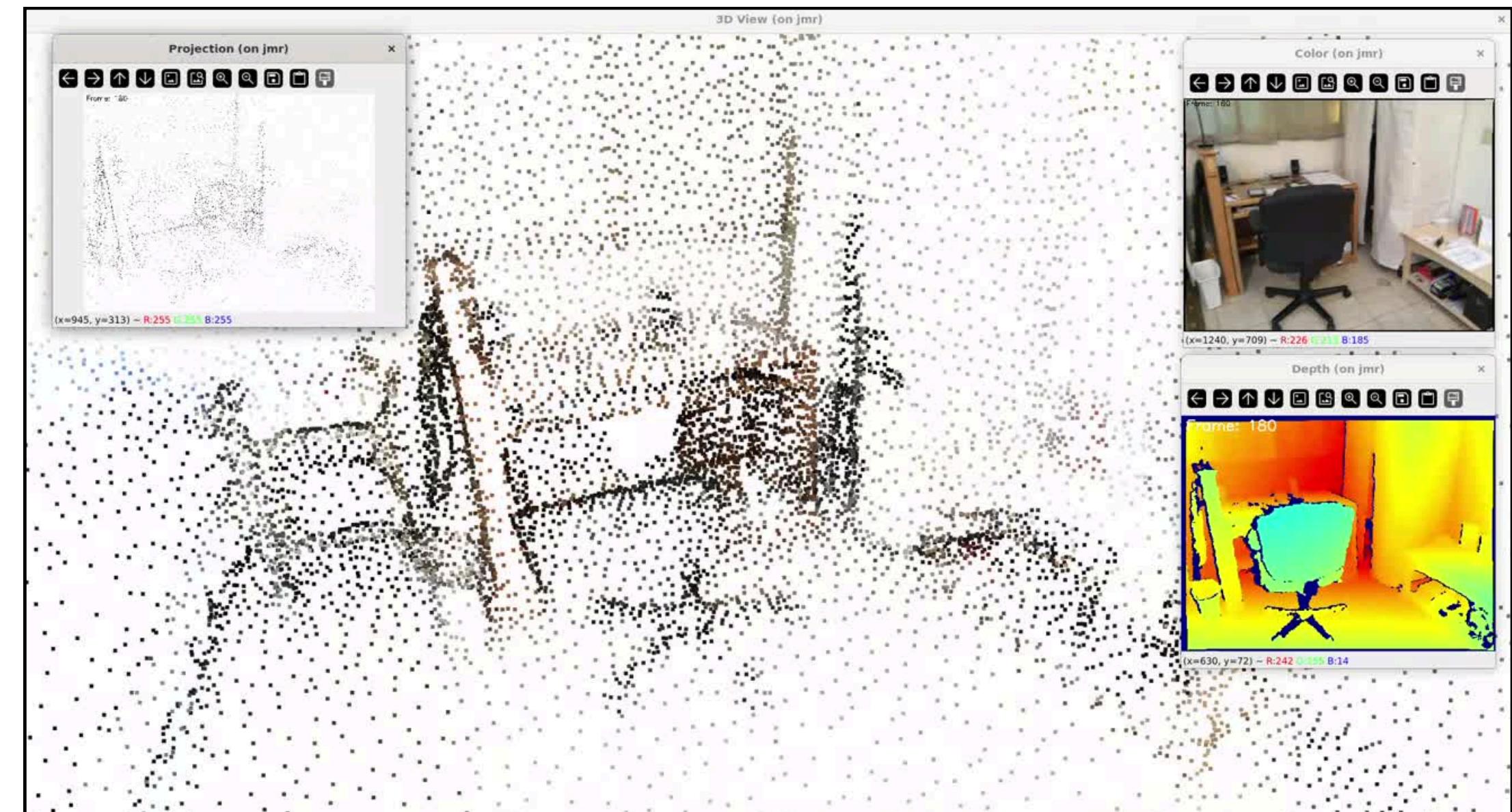
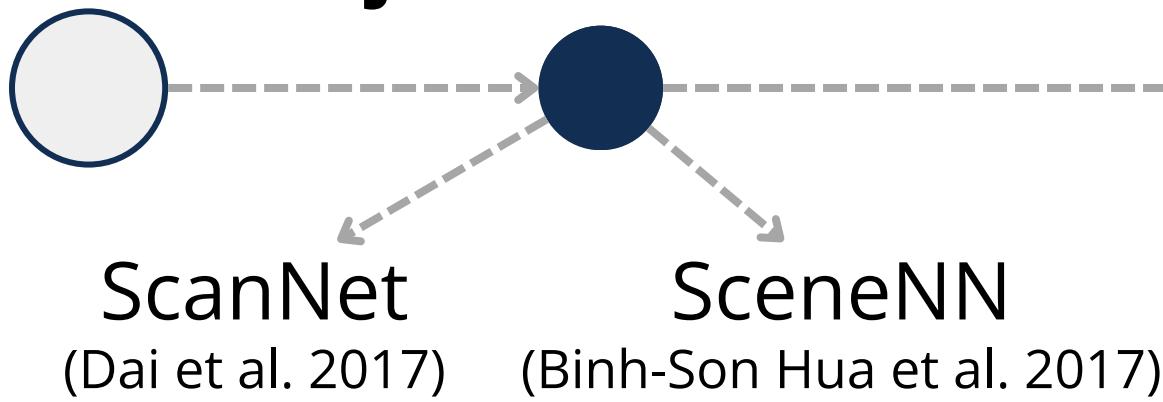
5

