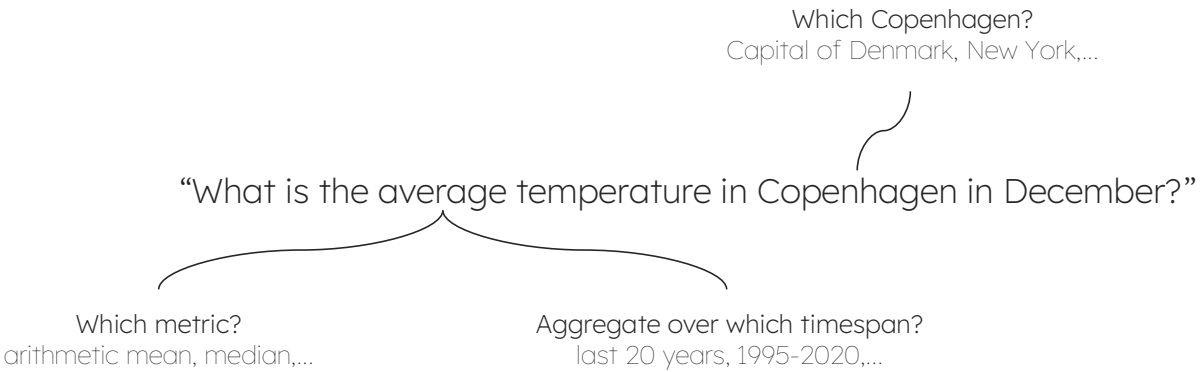
Are We Asking the Right Questions? On Ambiguity in Natural Language Queries for Tabular Data Analysis

EurIPS 2025 Presentation

Daniel Gomm, Cornelius Wolff, Madelon Hulsebos

Centrum Wiskunde & Informatica (CWI), University of Amsterdam

Ambiguities are everywhere in natural language interactions



Why do users formulate ambiguous queries?	How should systems and evaluations handle these ambiguities?	Are popular datasets fit to test system accuracy?

Users cooperate with systems

Which Copenhagen?
Capital of Denmark, New York,...

"What is the average temperature in Copenhagen in December?"

Which metric?
arithmetic mean, median,...

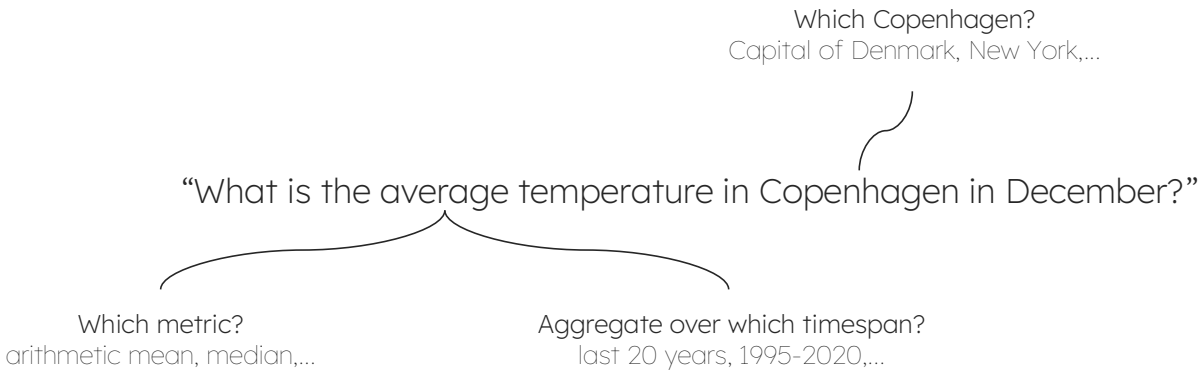
Aggregate over which timespan?
last 20 years, 1995-2020,...

Users cooperate with systems

Maxim of quality:
Users are *truthful*

Maxim of quantity:
Users provide *sufficient but not excessive information*

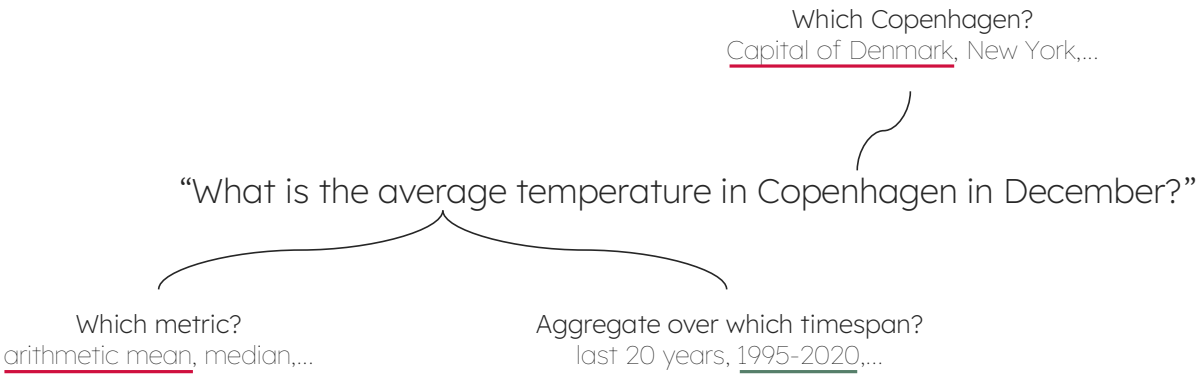
Users formulate cooperative queries



Users that follow the cooperative principle formulate **cooperative queries**.

