Atomic Propositions in NLP

From Applications to Evaluation

Luc Pommeret

Laboratoire Interdisciplinaire des Sciences du Numérique
Université Paris-Saclay

September 26, 2025

What are Atomic Propositions?

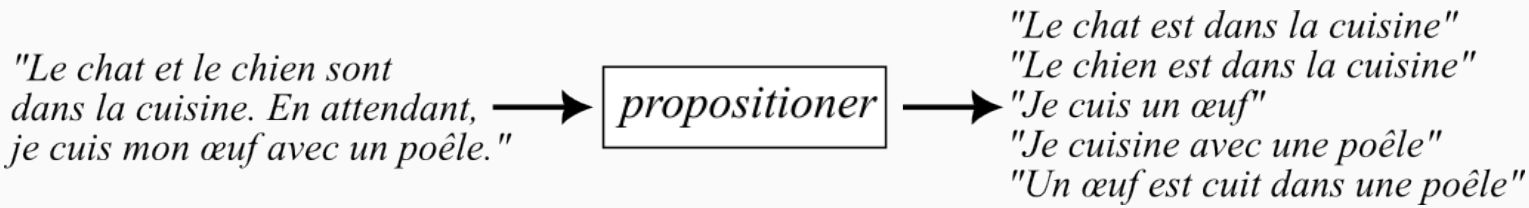

Figure 1: The propositionneur: transforming complex text into atomic propositions

Definition (Chen et al. 2024)

Atomic propositions are self-contained, factual statements that cannot be decomposed further without losing their semantic meaning.

Key Properties:

- **Autonomy:** Each proposition is interpretable in isolation
- **Atomicity:** Cannot be broken down into smaller meaningful units
- **Factuality:** Contains a single, verifiable fact

Applications of Atomic Propositions

Why is decomposition useful?

1. **Fact-Checking (FACTScore):**
   - Decomposes generated text into atomic facts
   - Verifies each fact independently against reliable sources
   - Enables fine-grained evaluation of factual accuracy

2. **Summary Evaluation (Herserant et al.):**
   - Breaks down summaries into atomic propositions
   - Evaluates coverage and precision at granular level
   - 75.1% faster than traditional evaluation methods

3. **Dense Retrieval (Chen et al. 2024):**
   - Uses atomic propositions as retrieval units in RAG systems
   - Improves recall by +10.1% (Recall@20)
   - Optimal granularity: 11.2 words per proposition

BUT: No one systematically evaluates whether these propositions are **truly atomic**!

AtomicEval: Evaluating Atomicity

No systematic evaluation of whether atomic propositions are truly autonomous!

Example from FACTScore:

- Proposition: *"He appeared as an actor."*
- Marked as "supported" by FACTScore
- **But:** Who is "He"? $\Rightarrow$ **Non-autonomous**

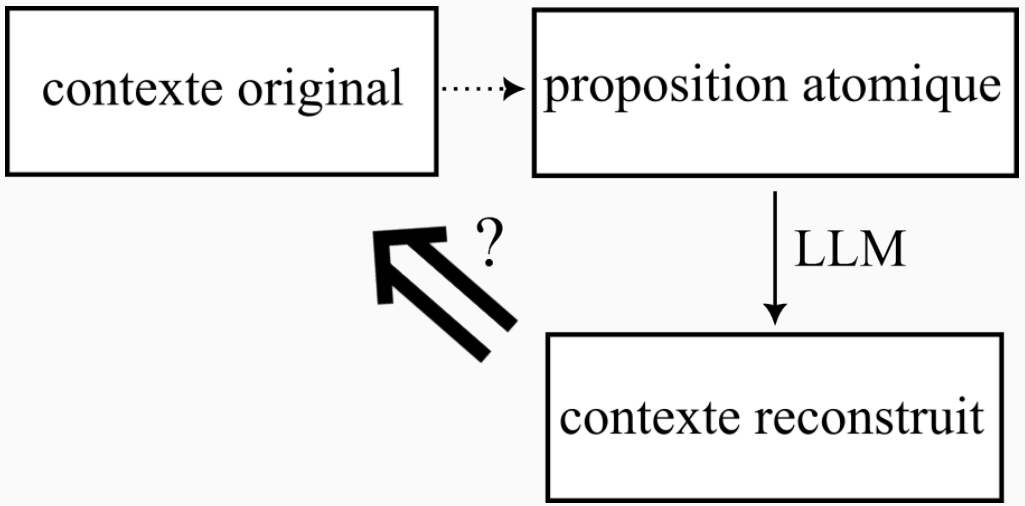

Figure 2: AtomicEval framework: Testing if reconstructed context implies original context