validación

Configuración experimental -----

5. Validación

Conjuntos de datos

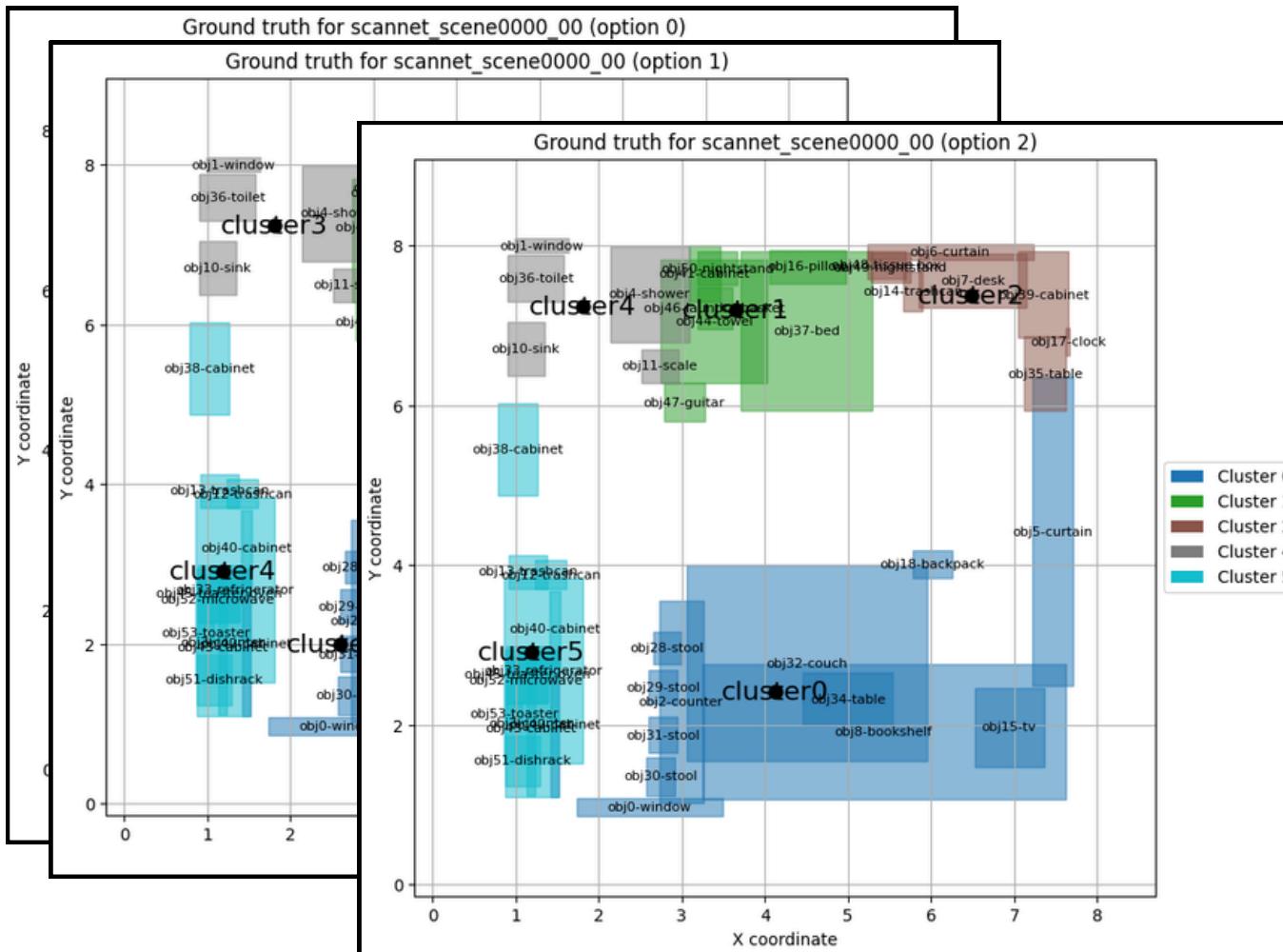


Configuración experimental

5. Validación

Ground truth de segmentaciones

$$\mathcal{P}_{\text{GT}}^{(j,\ell)} = \left\{ (P_k, t_k, d_k) \mid k \in K^{(j,\ell)} \right\}$$



Métricas de **clustering**

Adjusted Rand Index (ARI)

Normalized Mutual Information (NMI)

