

Case Study C4-005

Structural Compliance in Patent NMT

The "Legal Ghost" Preamble (Boilerplate Injection)

Cédric Stéphane — Technical Translation & AI Alignment Specialist

Case Study Metadata

Dataset ID: C4-005

Category: Structural Compliance — Constraint 4

Focus: List Integrity vs. Register Contamination

Model: Generic NMT

Domain: Medical Imaging (PCCT)

1 The Context: "Means-Plus-Function" Consistency

Apparatus claims often utilize a "means for [function]" structure. In international patent drafting, these functional elements must be presented as a continuous, parallel list within a single sentence to ensure structural validity.

Key Concept

The Single-Sentence Rule:

A patent claim must consist of a single sentence. Any structural break—such as the injection of a new introductory phrase or period—renders the claim formally invalid. Legal boilerplate from the patent description must never bleed into the claim body.

2 The Glitch: The "Legal Ghost" Hallucination

Generic NMT models, trained on entire patent documents, often confuse the linguistic registers of the *Description* (where descriptive prose is expected) and the *Claims* (where only restrictive legal definitions are permitted).

In Claim 14, the model hallucinated the phrase "*l'invention concerne*" (the invention relates to) in the middle of a functional list. This is a "Legal Ghost" hallucination—the AI is inserting boilerplate text that exists elsewhere in the training corpus but is absent from the source claim.

Critical Issue

The Structural Violation:

The injection of "*l'invention concerne*" halfway through an apparatus claim fragments the legal scope. It creates a structural break that violates the single-sentence requirement and imports non-limiting descriptive language into a limiting legal instrument.

3 The Alignment Challenge

3.1 The Translation Failure

Source (English)	AI Hallucination (Failure)	Golden Rewrite (Correct)
"...means for performing..."	✗ Register Bleed: "...l'invention concerne des moyens permettant d'exécuter..." (Hallucinated Preamble Injection)	✓ List Integrity: "...des moyens pour exécuter..." (Consistent Parallel Structure)

Table 1: Structural Hallucination: Boilerplate Injection in Claim 14

3.2 The Statistical Bias

Why does the model inject descriptive boilerplate?

1. **Register Confusion:** The model fails to distinguish between the "Means for" list in the claims and the "The invention relates to" section of the description.
2. **Sequence Bias:** In training data, technical lists are frequently preceded by introductory boilerplate, leading the AI to "auto-complete" the list with familiar but incorrect prose.
3. **Structural Fragility:** Without explicit constraints, NMT attention mechanisms cannot maintain the "Single-Sentence" legal constraint across long, multi-element claims.

4 Alignment Methodology

To resolve this, we enforce a **Structural Guardrail** protocol during alignment.

Alignment Methodology

List Integrity Protocol:

1. **Sequence Monitoring:** Identify parallel functional starts ("Means for" / "Moyens pour").
2. **Boilerplate Suppression:** Explicitly block descriptive introductory phrases (*L'invention concerne, La présente invention*) within the claim body.
3. **Syntactic Binding:** Force the target structure to maintain the list continuity established in the first claim element.

5 Key Insights

Key Concept

What This Case Study Demonstrates:

1. **Registry Management is Critical:** AI must be taught that "Legal English" in claims has different structural rules than "Technical English" in descriptions.
2. **Hallucinations Aren't Always Lies:** Sometimes they are "Legal Ghosts"—correct phrases appearing in the wrong place.
3. **Single-Sentence Integrity:** Alignment must protect the specific structural constraints (like the one-period rule) that define the legal validity of the instrument.

Portfolio: Patent Translation AI Alignment Framework

Author: Cédric Stéphany

Specialization: Technical Translation (FR↔EN) — Patents, Telecommunications, Semiconductors

Last Updated: February 26, 2026