→ Cooperative queries allow discriminating if interpretations are aligned with the user intent.

Query grounding in cooperative interaction



What do users expect a system does with ambiguities?

Conventional Grounding
There is a **convention** on what the user refers to

Selective Grounding
The user expects that the system can **select** a good option (division of labor)

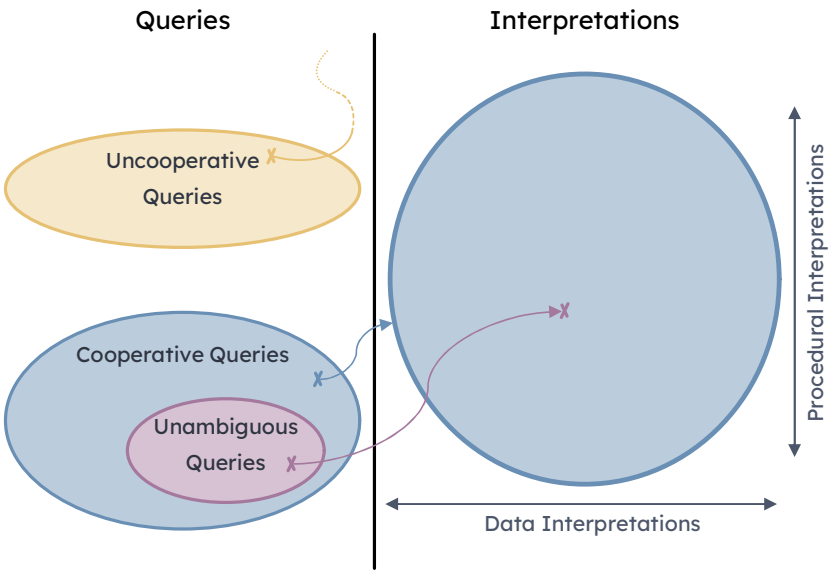
Uncooperative queries

If we cannot apply conventional or selective grounding to interpret ambiguities, users (unintentionally) violate the cooperate principle. We term this **uncooperative queries**.

"What is the average temperature?"

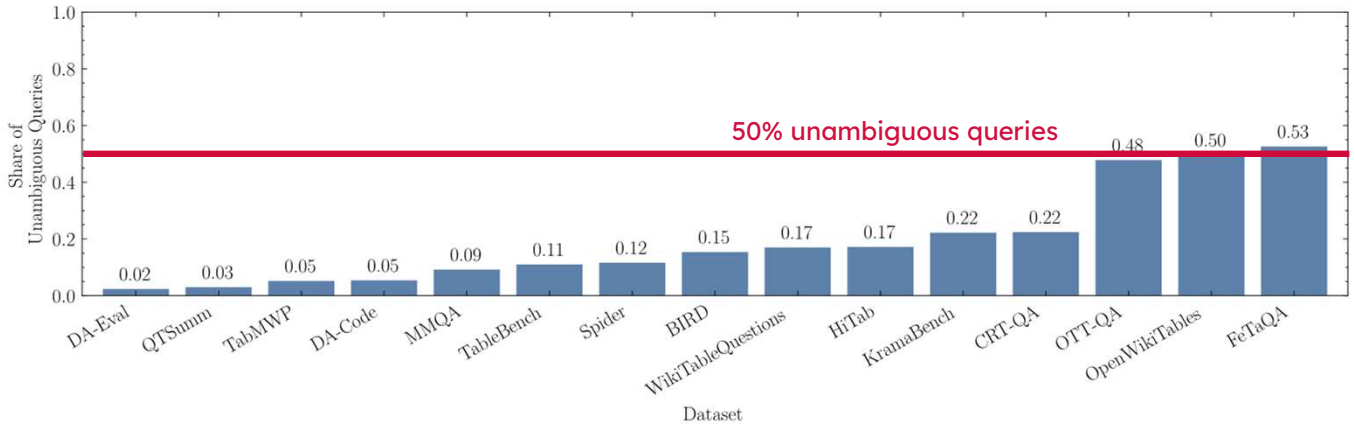
Where?
There is **no indication that any given selection of a location is aligned** with the user intent.

Grounding queries yields interpretations



→ (execution) **accuracy** against ground truth can only be evaluated with **unambiguous queries**

Are popular datasets fit to test system accuracy?



→ Benchmarks fail to isolate accuracy from systems interpretation capabilities.

Key take-aways

Let's build better datasets.

- Make ambiguities explicit in existing and new datasets
- Build datasets for iterative query specification to enable evaluations of interpretation capabilities

Let's build better systems.

- Empower systems to take proactive grounding decisions, dynamically bounded by the context
- Go beyond single-shot interactions towards more cooperative dialogs

We revisit ambiguity, understanding it as natural part of **cooperative interaction** instead of a problem to fix.

Blog Post

Paper