V-Measure

Evaluaciones

(Cuantitativa) Métodos de **segmentación**

Estudio de ablación

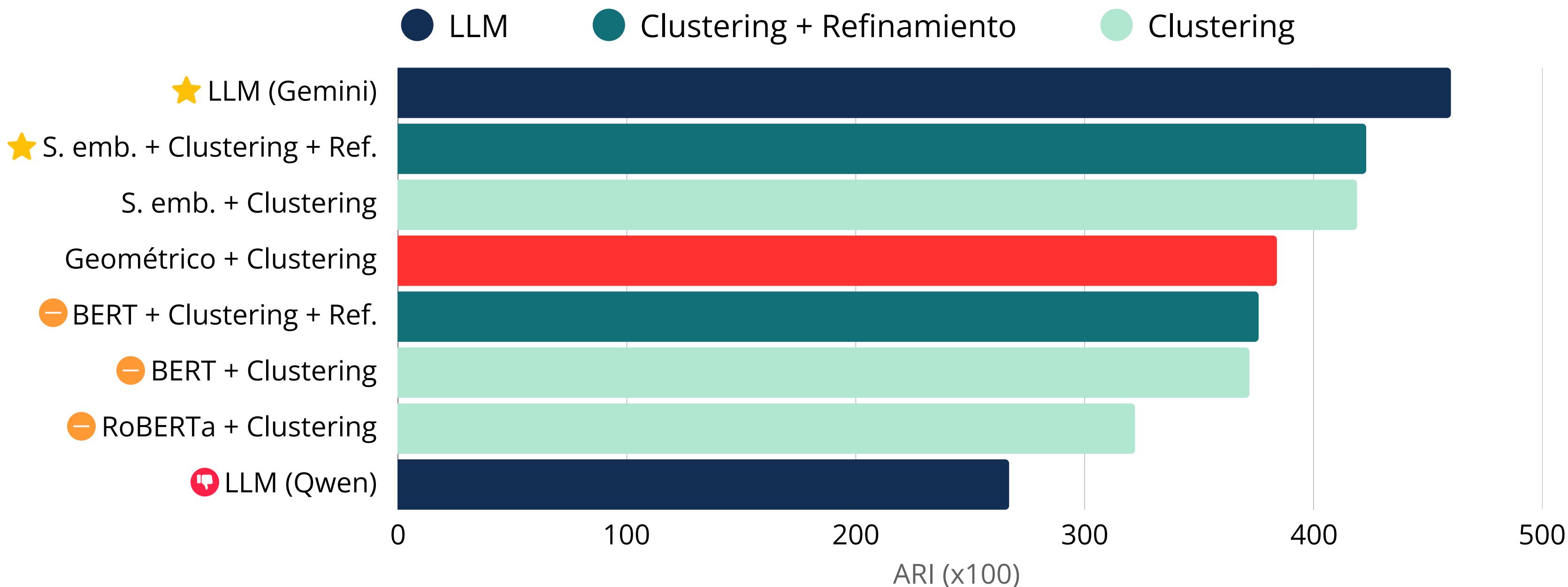
(Cualitativa) Métodos de **categorización**

(Cualitativa) Métodos de inferencia de **relaciones**

Desempeño de un robot con mapa semántico topológico

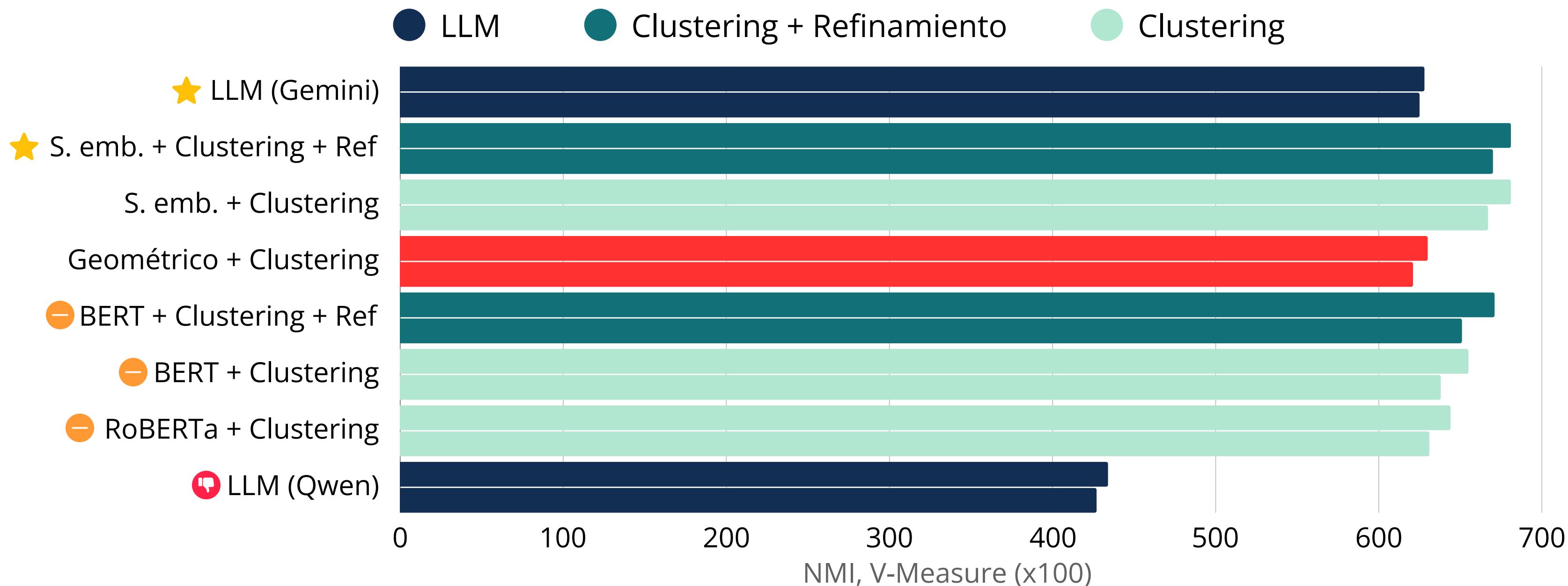
Métodos de segmentación ----- 5. Validación

