

**Cultural image (test set)**

**Step 1: VLM Spanish captioning**

Qwen2.5-VL-72B-Instruct (4-bit) or  
Qwen3-VL-8B

Prompt: culturally-aware, noun-first style

**Spanish caption (1 sentence per image)**

**Step 2: BM25 retrieval  
+ dev exemplar selection**

Retrieve 80 nearest ES-GN pairs  
Select 49 dev pairs (leave-one-out)  
No reranking in frozen config

Training  
**53K**

Dev  
**50**

**Step 3: Many-shot prompt assembly**

System: "translate ES-GN, match example style"  
49 gold dev pairs + 80 retrieved train pairs  
~3–5K tokens per prompt

**Step 4: Gemini 2.5 Flash translation**

temperature=0.0, thinking=off, max\_tokens=120  
Learns register + vocabulary from  
in-context examples

**Step 5: Clean + format**

Strip prefixes, normalize whitespace

**Guaraní caption (submission JSONL)**

Evaluated by chrF++  
+ human judges (top 5)

**Key: mBART replaced by Gemini  
many-shot ICL (+16 chrF++ gain)**