Resultados para ARI



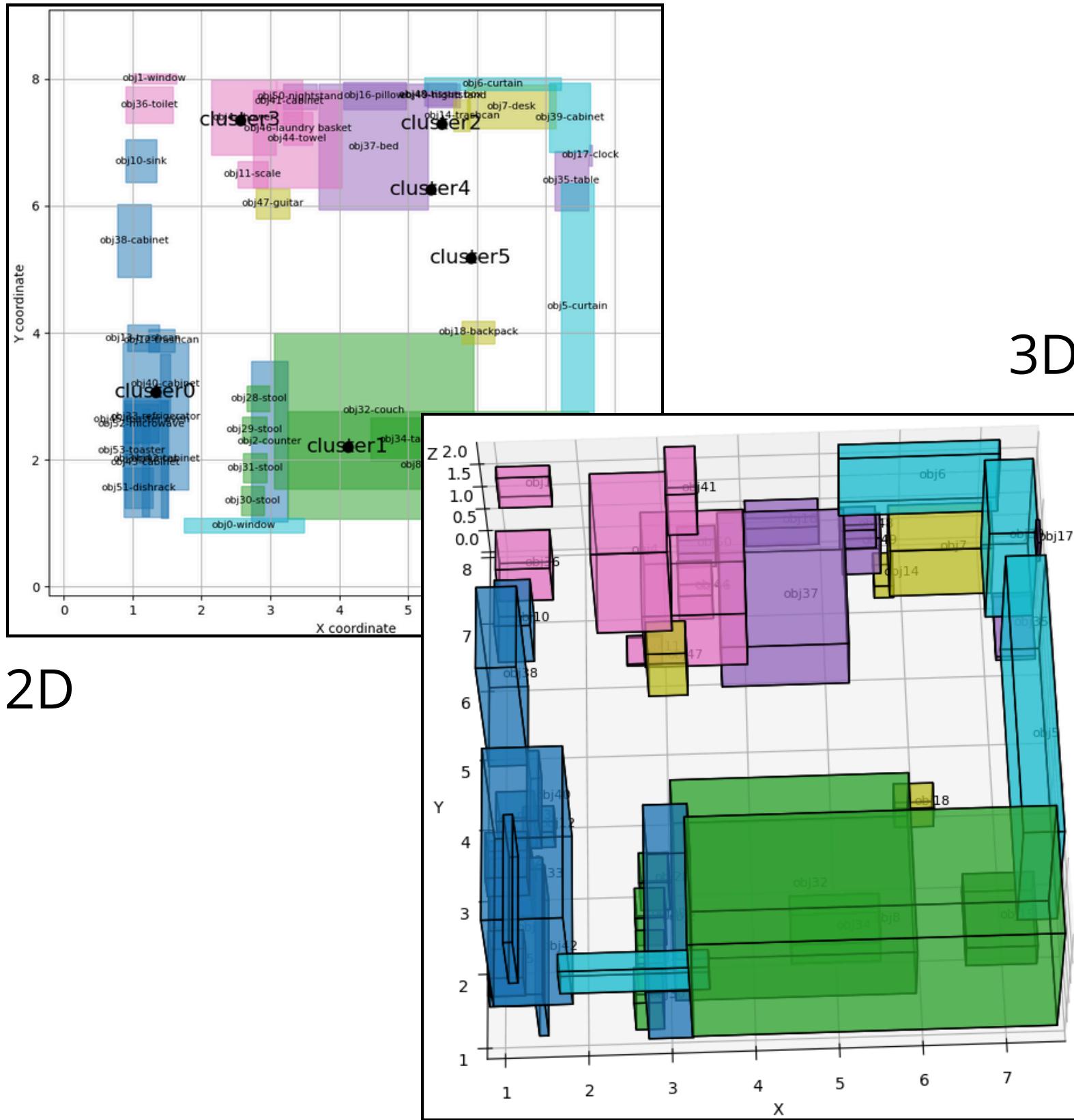
Métodos de segmentación ----- 5. Validación

Resultados para NMI, V-Measure



Métodos de segmentación

5. Validación



Métodos de categorización -----

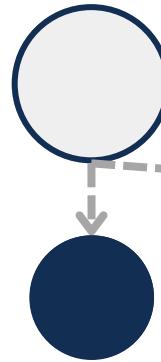
5. Validación



(scannet_scene0101_00)

Métodos de categorización -----

5. Validación



Problema: muy dependiente de la segmentación



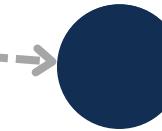
Sobre-segmentación



LLM
 $\mathcal{M}(\cdot)$

storage

A space primarily intended for storing or organizing items, characterized by a shelf likely used to hold supplies, tools, or personal belongings...



Sub-segmentación



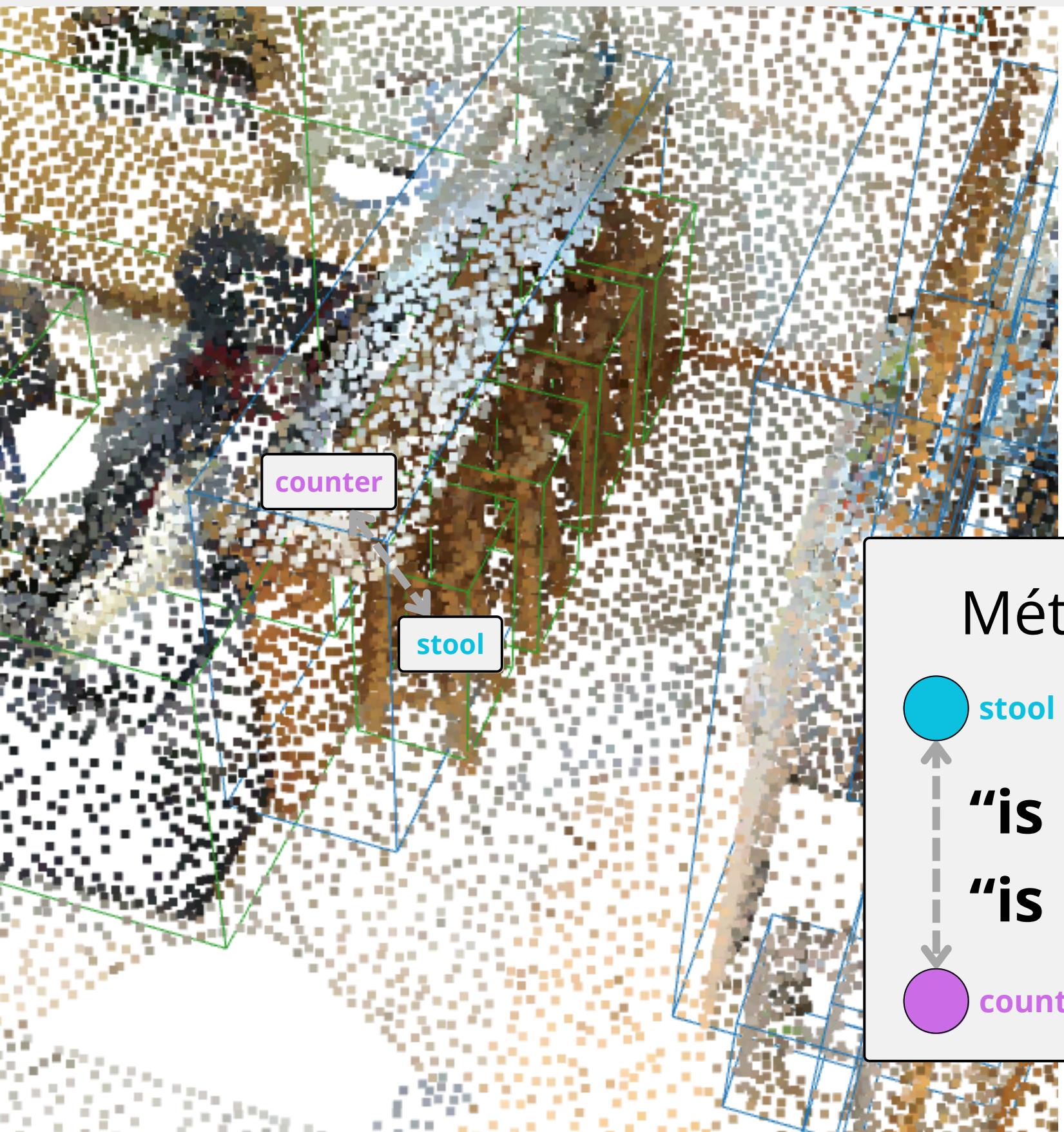
LLM
 $\mathcal{M}(\cdot)$

living_and_kitchen

This area serves as a multifunctional space combining living and kitchen functionalities, featuring seating arrangements like a couch and chairs, work areas such as desks and tables, and kitchen appliances including ...

Métodos de inferencia de relaciones

5. Validación

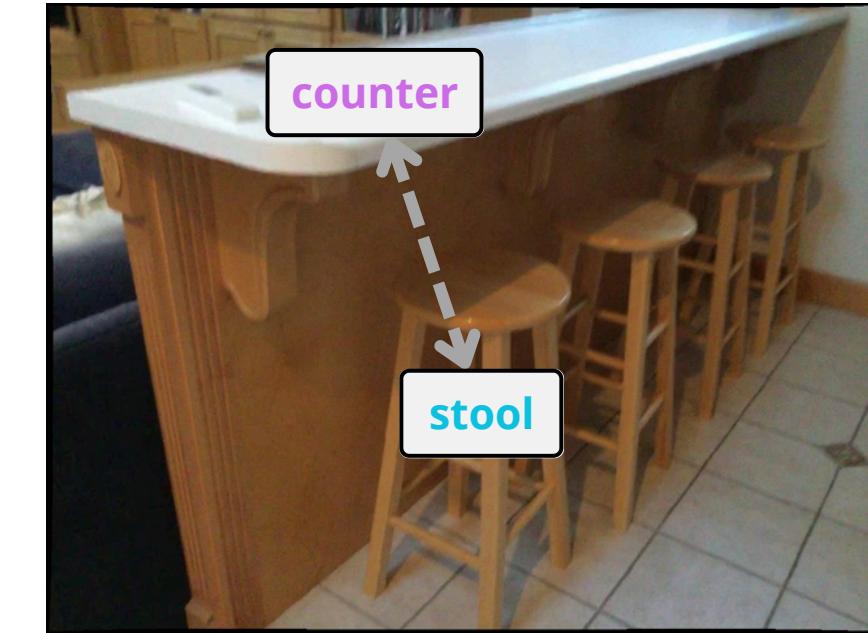


Método LLM

stool

“is next to”
“is used with”

counter



Método LVLM

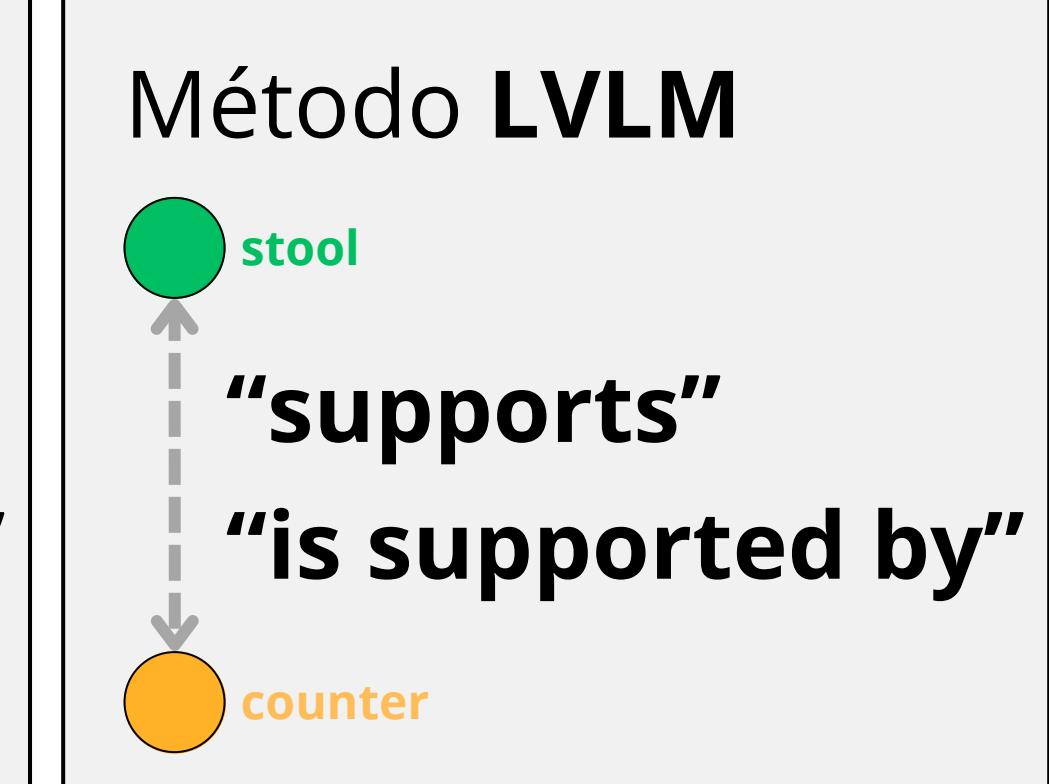
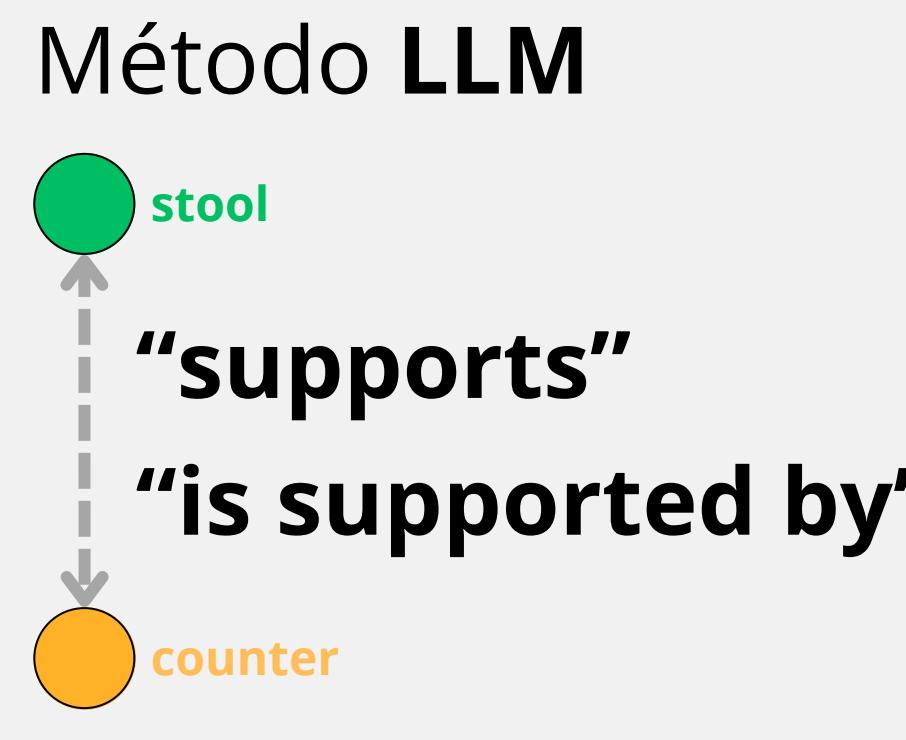
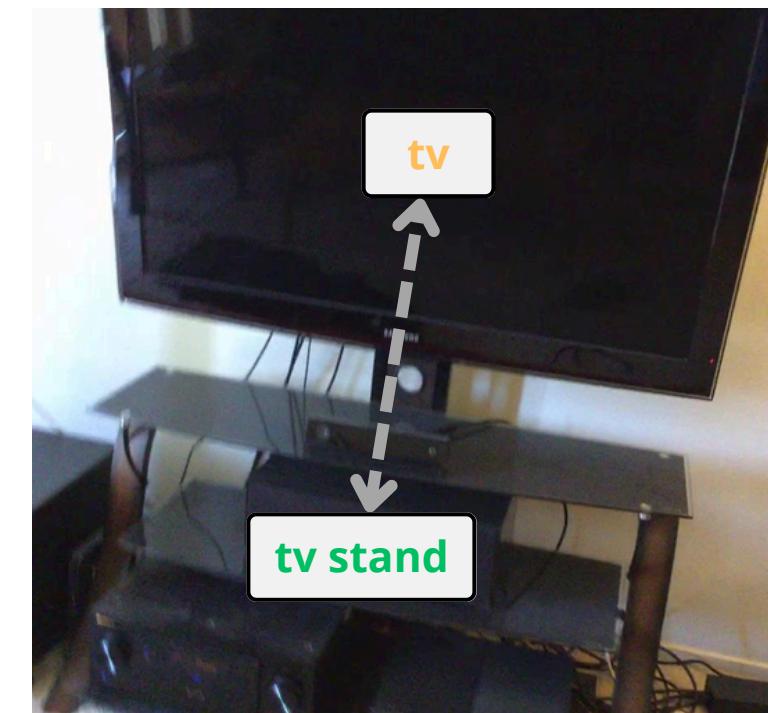
stool

“is partially under”
“provides seating area for”
“provides seating for”

counter

Métodos de inferencia de relaciones

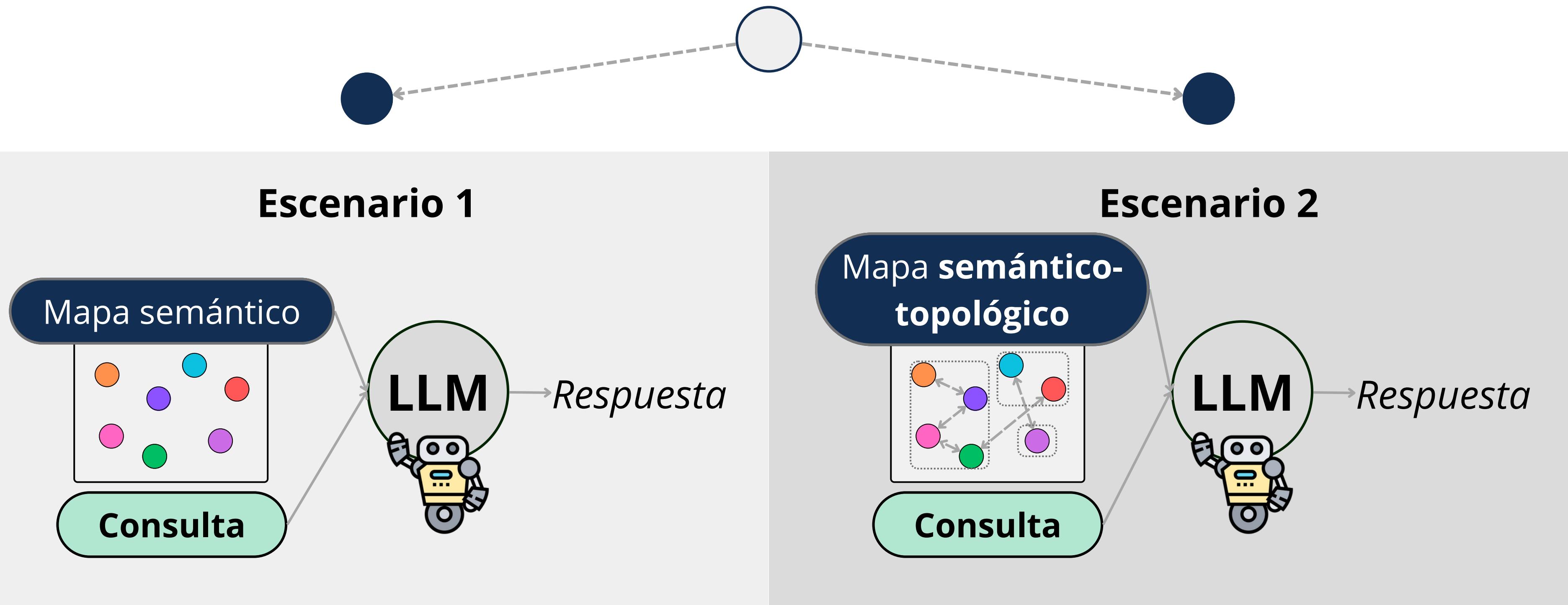
5. Validación



Desempeño con mapa semántico-topológico

5. Validación

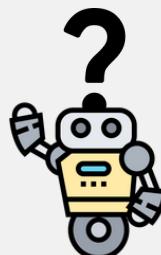
Robot equipado con LLM evaluado en **2 escenarios...**



Desempeño con mapa semántico-topológico

5. Validación

Razonamiento espacial



Which stool is between two other stools in the scene?

| Modelo | Escenario 1 (mapa básico) | Escenario 2 (mapa Enriquecido) |
|---------------------|--|--|
| Gemini 2.0 Flash | <i>The stool obj29 is between obj28 and obj31, as it is spatially next to both of them</i> | <i>The stool obj31 is between obj30 and obj29</i> |
| Qwen 2.5 14B | <i>Stool obj30 is between stool obj29 and stool obj31 in the scene</i> | <i>Stool obj29 is between stool obj28 and stool obj31 in the scene</i> |

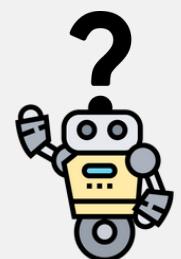
Desempeño con mapa semántico-topológico

5. Validación

Razonamiento estructural



| Modelo | Escenario 1 (mapa básico) | Escenario 2 (mapa enriquecido) |
|------------------------|---|---|
| Gemini 2.0 Flash | <p><i>The objects on the countertop are the dishrack (obj51), microwave (obj52), and toaster (obj53)</i></p> | <p><i>The objects on the countertop (obj2 and obj3) include the toaster oven (obj45), dishrack (obj51), and toaster (obj53)</i></p> |
| Qwen 2.5 14B | <p><i>The semantic map does not explicitly define a 'countertop' class, but instances obj2 and obj3 are labeled as 'counter'. No other objects are specified to be directly on these counters based on the given information</i></p> | <p><i>The objects on the countertop are obj45 (toaster oven), obj51 (dishrack), and obj53 (toaster)</i></p> |

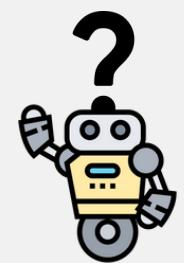
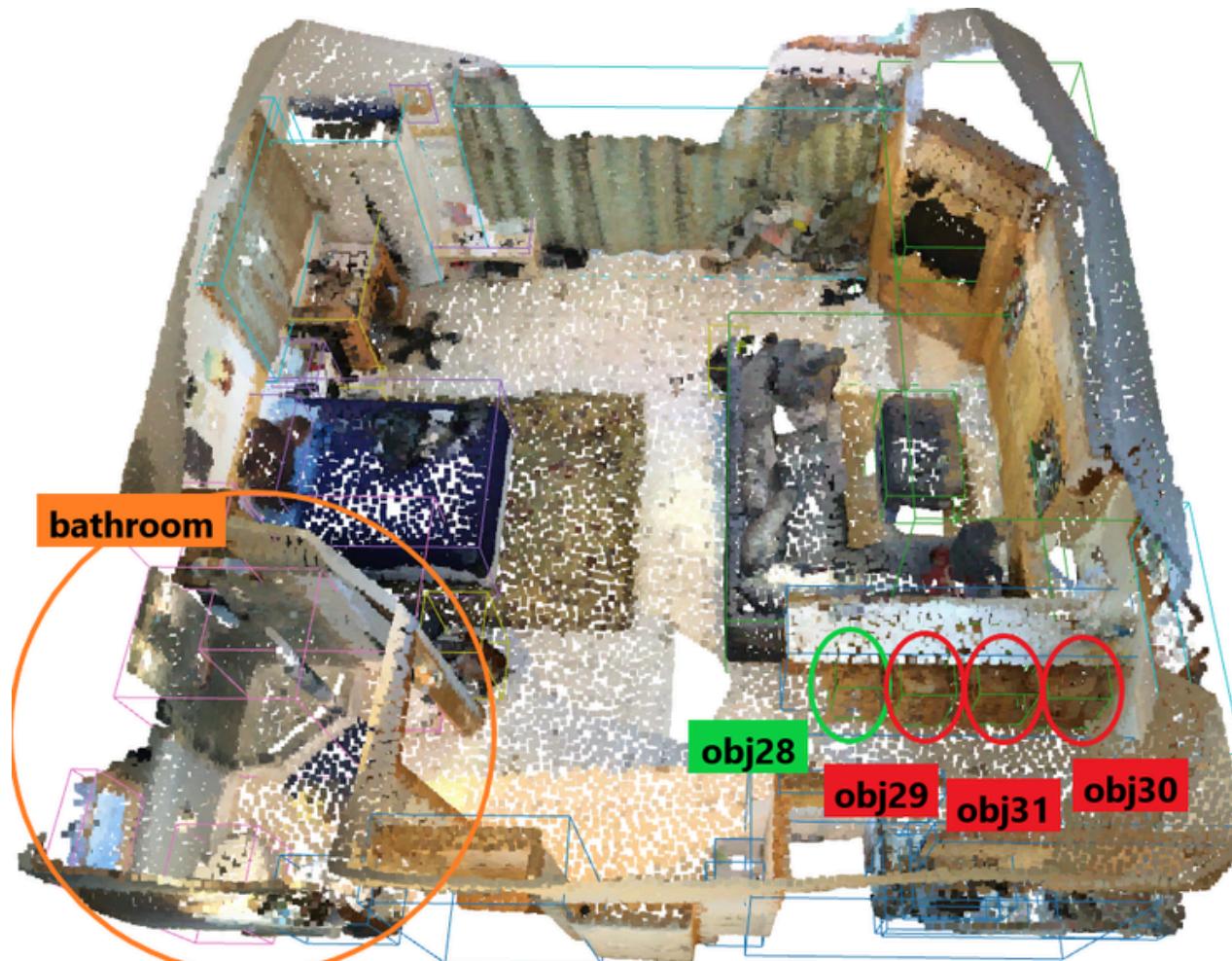


What objects are in the countertop?

Desempeño con mapa semántico-topológico

5. Validación

Razonamiento con lugares



Which stool is closest to the bathroom?

| Modelo | Escenario 1 (mapa básico) | Escenario 2 (mapa enriquecido) |
|-----------------------------|--|--|
| Gemini 2.0 Flash | <p><i>The closest stool to the bathroom is obj30, located near the counters and closest to the toilet (obj36) and shower (obj4)</i></p> | <p><i>The stool closest to the bathroom is obj29</i></p> |
| Qwen 2.5 14B | <p><i>Obj29, obj30, and obj32 are stools, but obj30 is closest to potential bathroom objects like the toilet (obj36) and shower (obj4)</i></p> | <p><i>The stool closest to the bathroom is obj28, which is located next to counter obj2 in the living room area</i></p> |

Desempeño con mapa semántico-topológico

5. Validación

Desempeño general

| Modelo | Escenario 1 | Escenario 2 |
|------------------|-------------|-------------|
| Gemini 2.0 Flash | Orange | Light Green |
| Qwen 2.5 14B | Red | Light Green |

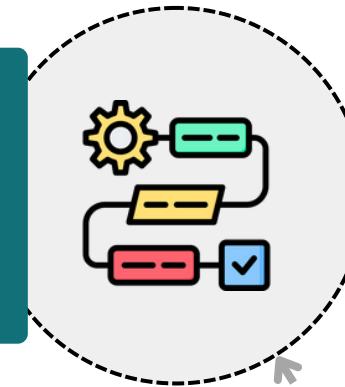
6

Conclusiones

Conclusiones

6. Conclusiones

Mapa semántico topológico



- Métodos de **segmentación** y **categorización** de lugares.
- Métodos de **inferencia de relaciones semánticas** entre elementos.
- **Validación** de las propuestas.
- Evaluación del desempeño de un robot equipado con LLM con mapa semántico-topológico

Limitaciones



- Coste económico y computacional de LLMs
- Solo se considera centroide de objetos
- Lugares que atraviesan paredes

Líneas futuras



- Evaluación cuantitativa del desempeño
- Categorización de lugares con LVLMs
- Elementos inamovibles para segmentación
- Integración en flujo de mapeo semántico completo

Creación de Mapas Semántico-Topológicos Para Robótica Móvil Utilizando Modelos de Aprendizaje Profundo

Building Semantic-Topological Maps for Mobile Robotics Using
Deep Learning Models

Realizado por:

Jesús Moncada Ramírez

Tutorizado por:

José Raúl Ruiz Sarmiento

Antonio Javier González Jiménez

Máster en Ingeniería Mecatrónica

Málaga, julio de 2025



UNIVERSIDAD DE MÁLAGA

